Title
Structure and Functional Studies of Telomerase RNA

Author
Cash, Darian Dallas

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Structure and Functional Studies of Telomerase RNA

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biochemistry and Molecular Biology

by

Darian Dallas Cash

2015

ABSTRACT OF THE DISSERTATION

Structure and Functional Studies of Telomerase RNA

by

Darian Dallas Cash
Doctor of Philosophy in Biochemistry and Molecular Biology
University of California, Los Angeles, 2015
Professor Juli Feigon, Chair

This dissertation consists of four projects, three focused on Nuclear Magnetic Resonance (NMR) structural analysis applied to telomerase RNA (TER) and one on optimizing expression of telomerase proteins. Telomerase is the enzyme necessary for synthesis of telomere repeats at the 3′ end of linear chromsomes. It has a minimal composition of TER and telomerase reverse transcriptase (TERT). TER includes the template used for copying the telomeric repeat and an adjacent pseudoknot which is important for telomerase function. The structure of the human TER pseudoknot revealed it contains a triple helix with tertiary interactions that are essential for telomerase activity. The goal of the first project was to further study pseudoknot structure to gain insights into conserved features and pseudoknot function in telomerase. To this end, the NMR solution structure of the yeast *K. lactis* TER pseudoknot was determined. This pseudoknot contains an extended pyrimidine motif triple helix with a C-G-C^+ triple and three bulge nucleotides to maintain continuous base pairing and stacking interactions through the stems. Despite differences in sequence and base triples, the human and yeast pseudoknots have a remarkably similar tertiary shape, indicating a conserved function. In the second project, the

NMR solution structure of the ciliate *Tetrahymena thermophila* TER pseudoknot was determined. The *Tetrahymena* pseudoknot (tetPK) is more compact than the pseudoknots of human and yeast, however it maintains the conserved features of stacked stems and base triple interactions. TetPK contains a unique A-G-C base triple, which is shown to be important for pseudoknot stability and telomerase activity. The folding of tetPK was studied in the context of full length TER, where its formation was shown to be disrupted by competition with alternate structures. The third project analyzed NMR structure calculations of A-form RNA helices. Two parameters were determined to be essential for getting accurate helical structures, distance restraints derived from sequential base-to-base NOEs and inclusion of accurate and sufficient residual dipolar couplings (RDCs). The fourth project involved protein expression of telomerase proteins for structure and functional study. For this, a new cloning system was designed termed diverse combination ligation indepenedent cloning (DC-LIC). DC-LIC combines the advantages of a number of gene cloning and protein expression techniques to simplify and optimize production of recombinant proteins from *E. coli*. DC-LIC was used to build a number of telomerase gene constructs with improved expression and solubility.

The dissertation of Darian Dallas Cash is approved.

Robert Clubb

Albert Courey

Douglas Black

Feng Guo

Juli Feigon, Committee Chair

University of California, Los Angeles

2015

This dissertation is dedicated to my family.

TABLE OF CONTENTS

List of Figures……...vii

List of Tables…….xvii

Acknowledgements……...xviii

Vita……xx

Publications……..xxi

Chapter 1 Introduction……1

References……..14

Chapter 2 A pyrimidine motif triple helix in the *Kluyveromyces lactis* RNA pseudoknot is essential for function in vivo……..20

Chapter 3 Structure and folding of the *Tetrahymena* telomerase RNA pseudoknot……33

References……..60

Chapter 4 Analysis and optimization of NMR structures of A-form RNA helices…….…….66

References……..87

Chapter 5 Diverse combination ligation independent cloning……...91

References…….…….107

Appendix DC-LIC constructs…….……109

LIST OF FIGURES

Chapter 1:

Figure 1.1: Mechanism of telomere 3′ end elongation by telomerase. The single strand 3′ end of the telomere base pairs with the integral RNA template (red) of telomerase (yellow). TERT processively adds new telomeric repeats (green). C-strand synthesis is performed by DNA polymerase (brown) with an RNA primer (blue)…….……..……3

Figure 1.2: Telomerase core domain architecture. Telomerase RNA architecture and known holoenzyme proteins in **(A)** *Tetrahymena*, **(B)** human, and **(C)** *Kluyveromyces lactis*...…5

Figure 1.3: TERT domain architecture. For human, medaka fish (*Oryzias latipes*), and *Tetrahymena*, TERT domains include telomerase essential N-terminus (TEN, green), RNA binding domain (RBD, red), reverse transcriptase (RT, blue), and C-terminal extension (CTE, orange). The beetle (*Tribolium castaneum*) TERT lacks the TEN domain…..…….……7

Figure 1.4: TERT structure and EM maps of human and *Tetrahymena* telomerase. **(A)** Crystal structure of beetle (*Tribolium castaneum*) TERT with domains labeled RBD (blue), RT (purple), CTE (cyan) and DNA:RNA hybrid (red:green) (PDB:3KYL). **(B)** EM structure of bilobal human telomerase dimer (EMDB: 2310). **(C)** EM structures of the 'open' (EMDB: 2311) monomer of human telomerase dimer with proposed region of dsRNA (TER) indicated with oval outline. **(D)** *Tetrahymena* telomerase holoenzyme (EMDB: 5804) with the RBD, RT, and CTE of TERT and the template of TER modeled in. The structures from *Tetrahymena* and human are shown with beetle TERT (PDB ID: 3KYL) in the same orientation for comparison. The arrows point to the putative binding site of TER CR4/5 in human telomerase **C** and the modeled binding site of TER Loop 4 in the *Tetrahymena* telomerase holoenzyme **D**…….……...……9

Chapter 2 (reprint):

Figure 2.1: (A) The t/PK (or core) domain of *Kluyveromyces lactis* telomerase RNA. Conserved sequences (CS) 3 and 4 make up the pseudoknot, which was truncated to the boxed nucleotides for structural studies. (B) Minimal *K. lactis* TER pseudoknot construct (kPKDU) for NMR studies with predicted secondary structure (*1*). Secondary structure elements are colored: stem 1 (red), loop 1 (gold), stem 2 (blue), loop 2 (green). (C) Southern analysis of telomere restriction fragments from *K. lactis* strains harboring WT or ΔU959 TER1 with BclI template mutation to mark telomerase action, shown in D. Genomic DNA samples were digested with EcoRI (-) or EcoRI+ BclI (+) restriction endonucleases, electrophoresed in a 1% agarose gel, blotted, and hybridized first with a BclI-specific telomere probe (top) and then with a WT telomere probe (bottom). (D) Schematic representation showing a telomere containing WT (blue) and BclI (green) repeats……22

Figure 2.2: (A) (Top Panel) AU and GC (inset) J*NN*-HNN-COSY spectra aligned with (Bottom Panel) Imino proton region of 2D H_2O NOESY at 278 K and 800 MHz. NOE crosspeaks of stem 1/stem 2 WC pairs are connected above the diagonal in black and loop1-stem 2 Hoogsteen pairs below the diagonal in red. Iminos from the stem junction are connected below the diagonal and colored green. (C) Determined secondary structure of kPKDU (colored as Fig. 1) with bulge nucleotides (U955, C957) in magenta…….22

Figure 2.3: (A) 20 lowest energy structures from NMR solution structure determination of kPKDU superpositioned over all heavy atoms. (B) Schematic representation of the kPKDU tertiary structure. Base pairs are indicated by dotted lines. (C) Junction and (D) Loop 1/stem 2 turn tertiary structures. Stereoviews are presented in Fig. S3…… ……23

Figure 2.4: 1D imino proton spectra at 283K and 500 MHz of RNA constructs of the *K. lactis* pseudoknot and stem 2-loop 2 hairpin in 10 mM Tris-D, pH 6.3 (green and blue) and with added

Mg^{+2} (red) (see Methods). Hairpin (green, bracketed) and pseudoknot sequences are shown below the spectra. (A) kPKDU and S2L2, (B) kPKFL and corresponding S2L2-FL. kPKFL and S2L2-FL have the full-length stem 1. (C) kPKFL(861-3CUU:UCC) and S2L2-FL, and (D) kPKFL(864-866UUU:CCC) and S2L2-FL. Dashed lines on the sequences indicated observed tertiary base pairs. Nucleotide substitutions in (C) and (D) were designed to prevent formation of the first 3 or last 3 triples. Spectra are labeled with the RNA construct and primary tertiary fold, hairpin (HP) or pseudoknot (PK), under the given conditions. ……23

Figure 2.5: (A, B) Two alternative base triple registers, 1 and 2, predicted for the *K. lactis* pseudoknot. Register 1 (A) is observed in the solution structure. (C, D) Compensatory mutations designed to affix (C) register 1 and (D) register 2 *in vivo*. Mutations were made in the top strand (S2A), bottom strand (S2B) and loop (L1, L1'). (E) Southern analysis of telomere restriction fragments prepared from WT, ΔTER1, and PK mutants (shown in red in A, C, D) (see Methods and Fig. S1)……24

Figure 2.6: (A) Minimal *S. cerevisiae* pseudoknot with predicted secondary structure and tertiary interactions. Interactions are validated by assays (blue), sequence conservation (green), or unconfirmed (red). (B) 1D imino spectra of *S. cerevisiae* pseudoknot with protonated C^{+} iminos (NH) indicated by arrows, recorded at 283K on 600MHz NMR. Sample buffer is 10mM Na phosphate, pH 6.3, 50mM KCl…….25

Figure 2.7: Solution structures of minimal TER pseudoknots from (A) *K. lactis* (kPK) and (B) *H. sapiens* (hPK), colored as in Figure 1. Tertiary structure schematics (left) and lowest energy structures (right) are shown. (C) Superposition of the backbones of the kPK (red) and hPK (green) tertiary structures, with bulges as stick/ball for kPK (black) and hPK (magenta)….25

Figure 2.S1: A BclI silent mutation is used to mark the nascent products of an investigated

telomerase. (A) A schematic representation showing a telomere containing WT (blue) and BclI (green) repeats. Telomeric restriction fragments are shown for WT-BclI (B) and telomerase null (C) strains, digested with either EcoRI or EcoRI+BclI. BclI-specific and WT hybridization probes for Southern analysis are shown as arrows above the fragments……..…28

Figure 2.S2: (A) kPKDU and stem2-loop 2 hairpin constructs studied by NMR. (B) 1D imino proton spectra of kPKDU in 10 mM Tris-D pH 6.3 titrated with $MgCl_2$ or KCl to the concentrations indicated. (C) 1D imino proton spectrum of the stem2-loop 2 hairpin in 10 mM Tris-D pH 6.3. Spectra were taken at 283K on a 600MHz NMR spectrometer……29

Figure 2.S3: Stereoviews of (A) Lowest 20 energy structures, (B) Junction and (C) Loop1/stem 2 turn tertiary structures…….29

Figure 2.S4: ^{1}H-^{13}C HSQC spectra of kPKDU with Mn^{2+} (red) and without Mn^{2+} (black). Resonances of residues close to bound Mn disappear due to paramagnetic relaxation. (A) C6H6/C8H8 HSQC overlay (red = 100 μM Mn^{2+}). (B) C1′H1′ HSQC overlay (red = 40 μM Mn^{2+}). (C) Atoms which disappear due to broadening are indicated by red (C6H6/C8H8) or gold (C1′H1′) spheres on the pseudoknot tertiary structure. Spectra were taken at 298K on an 800MHz NMR spectrometer…….30

Figure 2.S5: (A) H2O NOESY of imino region of kPKWT. Residues colored in red have chemical shift changes > 0.1ppm. Secondary structure is maintained compared to kPKDU as seen by similar crosspeak patterns and NOE connectivities. (B) kPKWT secondary structure schematic with residues circled that have chemical shift changes > 0.1ppm. Spectra were taken at 283K on an 800MHz NMR spectrometer…….30

Figure 2.S6: Colony morphology of WT, ΔTER, and mutant PK strains of *K. lactis*. Phenotypes categorized as normal (blue), mild (purple), or rough (red)……..…….…...31

Figure 2.S7: (A) Sequence alignment of the pseudoknot domain of *Kluyveromyces marxianus* species telomerase RNA. (B) Secondary structure prediction of pseudoknots based on kPKDU structure…….31

Figure 2.S8: Sequence and secondary structure of the *S. cerevisiae* telomerase RNA pseudoknot based on reference 1. The three U-A-U triples predicted by reference 2 are shown by black dashed lines, and additional predicted U-A-U triples are shown by red dashed line.
1. Liu F, Kim Y, Cruickshank C, Theimer CA (2012) Thermodynamic characterization of the Saccharomyces cerevisiae telomerase RNA pseudoknot domain in vitro. RNA 18(5):973–991.
2. Qiao F, Cech TR (2008) Triple-helix structure in telomerase RNA contributes to catalysis. Nat Struct Mol Biol 15(6):634–640……..32

Chapter 3:

Figure 3.1: *Tetrahymena thermophila* telomerase RNA. **(A)** *Tetrahymena* telomerase RNA, with base paired helices labeled I-IV. Template (magenta), template boundary element (TBE), template recognition element (TRE, cyan), and pseudoknot (PK) are also labeled. **(B)** *Tetrahymena* TER pseudoknot NMR construct. The secondary elements are colored as Stem A (red), Loop A (gold), Stem B (blue), and Loop B (green). Two G's (gray) were added to enhance *in vitro* transcription……..….36

Figure 3.2: Structure of tetPK. **(A)** Diagram of determined secondary structure of tetPK. Solid lines represent backbone connectivity, and dotted lines represent hydrogen bonds. **(B)** Family of lowest 10 energy NMR structures of tetPK. **(C)** Lowest energy NMR structure of tetPK. **(D)** G84-C72-A91 minor groove base triple and **(E)** A80-G95-C75 major groove base triple, with

hydrogen bonds shown as dotted lines. Pseudoknot elements are colored as in Figure 1, except A80 is black and G95 is cyan, for clarification when viewing structures.....…….43

Figure 3.3: Telomerase activity and 1D imino spectra of pseudoknot variants. **(A)** TER substitutions with associated stem-loop base triple configuration. **(B)** Telomerase primer extension assay of WT and mutant TER constructs. The number of telomeric repeats and recovery control (RC) are indicated on the right. **(C)** Relative telomerase activity of WT and mutant TER constructs at pH 7 (dark gray), pH 8 (light gray), and with p65 at pH 7 (white). From the assay image (in **B**), the density of each lane was integrated and normalized to WT pH7 (as 100%). **(D)** (Top Panel) 1D imino spectra of WT tetPK at varying pH. (Bottom Panel) 1D imino spectra of Stem IIIb hairpin and tetPK variants. Imino resonances of Stem IIIa residues are shown in red. Stem IIIa forms in all constructs except Stem IIIb and C75A……..…….47

Figure 3.4: Folding of the TER Core domain. **(A)** Overlay of 2D imino NOESY spectra from Core domain (green), Stem I (red), and Stem II (blue). Template-TRE hairpin resonances are indicated by cyan box. Imino assignment and connectivity is diagrammed for Stem I elongation (black). **(B)** Stem I, Stem II, and Core domain NMR constructs. Red nucleotides indicate changes from WT sequence. **(C)** Overlay of 2D imino NOESY spectra from tetPK (black) and Template-TRE-PK (T-T-PK, green) with Template-TRE hairpin resonances indicated by cyan box. **(D)** Diagram of T-T-PK structure. **(D)** Sequence of Template-TRE (magenta-cyan) NMR construct and corresponding 1D imino spectra. **(E)** Secondary structure of the Core domain in absence of TERT. The potential Stem IIIa hairpin is gray.……..49

Figure 3.5: Secondary structure models of *Tetrahymena* TER. **(A)** TER model based on SHAPE data. **(B)** TER model based on NMR analysis. The potential Stem IIIa hairpin is shown with gray lines based on the evidence of transient structures forming in the pseudoknot region. In addition,

footprinting analysis indicated that Stem IIIa and IIIb are in equilibrium with each other, with Stem IIIa being the major species. Coloring scheme is the same as Figure 1……..…....51

Figure 3.6: *Tetrahymena* TER pseudoknot formation is dependent on closed vs. open TER circle. **(A)** TER NMR constructs with different combinations of TER elements, including PK-StemI-StemII (PK-SI-SII), Template-TRE-PK (T-T-P-K), StemI-Template-TRE-PK (SI-T-T-P-K), StemI-PK (SI-PK), StemI (SI), and Core Domain. Whether each construct forms a pseudoknot (check mark) or not (X mark) is indicated. **(B)** 1D imino spectra of TER NMR constructs. A80 aminos are indicated by green arrows (or absence "X")…….....52

Figure 3.7: Comparison of ciliate, human, and yeast TER pseudoknots. **(A)** Diagram of TER pseudoknot secondary structure from *K. lactis*, human and *Tetrahymena*. **(B)** NMR solution structures of TER pseudoknots from human (red) and *K. lactis* (backbone only, blue) which are overlayed. TetPK is (green). Human and tetPK structures are aligned according to their similar base triples, with the corresponding U-A-U triple in each colored black…….....56

Fig. 3.S1: TetPK is unstable at higher temperature and pH. 2D (H5-H6) TOCSY spectra of tetPK at varying temperature and pH. The increasing number and intensity of TOCSY crosspeaks indicates higher population of unfolded pseudoknot..……...58

Figure 3.S2: Evidence for N1 protonation of A80. **(A)** ^{13}C-^{1}H HSQC of C2-H2 region of tetPK. The C2 resonance of A80 is upfield shifted, boxed in green. **(B)** 2D imino NOESY of tetPK. The pseudoknot imino resonances are assigned and labeled on the diagonal in black. The A80 amino resonances are downfield shifted and split, boxed in green. The A80 amino resonances have NOE crosspeaks to various other pseudoknot imino resonances, indicated in red……59

Chapter 4:

Figure 4.1: *Tetrahymena* telomerase RNA. **(A)** Diagram of *Tetrahymena* TER, with helical segments numbered I-IV. Stemloops 2 and 4 (SL2 and SL4) are boxed in red. NMR constructs of **(B)** SL2 and **(C)** SL4 are shown, with corresponding lowest energy NMR structures (from original data). Non-native residues are colored gray.……69

Figure 4.2: Effect of including more distance restraints on SL4 structure. **(A)** The average structure, from the lowest 10 energy, is shown for each restraint dataset. **(B)** Statistics are given for each dataset for the NOE structure and RDC-refined structure……..……75

Figure 4.3: Influence of sequential base-to-base distance on structure. For these structure calculations, the original distance restraint list was used, except the sequential base-to-base distances were either left unchanged (Original data, red) or modified to the correct values **(**4.5Å, black). **(A)** SL2 and **(B)** SL4 average structures (from lowest 10 energy structures) are shown with corresponding major groove width (D_{p-p})……...…….……..…78

.

Figure 4.4: Fitting of NOE-based structures to RDCs. For **(A)** SL2 and **(B)** SL4, back calculated RDCs for NOE-based structures are plotted against experimental RDCs. The R^2 for RDC fitting, and Q values are indicated……...……..79

Figure 4.5: Non-redundant bond types are essential for RDC refinement. NOE structures were calculated using the **(A)** Original data, red, **(B)** Dataset 2, black, or **(C)** Dataset 5, gray. The NOE structures (middle) were then refined using the entire RDC data (left) or base RDC data only (right). Resulting major groove widths are indicated (D_{p-p})……...81

Figure 4.6: Fitting A-form helices to bulged RNA structures. Secondary structure diagrams of **(A)** classical swine fever virus internal ribosome entry site (CSFV IRES) RNA and **(B)**

Tetrahymena TER Stem 4. The Watson-Crick stems (red box) of the NMR structures of the RNAs (red) are fit to A-form helices (green)…….84

Figure 4.7: *Tetrahymena* telomerase holoenzyme. **(A)** (Top) Representative class averages of negative staining electron microscopy and cryoelectron microscopy images of TERT–Flag (F) telomerase. (Bottom) 3D reconstruction of Teb1–F telomerase (front and side views) and class averages of affinity-labelled telomerase particles. Lines with circle heads indicate attachment point of Fab (red arrows) and MS2cp (white arrow). **(B)** Telomerase subunit schematic (front view). **(C)** 3D reconstruction of Teb1–F telomerase with TERT (blue), p65 (green) and TER (black), plus Teb1C (orange) modelled into the electron microscopy density. **(D)** TER model structure (well determined = magenta; remaining = black) and interactions with TERT TRBD (blue) and TEN (cyan), and p65 La, RRM1, and xRRM2 domains (green). **(E)** Secondary structure schematic of TER with TRBD……86

Chapter 5:

Figure 5.1: LIC cloning schematic. **(A)** Components of PCR, including forward (fwd) and reverse (rev) primers, and template (target gene). **(B)** PCR product with added LIC sequences. The 3′ end of the gene may encode a stop codon or readthrough sequence (i.e. TEV site) to add C-terminal tag. **(C)** T4 DNA polymerase treatment with corresponding dNTP (green) creates matching overhangs in the insert and vector. **(D)** Insert and vector are annealed to create the final recombinant plasmid for *E. coli* expression……..93

Figure 5.2: Strategy for C-GFP protein expression optimization. **(A)** Empty vector with LIC sequence displayed, where after cleavage results in 5′ (red) and 3′ (blue) halves. The cloning site is designed so that the C-terminal GFP is frame-shifted with a termination codon, so that empty vector does not express GFP. C-GFP is colored based on its expected expression/fluorescence levels, where gray indicates no expression/fluorescence and increasing green color indicates

higher expression/fluorescence. **(B)** Recombinant vector with cloned insert. The cloning site is designed so that only correctly cloned genes will place C-GFP in the proper frame to express (green). **(C)** The target gene can be modified, randomly or systematically, and cloned into C-GFP vectors. The amount of GFP expression serves as an indicator for protein expression levels of each variant. **(D)** Example results for LB-agar plate (top) or liquid culture (bottom) C-GFP expression tests. ……98

Figure 5.3: Vector for RNA expression in *E. coli.* **(A)** The tRNA scaffold assists RNA stability in the cell. The RNA of interest is cloned into the anticodon stem of the tRNA. **(B)** Two cloning sites with different restriction enzymes can be used to exclude (R1/R1*) or include (R2/R2*) the optional streptavidin aptamer when building RNA expression constructs. **(C)** The streptavidin RNA aptamer is a 42nt bulged stemloop that has high affinity for streptavidin (S), allowing for affinity purification……100

Figure 5.4: DC-LIC cloning system applied to telomerase proteins. **(A)** Amino acid sequence and corresponding secondary structure (E = beta sheet, H = alpha helix) for *Tetrahymena thermophila* (tt) TERT (C-terminus). Different TERT constructs were made according to the secondary structure, ending at the indicated residues (green highlight), and tested for expression using C-GFP. ttTERT (1-1015) was the largest expressing construct. SDS-PAGE gels show expression and purification of **(B)** *Tetrahymena* TERT (1-1015), His6-MBP-(C-GFP), **(C)** Medaka fish TERT (full length), MBP-(C-His6), and **(D)** Human TERT (319-1132), His6-MBP-(C-GFP). The lanes are labeled for the molecular weight marker (M), and purification step as His column (H), Amylose resin (A) or Ion exchange (IEX). The target protein location on the gel is indicated by (*)……..……102

LIST OF TABLES

Chapter 2:

Table 2.S1: Restraint and structure statistics for kPKDU. *Two per hydrogen bond. ** Pairwise RMSD was calculated for the lowest 10 energy structures……..……32

Chapter 3:

Table 3.1: Restraint and structure statistics for tetPK. *Two hydrogen bond restraints were used for each hydrogen bond. ** Pairwise RMSD was calculated for the lowest 10 energy structures……44

Chapter 4:

Table 4.1: NMR restraints. Ideal values used for dihedral angles and inter-proton distances are listed. Distance types were categorized (starting with commonly used data) for use in structure calculations. For the mock calculations, each dataset includes all of the data from the previous sets……..73

Table 4.2: Helical parameters determined using mock calculations of SL4. For the NOE and RDC refined structures, the helical parameters shift, slide, rise, tilt, roll and twist are indicated for the given dataset with average and standard deviation compared to those of an A-form RNA helix……..…….76

Chapter5:

Table 5.1: DC-LIC vectors. Table of DC-LIC vectors with different combinations of N-terminal and C-terminal fusion tags. N/a indicates vectors that were not made……..……96

ACKNOWLEDGEMENTS

I would like to thank my advisor Juli Feigon for continuing support and guidance throughout my graduate career. Thanks for helping me become a better researcher and scientific writer. Thanks for motivating me and becoming the equivalent of my west coast mother. I would also like to thank the members of the Feigon Lab who have thoroughly trained me and engaged me in thoughtful scientific and other entertaining discussions. Thanks to Dr. Nak-Kyoon Kim and Dr. Qi Zhang for training me in NMR methodology for nucleic acids. Thanks to Dr. Jing Zhou for getting me started with my first project. Thanks to Dr. Robert Peterson, Dr. Mariusz Kamionka, Dr. Mahavir Singh, Dr. Edward Miracco, Dr. Jiansen Jiang, Dr. Mijeong Kang, Dr. Elon Hartman, Dr. Yaqiang Wang, Dr. Catherine Eichhorn, Dr. Jing Xue, and Master Henry Chan.

I would like to thank Dr. Yehuda Tzfati, Dr. Nikolai Ulyanov, Dr. Kinneret Shefer, and Osnat Cohen-Zontag for their collaboration with the work published in Chapter 2.

I would like to acknowledge the Cellular and Molecular Biology Training Grant program at UCLA, where I was supported by NIH grant GM007185.

I would like to thank the University of Maryland, Baltimore County Meyerhoff Scholarship Program for training me and inspiring me to pursue a graduate degree. In particular Dr. Freeman Hrabowski, Dr. Michael Summers, and Lamont Toliver who were instrumental in my success.

I would like to thank my parents for bringing me into this world and providing me with the inspiration and tools to achieve my goals; and for nurturing my interests in science by sending me to various summer programs. Thanks to my brothers and sister, Earl Cash, Chris Cash, and Kim Cash Tate for their continuing support and friendship. Thanks to my brother in law Willam Tate for motivating me to finish my PhD degree. And thanks to Chica Yaegashi for providing me support and inspiration.

The text of Chapter 2 is a reprint from Cash, D. D., Cohen-Zontag, O., Kim, N. K., Shefer, K., Brown, Y., Ulyanov, N. B., Tzfati, Y., and Feigon, J. (2013) Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo, *Proc Natl Acad Sci U S A 110*, 10970-10975 with permission from National Academy of Sciences.

Chapter 4 includes information from work where I was co-author. Jiang, J., Miracco, E. J., Hong, K., Eckert, B., Chan, H., Cash, D. D., Min, B., Zhou, Z. H., Collins, K., and Feigon, J. (2013) The architecture of *Tetrahymena* telomerase holoenzyme, *Nature 496*, 187-192 with permission from my co-authors.

VITA

Darian Dallas Cash

Education:

M.S. Biochemistry, University of California, Los Angeles (UCLA)	Mar 2012
B.S. Chemistry, University of Maryland, Baltimore County (UMBC)	Dec 2002

Research and Work Experience:

Graduate Researcher, UCLA	Sep 2009 – Dec 2014
QC Chemist, United Pharma LLC, Irvine, CA	Dec 2008 – Aug 2009
QA/QC Chemist, Baxter Healthcare, Los Angeles, CA	Oct 2007 – Oct 2008
Research Associate, Xencor Inc., Monrovia, CA	Mar 2005 – Jul 2007
Graduate Researcher, UCLA	Sep 2002 – Jan 2005
Undergraduate Researcher, Howard Hughes Medical Institute, UMBC	Oct 1999 – Sep 2002

Honors and Awards:

Cellular and Molecular Biology Training Grant, UCLA	Sep 2010 – Sep 2013
Meyerhoff Scholarship, UMBC	Jun 1998 – Sep 2002

PUBLICATIONS

1. Cash, D. D., Cohen-Zontag, O., Kim, N. K., Shefer, K., Brown, Y., Ulyanov, N. B., Tzfati, Y., and Feigon, J. Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo, *Proc Natl Acad Sci U S A* ***110***, 10970-10975. (2013)

2. Jiang, J., Miracco, E. J., Hong, K., Eckert, B., Chan, H., Cash, D. D., Min, B., Zhou, Z. H., Collins, K., and Feigon, J. The architecture of Tetrahymena telomerase holoenzyme, *Nature* ***496***, 187-192. (2013)

3. Miracco, E. J., Jiang, J., Cash, D. D., and Feigon, J. Progress in structural studies of telomerase, *Curr Opin Struct Biol* ***24***, 115-124. (2014)

CHAPTER 1

Introduction

Telomeres and Telomerase

Telomeres are nucleoprotein complexes at the ends of eukaryotic linear chromosomes (*2*). Telomeric DNA consists of a repetitive short (usually 6-8 nt) non-coding sequence, that is double stranded with a 3′ single strand overhang (*3*). The telomeric length and sequence of the repeat vary between species. Telomere length spans <1000 nt in yeast and ciliates, 10-15 kB in humans, and 25-50 kB in mice (*4*). The single strand overhang is typically 150-200 nt long and is G-rich (G-strand), with a TTAGGG repeat in vertebrates. The complementary strand is termed the C-strand. Telomeres help distinguish chromosome ends from double-stranded breaks, protecting them from DNA repair machinery, degradation and fusion (*4, 5*). They are essential for chromosomal stability. However despite their importance, telomeres are progressively shortened due to the "end replication problem." This refers to the problem that arises in conventional DNA replication where the 3′ end cannot be fully replicated due to the necessity of a 3′ hydroxyl (OH) for polymerase catalysis. The 3′-OH is supplied by RNA primers for 3′ end replication, which are subsequently removed. Since DNA inevitably shortens with each round of cell division, telomeres can be seen as an expendable first defense against DNA loss, protecting the more important central genetic material. Telomere shortening past a critical length results in chromosomal damage and cellular senescence (*6, 7*).

Telomerase is the ribonucleoprotein (RNP) complex that maintains telomere length. Telomerase is composed of the specialized telomerase reverse transcriptase (TERT), telomerase RNA (TER), and other species specific proteins (*8, 9*). TERT and TER alone are sufficient for activity *in vitro,* while the other proteins are necessary for function *in vivo* (*2, 10, 11*). After telomerase elongates the telomeric G-strand, the C-strand is synthesized by DNA polymerase (Fig. 1) (*12*). While telomerase seems to be an essential factor in cell proliferation, its activity is undetectable in most somatic cells (*13*). As such, telomeres shorten with each cell division, creating a correlation between telomere length and aging. On the other hand, telomerase is expressed in cells that need to divide regularly, such as epithelial, haemopoietic, germ line, and

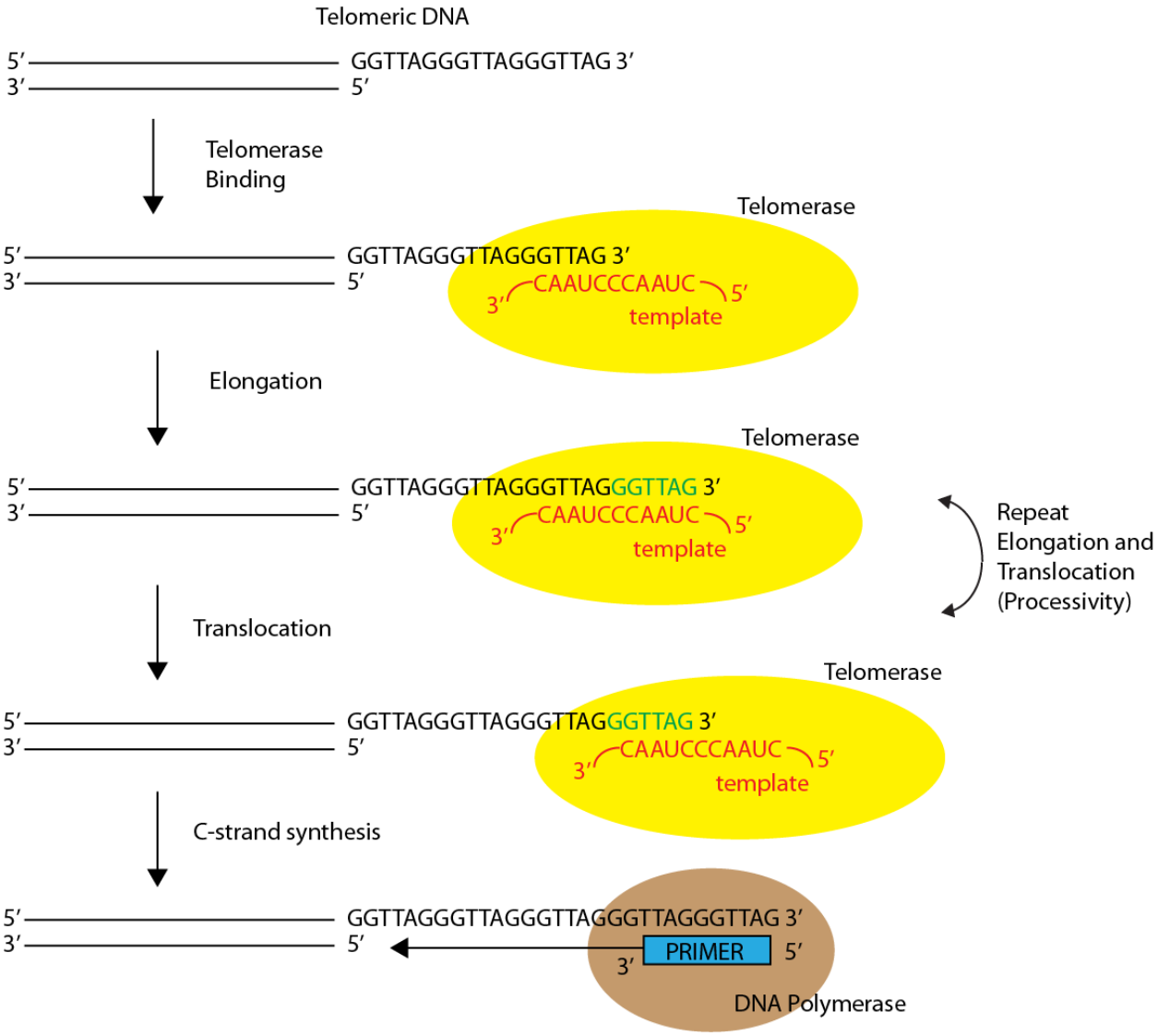


Figure 1: Mechanism of telomere 3′ end elongation by telomerase

The single strand 3′ end of the telomere base pairs with the integral RNA template (red) of telomerase (yellow). TERT processively adds new telomeric repeats (green). C-strand synthesis is performed by DNA polymerase (brown) with an RNA primer (blue)

stem cells. Telomerase allows these cells to divide indefinitely, essentially becoming "immortal" (*14*). This immortality is also what links telomerase to cancer, as ~90% of all human cancer cell lines have highly active telomerase. Thus, telomerase is a potential target for cancer therapy and an area of intense study.

Telomerase and Disease

While cancer is related to increased telomerase activity and extended telomeres, there are also a number of inherited diseases marked by short telomeres, such as dyskeratosis congenita (DKC), aplastic anemia (AA), and idiopathic pulmonary fibrosis (IPF) (*15-17*). Dyskeratosis means irreversible degeneration of skin cells and congenita means "born with." Originally thought to be a skin disease, DKC is characterized by abnormal skin, nail dystrophy, and oral leukoplakia (white patches) (*18*). DKC patients have a predisposition for cancer and an increased risk for developing other life-threatening conditions. A majority of the population of DKC patients has mutations in one of six telomerase related genes, including TERT, TER, and dyskerin (a human telomerase holoenzyme protein) (Fig. 2) (*19*). These mutations generally result in telomerase dysfunction and lower levels of telomerase (*20*). DKC patients may also develop AA, which is a bone marrow disorder. AA occurs when the bone marrow does not make enough new blood cells. Mutations in TERT, TER, and related co-factors have been linked to AA, explaining the apparent short telomeres in AA patients (*21*). Treatments for these telomere biology disorders are aimed at restoring telomere length. Implanted stem cells have been shown to increase TER expression and telomerase activity in DKC patients, providing a basis for therapies against the wide array of telomere maintenance diseases (*22*).

TER and TERT Domains

TER varies in size considerably between species, ranging from 159 nt in ciliates, 451 nt in human, to over 2000 nt in some yeast (Fig. 2) (*23-25*). In addition to the template which is used to copy the telomeric repeat, TER has a number of motifs necessary for function, including

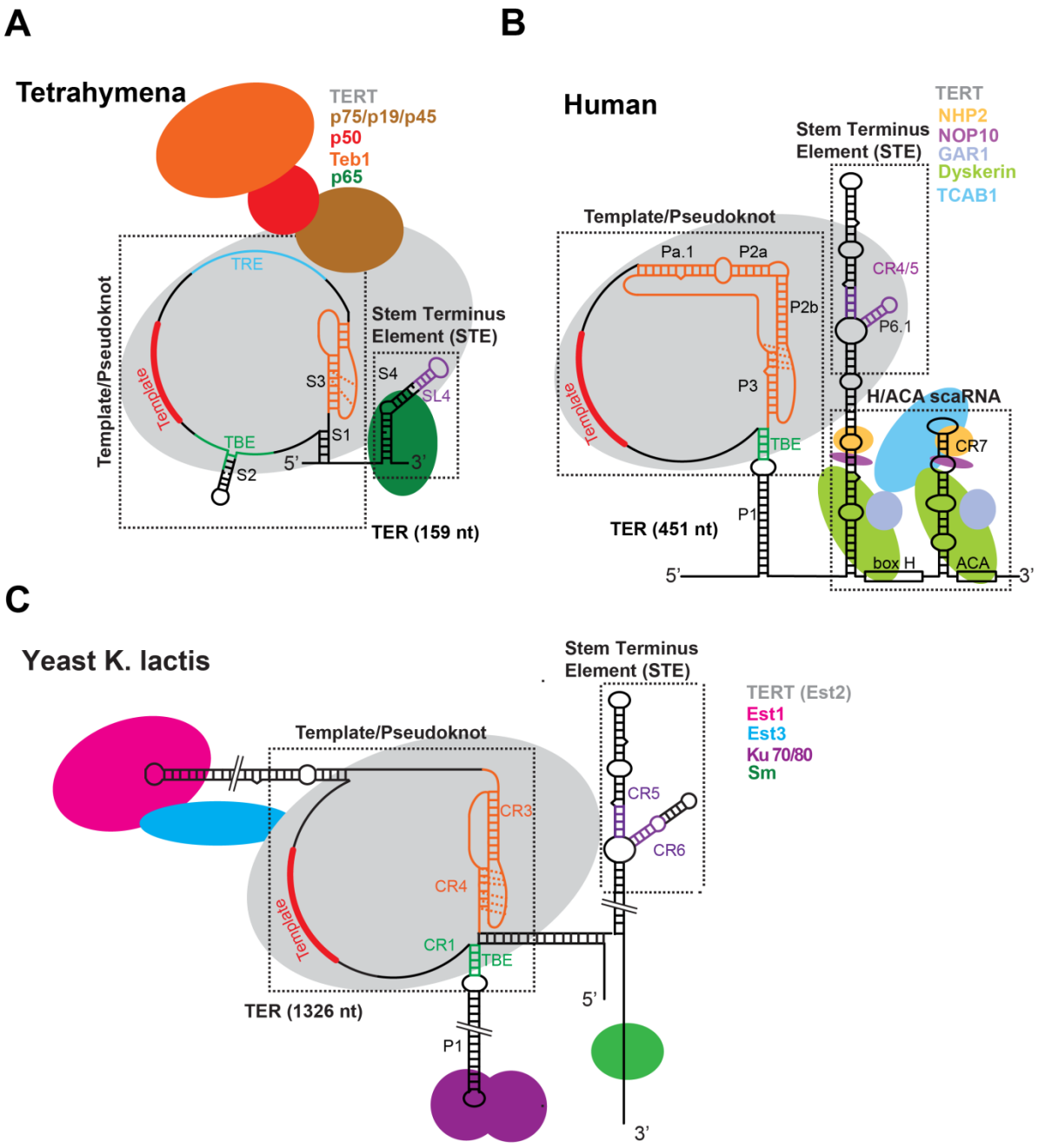


Figure 2: Telomerase core domain architecture

Telomerase RNA architecture and known holoenzyme proteins in **(A)** *Tetrahymena*, **(B)** human, and **(C)** *Kluyveromyces lactis*.

the pseudoknot, template boundary element (TBE) and stem-terminus element (STE) (Fig. 2) (*26, 27*). The pseudoknot is template-adjacent and is important for TER folding in addition to a number of other potential roles, such as template positioning within the active site and caatalysis (*28, 29, 69*). The TBE is typically a stem or hairpin, immediately upstream of the template, which prevents copying beyond the template sequence (*30, 31*). The STE can either be a terminal hairpin, three-way junction, or combination of the two. The STE binds TERT and stimulates catalytic activity (*32*). The TER core domain, or template/pseudoknot (t/PK) domain, is an enclosed region of the RNA consisting of the template, pseudoknot and TBE (Fig. 2). The core domain is the most conserved part of TER, while the remaining elements vary between species.

Ciliates, which have a relatively small TER, only have an additional stem-loop 4 (SL4) outside of the core domain, which serves as the STE equivalent. SL4 contains a GA bulge that binds accessory protein p65, causing a large conformational change in TER that promotes holoenzyme assembly (Fig. 2A) (*33-35*). The vertebrate STE is a three-way junction, termed conserved regions (CR) 4/5, that is essential for function in conjunction with the core domain (*36, 37*). Vertebrates also contain a 3′ H-box/ACA (H/ACA) domain found in H/ACA scaRNAs consisting of two hairpins and the conserved H/ACA sequence that binds the proteins dyskerin, Nop10, Nhp2, Gar1 (Fig. 2B) (*38, 39*). While most H/ACA RNPs function in site-directed pseudouridylation, the telomerase H/ACA motif is necessary for TER accumulation and biogenesis *in vivo* (*40, 41*). Yeast have large TERs, with three long stem structures ("arms") extending from the core domain that serve as protein scaffolds (*24, 42, 43*). The three arms of yeast TER bind Est1/Est3 (ever shorter telomere) proteins which regulate telomerase activity (*44*), the Ku heterodimer which enhances RNP accumulation and recruitment to telomeres (*45, 46*), and Sm proteins that are important for TER maturation and stability (Fig. 2C) (*47*). The 3′ Sm binding arm of yeast TER also contains the three-way junction STE.

In contrast to TER, TERT varies considerably less between species. TERT is composed of the telomerase essential N-terminus (TEN) domain, RNA binding domain (RBD), reverse

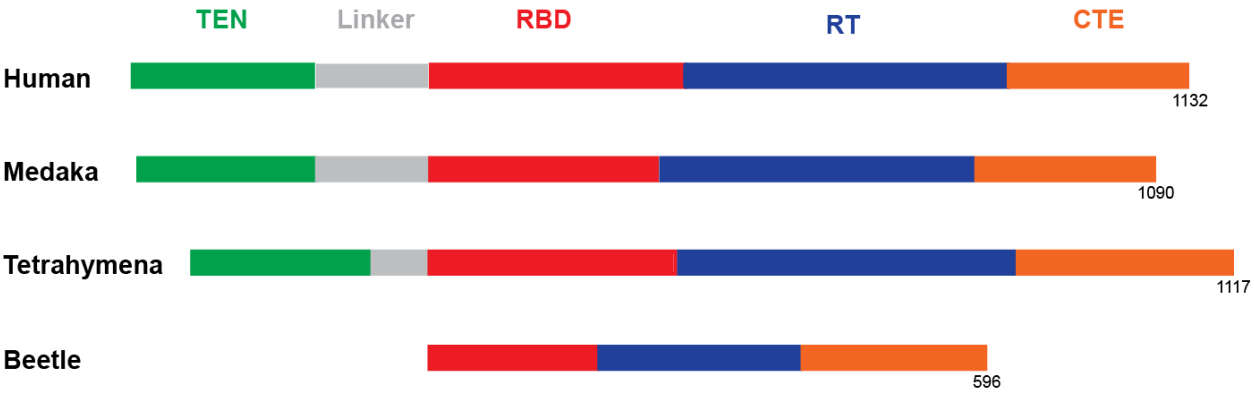


Figure 3. TERT domain architecture

For human, medaka fish (*Oryzias latipes*), and *Tetrahymena*, TERT domains include telomerase essential N-terminus (TEN, green), RNA binding domain (RBD, red), reverse transcriptase (RT, blue), and C-terminal extension (CTE, orange). The beetle (*Tribolium castaneum*) TERT lacks the TEN domain.

transcriptase (RT), and C-terminal extension (CTE) (Fig. 3). The TEN domain interacts with single-stranded DNA as well as TER, linking the enzyme and substrate to provide high repeat addition processivity (RAP) (*48, 49*). The RBD has specific interactions with TER, binding the STE in vertebrates and stem-loops 2 and 4 in ciliates (*50, 51*). The RBD is essential for enzyme assembly and activity (*52*). The RT domain which forms the active site is similar to retroviral RTs, containing the "palm" and "finger" subdomains (*53, 54*). The CTE adopts a novel protein fold and is the putative "thumb" domain of telomerase (*53*). The CTE has species specific roles *in vivo*, affecting RAP and subcellular localization in humans (*55*).

Telomerase Structure

The crystal structure of beetle (*Tribolium castaneum)* TERT was a significant advancement in the telomerase field (Fig. 4A) (*53*). While beetle TERT lacks the TEN domain, the structure showed that TERT forms a ring with the RBD contacting the CTE. Substrate binding and catalytic motifs are on the interior surface of the ring. Subsequently, the beetle TERT structure was solved in the presence of an RNA-DNA hairpin, mimicking the template/telomeric primer duplex (*56*). The structures reveal a strong similarity between TERT and HIV RT domain organization and nucleic acid binding, suggesting a common mechanism. The palm, finger, and thumb domains mediate specific nucleic acid interactions, causing conformational changes in TERT to precisely position the DNA 3′ end at the active site for catalysis (*57*). The DNA-RNA substrate is located at the center of the TERT ring (Fig. 4A).

Recent EM structures of *Tetrahymena* and human telomerase holoenzymes have provided insight into telomerase architecture (Fig. 4B, C, D) (*58, 59*). Endogenously assembled *Tetrahymena* telomerase was studied using affinity labels to localize each component within the structure. This localization revealed the network of interactions necessary for telomerase assembly and function and allowed for fitting of high-resolution structures (and models when necessary) into the EM map. This modeling provided a view of the organization of the catalytic core. Furthermore, the structure is consistent with biochemical data that indicates the accessory

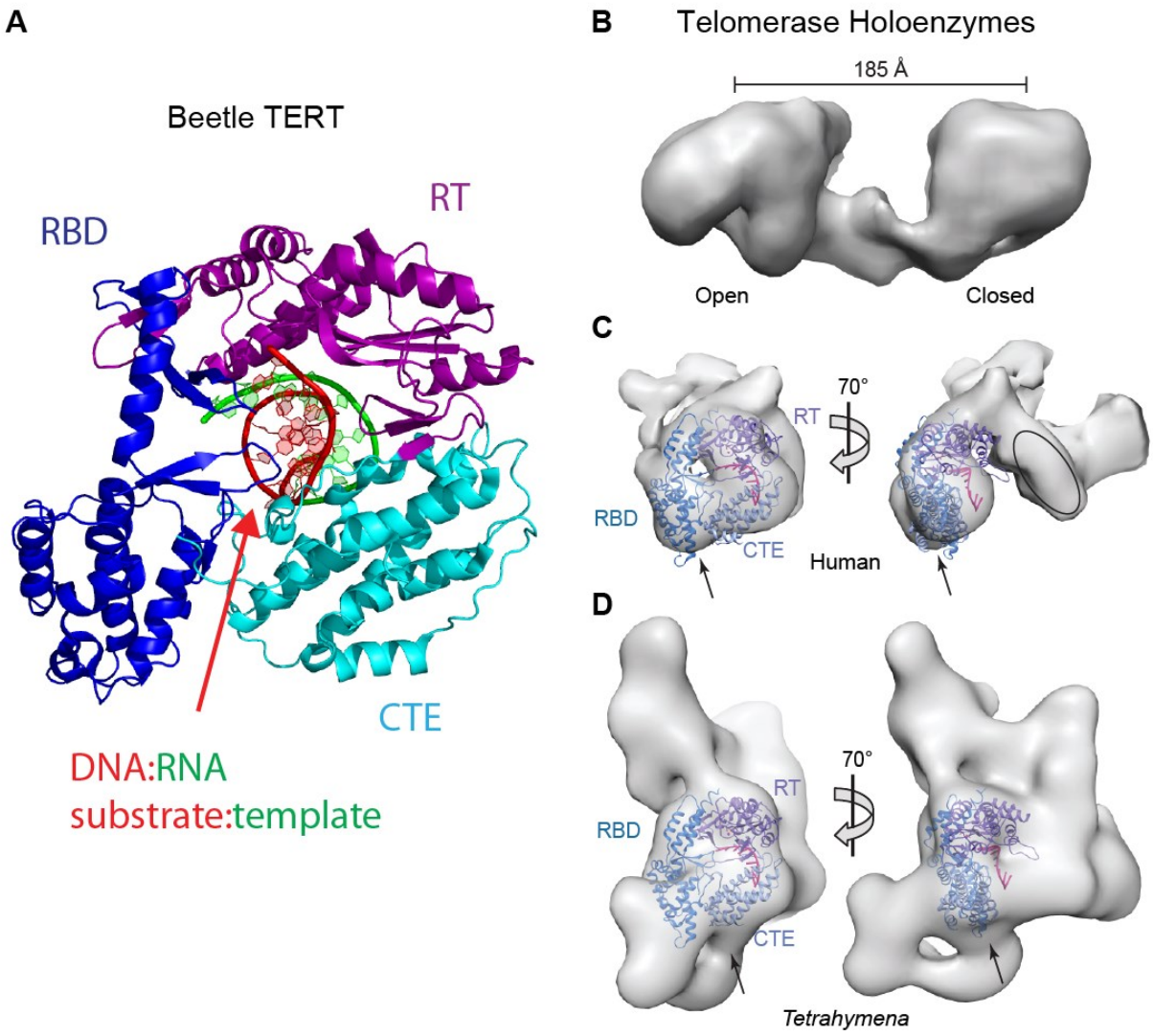


Figure 4. TERT structure and EM maps of human and *Tetrahymena* telomerase

(A) Crystal structure of beetle (*Tribolium castaneum*) TERT with domains labeled RBD (blue), RT (purple), CTE (cyan) and DNA:RNA hybrid (red:green) (PDB:3KYL). **(B)** EM structure of bilobal human telomerase dimer (EMDB: 2310). **(C)** EM structures of the 'open' (EMDB: 2311) monomer of human telomerase dimer with proposed region of dsRNA (TER) indicated with oval outline. **(D)** *Tetrahymena* telomerase holoenzyme (EMDB: 5804) with the RBD, RT, and CTE of TERT and the template of TER modeled in. The structures from *Tetrahymena* and human are shown with beetle TERT (PDB ID: 3KYL) in the same orientation for comparison. The arrows point to the putative binding site of TER CR4/5 in human telomerase **C** and the modeled binding site of TER Loop 4 in the *Tetrahymena* telomerase holoenzyme **D**.

protein p50 is a central hub between the catalytic core, DNA-binding Teb 1, and the p75-p19-p45 (7-1-4) subcomplex.

A human telomerase EM structure was determined from "super-telomerase," purified from kidney HEK293T cancer cells transiently transfected with hTERT and hTER plasmids (*59*). The structure reveals human telomerase as a dimer with a flexible hinge interface. The components in the complex determined by mass spectrometry are TERT, TER, dyskerin and Nop10. The individual monomers were proposed to have two possible conformations in regards to the TERT ring, either closed or open with a V-shaped partition. Nanogold was used to label the location of the primer substrate and TERT-His tag. The crystal structure of beetle TERT was fit into the open monomer, providing insights into TER location, template position, and domain organization of the telomerase dimer. Further experiments determined that the two catalytic sites, which are ~190Å apart, must cooperate for catalysis, since mutations in one subunit abolished activity. The authors hypothesize that the dimeric telomerase functions to extend aligned sister telomeres in parallel to maintain equal lengths after extension (*59*).

In addition to EM studies, telomerase protein and RNA structure has been extensively studied through a modular approach, i.e. determination of structures of isolated domains. The *Tetrahymena* TEN domain crystal structure was solved, revealing a novel protein fold with the ability to bind both RNA and telomeric DNA (*60*). There are four crystal structures of the RBD of telomerase including the beetle RBD as part of TERT, *Tetrahymena* RBD, *Takifugu* fish RBD and medaka fish RBD complex with CR4/5 (*51, 53, 61, 62*). The TERT RBD structure is well conserved, except that vertebrates have an additional N-terminal linker termed "TFLY" (for the corresponding amino acids) that binds the TBE of TER. The medaka fish RBD-CR4/5 complex shows highly specific sequence recognition and conformational changes when compared to the free CR4/5 NMR structure, suggesting a role in telomerase regulation similar to the *Tetrahymena* SL4 (*51, 63*).

Many TER stem and hairpin structures have been solved, including human P6, P6.1, P8, P2b, and ciliate SL2 and SL4 (*64-70*). The first structure of a TER pseudoknot was the minimal

human pseudoknot reported in 2005 (*71*). The pseudoknot contains a stabilizing triple helix with three major groove triples, two minor groove triples, and loop-loop base pair inserted between the two stems. Through mutational analysis, these tertiary interactions of the pseudoknot were shown to be essential for telomerase activity. This was the first observation of an RNA pyrimidine motif triple helix with biological function. Furthermore, the structure explains phylogenetic sequence conservation to maintain the network of tertiary interactions, as well as mutations linked to the diseases DKC and AA (*17*, *71*) .

Thesis Synopsis

The human TER pseudoknot structure and functional studies were a significant influence for the majority of this thesis work. The first project, covered in Chapter 2, had two general aims. First, determine if the pseudoknot triple helix is conserved in yeast and what structural features are most important. Second, provide further correlation between telomerase activity and pseudoknot tertiary structure. This study was designed to gain further insight into the role of the pseudoknot in telomerase activity. For this work, I determined the NMR solution structure of the yeast *Kluyveromyces lactis* TER pseudoknot, which revealed a number of similarities and differences to that of human. The yeast pseudoknot contains an unexpectedly long pyrimidine motif triple helix, with three bulge nucleotides between the triples near the junction. The bulges allow for base pairing and continuous stacking interactions through the junction, which is also a feature of the human pseudoknot. Despite sequence differences, the overall tertiary shape is remarkably similar between the yeast and human pseudoknot, implying a conserved function. *In vivo* functional studies of the yeast pseudoknot revealed that the determined structure is physiologically relevant, since the base paired junction and base triple conformation were essential for telomere length maintenance.

The second project (Chapter 3) involves the study of the structure and folding of the ciliate *Tetrahymena* TER pseudoknot. Ciliate telomerase and telomeres have been extensively studied due to their telomere-rich macronucleus (*72*). The recent *Tetrahymena* telomerase EM

structure further signifies the importance of characterizing the *Tetrahymena* telomerase RNA. Therefore, I determined the NMR solution structure of the *Tetrahymena* TER pseudoknot and examined its folding in the context of full length TER. The structure revealed that it contains a novel A-G-C base triple which is important for pseudoknot folding and stability as well as telomerase activity. Previous studies determined that the ciliate pseudoknot does not form in the context of the full length TER without TERT. My work revealed that alternate structures, including a hairpin around the template region and extended Stem I, form in the full length TER core domain and prevent formation of the intrinsically unstable pseudoknot..

Chapter 4 was initiated in response to a 2010 analysis of X-ray and NMR RNA structures, focusing on A-form helical segments (*73*). A survey of deposited structures revealed a discrepancy in structural precision between NMR and crystallography, where crystallography structures were more compact while NMR structures had many outliers with extended helices. Previously, AMBER refinement and residual dipolar couplings (RDCs) were shown to improve the NMR structures. In this thesis work, I used the Xplor-NIH structure calculation method to determine the important NMR restraints for accurately determining RNA helices. The most essential data was distance restraints derived from sequential base-to-base NOEs (between neighboring H5, H6, H8 protons) and application of an extensive set of non-redundant RDCs. This optimization was applied to *Tetrahymena* TER SL2 and SL4 structures, which were re-calculated using an improved set of distance restraints and RDCs. These refined structures were used for fitting into the EM density map of the *Tetrahymena* telomerase holoenzyme (*58*).

Chapter 5 begins an effort to study telomerase protein-RNA interactions by developing a cloning system to simplify and optimize protein expression. Diverse combination ligation independent cloning (DC-LIC) builds on the original LIC method developed in the early 90's, which uses the $3' > 5'$ exonuclease activity of T4 DNA polymerase to create matching overhangs in the insert and linear vector (*74*). The DC-LIC system expands upon this basic technique by making a wide variety of vectors that all include the same LIC site and can thus be simultaneously screened with the same insert (PCR product). Highlights of the DC-LIC system

include the 6x His tag for purification, a variety of fusion tag partners, a TEV cleavage site, simplified co-expression, and GFP expression and solubility reporter. Using the DC-LIC cloning vectors, a number of telomerase protein constructs from human, medaka fish, and *Tetrahymena* were able to be solubly expressed.

References

1. Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N. B., and Tzfati, Y. (2007) A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA, *Mol. Cell. Biol. 27*, 2130-2143.

2. Nandakumar, J., and Cech, T. R. (2013) Finding the end: recruitment of telomerase to telomeres, *Nat Rev Mol Cell Biol 14*, 69-82.

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

4. Blasco, M. A. (2007) Telomere length, stem cells and aging, *Nat. Chem. Biol. 3*, 640-649.

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

6. Zhang, X., Mar, V., Zhou, W., Harrington, L., and Robinson, M. O. (1999) Telomere shortening and apoptosis in telomerase-inhibited human tumor cells, *Genes Dev 13*, 2388-2399.

7. Greider, C. W. (1998) Telomerase activity, cell proliferation, and cancer, *Proc Natl Acad Sci U S A 95*, 90-92.

8. Collins, K. (2006) The biogenesis and regulation of telomerase holoenzymes, *Nat. Rev. Mol. Cell. Biol. 7*, 484-494.

9. Harrington, L. (2003) Biochemical aspects of telomerase function, *Cancer Lett 194*, 139-154.

10. Weinrich, S. L., Pruzan, R., Ma, L., Ouellette, M., Tesmer, V. M., Holt, S. E., Bodnar, A. G., Lichtsteiner, S., Kim, N. W., Trager, J. B., Taylor, R. D., Carlos, R., Andrews, W. H., Wright, W. E., Shay, J. W., Harley, C. B., and Morin, G. B. (1997) Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT, *Nat Genet 17*, 498-502.

11. Licht, J. D., and Collins, K. (1999) Telomerase RNA function in recombinant Tetrahymena telomerase, *Genes Dev 13*, 1116-1125.

12. Price, C. M. (1997) Synthesis of the telomeric C-strand. A review, *Biochemistry (Mosc) 62*, 1216-1223.

13. Giardini, M. A., Segatto, M., da Silva, M. S., Nunes, V. S., and Cano, M. I. (2014) Telomere and telomerase biology, *Prog Mol Biol Transl Sci 125*, 1-40.

14. Maqsood, M. I., Matin, M. M., Bahrami, A. R., and Ghasroldasht, M. M. (2013) Immortality of cell lines: challenges and advantages of establishment, *Cell Biol Int 37*, 1038-1045.

15. Artandi, S. E. (2006) Telomeres, telomerase, and human disease, *N Engl J Med 355*, 1195-1197.

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

17. Fossel, M., and Flanary, B. (2009) Telomerase and human disease: the beginnings of the ends?, *Rejuvenation Res 12*, 333-340.

18. Fernandez Garcia, M. S., and Teruya-Feldstein, J. (2014) The diagnosis and treatment of dyskeratosis congenita: a review, *J Blood Med 5*, 157-167.

19. Machado-Pinilla, R., Carrillo, J., Manguan-Garcia, C., Sastre, L., Mentzer, A., Gu, B. W., Mason, P. J., and Perona, R. Defects in mTR stability and telomerase activity produced by the Dkc1 A353V mutation in dyskeratosis congenita are rescued by a peptide from the dyskerin TruB domain, *Clin Transl Oncol 14*, 755-763.

20. Vulliamy, T. J., Walne, A., Baskaradas, A., Mason, P. J., Marrone, A., and Dokal, I. (2005) Mutations in the reverse transcriptase component of telomerase (TERT) in patients with bone marrow failure, *Blood Cells Mol Dis 34*, 257-263.

21. Savage, S. A., and Bertuch, A. A. (2010) The genetics and clinical manifestations of telomere biology disorders, *Genet Med 12*, 753-764.

22. Agarwal, S., Loh, Y. H., McLoughlin, E. M., Huang, J., Park, I. H., Miller, J. D., Huo, H., Okuka, M., Dos Reis, R. M., Loewer, S., Ng, H. H., Keefe, D. L., Goldman, F. D., Klingelhutz, A. J., Liu, L., and Daley, G. Q. (2010) Telomere elongation in induced pluripotent stem cells from dyskeratosis congenita patients, *Nature 464*, 292-296.

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

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

25. Kuprys, P. V., Davis, S. M., Hauer, T. M., Meltser, M., Tzfati, Y., and Kirk, K. E. (2013) Identification of telomerase RNAs from filamentous fungi reveals conservation with vertebrates and yeasts, *PLoS One 8*, e58661.

26. Egan, E. D., and Collins, K. (2012) Biogenesis of telomerase ribonucleoproteins, *RNA 18*, 1747-1759.

27. Ly, H., Blackburn, E. H., and Parslow, T. G. (2003) Comprehensive structure-function analysis of the core domain of human telomerase RNA, *Mol. Cell. Biol. 23*, 6849-6856.

28. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H., and James, T. L. (2002) A molecular switch underlies a human telomerase disease, *Proc Natl Acad Sci U S A 99*, 16998-17003.

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

30. Tzfati, Y., Fulton, T. B., Roy, J., and Blackburn, E. H. (2000) Template boundary in a yeast telomerase specified by RNA structure, *Science 288*, 863-867.

31. Lai, C. K., Miller, M. C., and Collins, K. (2002) Template boundary definition in Tetrahymena telomerase, *Genes Dev 16*, 415-420.

32. Blackburn, E. H., and Collins, K. (2011) Telomerase: an RNP enzyme synthesizes DNA, *Cold Spring Harb Perspect Biol 3*.

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

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

35. Singh, M., Wang, Z., Koo, B. K., Patel, A., Cascio, D., Collins, K., and Feigon, J. (2012) Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65, *Mol Cell 47*, 16-26.

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

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

38. Collins, K. (2008) Physiological assembly and activity of human telomerase complexes, *Mech. Ageing. Dev. 129*, 91-98.

39. Egan, E. D., and Collins, K. (2010) Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled in vivo, *Mol Cell Biol 30*, 2775-2786.

40. Fu, D., and Collins, K. (2003) Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs, *Mol. Cell 11*, 1361-1372.

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

42. Cech, T. R. (2004) Beginning to understand the end of the chromosome, *Cell 116*, 273-279.

43. Chappell, A. S., and Lundblad, V. (2004) Structural elements required for association of the Saccharomyces cerevisiae telomerase RNA with the Est2 reverse transcriptase, *Mol. Cell. Biol. 24*, 7720-7736.

44. Osterhage, J. L., and Friedman, K. L. (2009) Chromosome end maintenance by telomerase, *J Biol Chem 284*, 16061-16065.

45. Stellwagen, A. E., Haimberger, Z. W., Veatch, J. R., and Gottschling, D. E. (2003) Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends, *Genes Dev 17*, 2384-2395.

46. Williams, J. M., Ouenzar, F., Lemon, L. D., Chartrand, P., and Bertuch, A. A. (2014) The principal role of ku in telomere length maintenance is promotion of est1 association with telomeres, *Genetics 197*, 1123-1136.

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

48. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) A mutant of Tetrahymena telomerase reverse transcriptase with increased processivity, *J Biol Chem 275*, 24199-24207.

49. Collins, K. (2011) Single-stranded DNA repeat synthesis by telomerase, *Curr Opin Chem Biol 15*, 643-648.

50. Akiyama, B. M., Gomez, A., and Stone, M. D. (2013) A conserved motif in Tetrahymena thermophila telomerase reverse transcriptase is proximal to the RNA template and is essential for boundary definition, *J Biol Chem 288*, 22141-22149.

51. Huang, J., Brown, A. F., Wu, J., Xue, J., Bley, C. J., Rand, D. P., Wu, L., Zhang, R., Chen, J. J., and Lei, M. (2014) Structural basis for protein-RNA recognition in telomerase, *Nat Struct Mol Biol 21*, 507-512.

52. Robart, A. R., and Collins, K. (2011) Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation, *Mol Cell 42*, 308-318.

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

54. Sarafianos, S. G., Clark, A. D., Jr., Das, K., Tuske, S., Birktoft, J. J., Ilankumaran, P., Ramesha, A. R., Sayer, J. M., Jerina, D. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2002) Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA, *EMBO J 21*, 6614-6624.

55. Hossain, S., Singh, S., and Lue, N. F. (2002) Functional analysis of the C-terminal extension of telomerase reverse transcriptase. A putative "thumb" domain, *J Biol Chem 277*, 36174-36180.

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

57. Steitz, T. A. (1997) DNA and RNA polymerases: structural diversity and common mechanisms, *Harvey Lect 93*, 75-93.

58. Jiang, J., Miracco, E. J., Hong, K., Eckert, B., Chan, H., Cash, D. D., Min, B., Zhou, Z. H., Collins, K., and Feigon, J. (2013) The architecture of Tetrahymena telomerase holoenzyme, *Nature 496*, 187-192.

59. Sauerwald, A., Sandin, S., Cristofari, G., Scheres, S. H., Lingner, J., and Rhodes, D. (2013) Structure of active dimeric human telomerase, *Nat Struct Mol Biol 20*, 454-460.

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

61. Harkisheimer, M., Mason, M., Shuvaeva, E., and Skordalakes, E. (2013) A motif in the vertebrate telomerase N-terminal linker of TERT contributes to RNA binding and telomerase activity and processivity, *Structure 21*, 1870-1878.

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

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

64. Chen, Y., Fender, J., Legassie, J. D., Jarstfer, M. B., Bryan, T. M., and Varani, G. (2006) Structure of stem-loop IV of Tetrahymena telomerase RNA, *Embo Journal 25*, 3156-3166.

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

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

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

68. Richards, R. J., Wu, H., Trantirek, L., O'Connor, C. M., Collins, K., and Feigon, J. (2006) Structural study of elements of Tetrahymena telomerase RNA stem-loop IV domain important for function, *Rna-a Publication of the Rna Society 12*, 1475-1485.

69. Theimer, C. A., Finger, L. D., Trantirek, L., and Feigon, J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA, *Proc. Natl. Acad. Sci. USA 100*, 449-454.

70. Theimer, C. A., Jady, B. E., Chim, N., Richard, P., Breece, K. E., Kiss, T., and Feigon, J. (2007) Structural and functional characterization of human telomerase RNA processing and Cajal body localization signals, *Mol. Cell 27*, 869-881.

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

72. Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging, *Nature Medicine 12*, 1133-1138.

73. Tolbert, B. S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D. A., and Summers, M. F. (2010) Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement, *J Biomol NMR 47*, 205-219.

74. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic Acids Res 18*, 6069-6074.

CHAPTER 2

A pyrmidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo

Pyrimidine motif triple helix in the *Kluyveromyces lactis* telomerase RNA pseudoknot is essential for function in vivo

Darian D. Cash[a], Osnat Cohen-Zontag[b], Nak-Kyoon Kim[a], Kinneret Shefer[b], Yogev Brown[b], Nikolai B. Ulyanov[c], Yehuda Tzfati[b,1], and Juli Feigon[a,1]

[a]Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, CA 90095-1569; [b]Department of Genetics, The Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem 91904, Israel; and [c]Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94158-2517

Contributed by Juli Feigon, May 23, 2013 (sent for review April 17, 2013)

Telomerase is a ribonucleoprotein complex that extends the 3′ ends of linear chromosomes. The specialized telomerase reverse transcriptase requires a multidomain RNA (telomerase RNA, TER), which includes an integral RNA template and functionally important template-adjacent pseudoknot. The structure of the human TER pseudoknot revealed that the loops interact with the stems to form a triple helix shown to be important for activity in vitro. A similar triple helix has been predicted to form in diverse fungi TER pseudoknots. The solution NMR structure of the *Kluyveromyces lactis* pseudoknot, presented here, reveals that it contains a long pyrimidine motif triple helix with unexpected features that include three individual bulge nucleotides and a C^+•G-C triple adjacent to a stem 2–loop 2 junction. Despite significant differences in sequence and base triples, the 3D shape of the human and *K. lactis* TER pseudoknots are remarkably similar. Analysis of the effects of nucleotide substitutions on cell growth and telomere lengths provides evidence that this conserved structure forms in endogenously assembled telomerase and is essential for telomerase function in vivo.

RNA triplex | yeast | RNA structure | Hoogsteen

Telomerase is a ribonucleoprotein complex that extends the 3′ ends of eukaryotic chromosomes by adding successive telomere DNA repeats using an internal RNA template and a specialized reverse transcriptase (1, 2). Telomeres are the protein–DNA complexes that form the ends of linear chromosomes and protect them from end-to-end fusion and degradation (3, 4). Telomerase is of significant medical interest owing to the correlation between telomere length and human health and the association of telomerase activity with cancer (5, 6). In the absence of telomerase activity, telomeres shorten with each round of cell division because of exonuclease digestion and the inability of conventional DNA polymerases to fully replicate linear chromosomes. Shortening past a critical length leads to cell cycle arrest and/or apoptosis (7). Telomerase activity is undetectable in most somatic cells, resulting in telomere attrition with each cell cycle (8, 9). On the other hand, telomerase is active in, and essential for the proliferation of, the germ line, some epithelial, haemopoietic, and stem cells, as well as ∼90% of cancer cell lines (10, 11). A number of inherited diseases are associated with telomere shortening due to telomerase insufficiency, such as dyskeratosis congenita, aplastic anemia, and pulmonary fibrosis (12–15).

The telomerase holoenzyme consists of the telomerase reverse transcriptase (TERT) and telomerase RNA (TER), which are essential and sufficient for catalytic activity in vitro (16), and several species-specific accessory proteins. TERs are highly divergent in size and sequence between species, ranging from ∼150 nt in ciliates, ∼450 nt in vertebrates, to more than 2,000 nt in some fungi (17). TERs provide the template for telomeric DNA synthesis but also contain other domains that are essential for telomerase assembly and function. The most well-conserved region of TER is the template/pseudoknot (t/PK), or core domain, which contains a pseudoknot, the template, and a template boundary element (TBE) (18, 19) (Fig. 1). The t/PK interacts with TERT and together with another conserved region of TER, the stem terminus element, is critical for telomerase activity (17, 20–22).

The solution NMR structure of minimal human TER pseudoknots revealed that the loops interact with the stems to form a triple helix surrounding the helical junction (22, 23). Loop 1 interacts with the major groove of stem 2 to form a short pyrimidine motif triple helix extended by a loop 1–loop 2 Hoogsteen base pair, whereas loop 2 interacts with stem 1 to form two minor groove triples. Both DNA and RNA polymers and oligonucleotides readily form pyrimidine motif triplexes with U•A-U Watson–Crick (WC)/Hoogsteen paired triples (24–26), but this was the first observation of a pyrimidine motif triplex in a biologically functional RNA. The triple-helical interactions were shown to be important for catalytic activity in vitro. Yeast TERs are substantially larger than vertebrate TERs and have three long "arms" extending from the catalytic core at the center (4, 27–29) that serve as a scaffold for the binding of accessory and regulatory proteins (30–32). The reverse transcriptase, termed Est2 in yeast, binds to the central core domain (33). Mutational analysis and structure modeling has provided evidence for functionally important pyrimidine motif triplexes in yeast TER pseudoknots (34, 35), similar to the human pseudoknot loop 1 and stem 2 interactions (22, 23).

Here we report the NMR solution structure of a minimal *Kluyveromyces lactis* TER pseudoknot and provide evidence that the tertiary interactions observed are present in the endogenously assembled telomerase and are essential for telomerase function in vivo. Comparison of the *K. lactis* and human TER pseudoknots shows that despite different sequences, junctions, and base triples, the 3D structures of the triple helices are remarkably similar, presumably serving a conserved role in telomerase function.

Results and Discussion

Minimal Pseudoknot Design and Secondary Structure. For structural studies, a minimal *K. lactis* pseudoknot ΔU959 construct (kPKDU) was designed according to a proposed model derived from in vivo mutational studies (Fig. 1*B*) (34). The predicted bulged U959 was deleted and a terminal A added, to increase the stability of the pseudoknot. Deletion of the bulged U in the context of full-length TER has little or no effect on telomerase activity in vivo, as revealed by the BclI-tagged telomeric repeats incorporated by this mutant telomerase and the normal telomere length observed by

Author contributions: D.D.C., N.B.U., Y.T., and J.F. designed research; D.D.C., O.C.-Z., N.-K.K., K.S., and Y.B. performed research; D.D.C., O.C.-Z., K.S., N.B.U., Y.T., and J.F. analyzed data; and D.D.C., Y.T., and J.F. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 2M8K).

[1]To whom correspondence may be addressed. E-mail: tzfati@mail.huji.ac.il or feigon@mbi.ucla.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309590110/-/DCSupplemental.

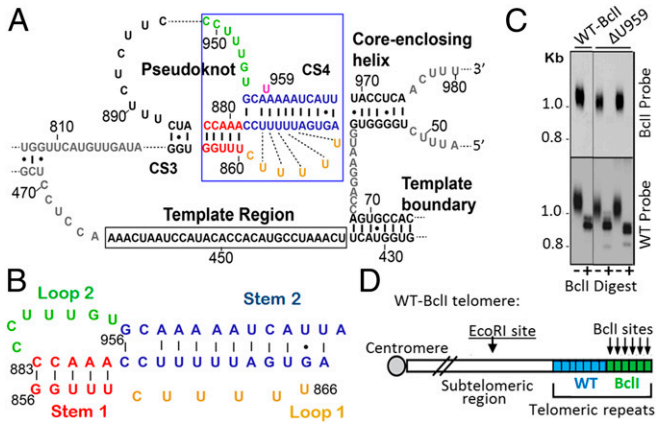


Fig. 1. (*A*) t/PK (or core) domain of *K. lactis* TER. Conserved sequences (CS) 3 and 4 make up the pseudoknot, which was truncated to the boxed nucleotides for structural studies. (*B*) Minimal *K. lactis* TER pseudoknot construct (kPKDU) for NMR studies with predicted secondary structure (34). Secondary structure elements are colored: stem 1 (red), loop 1 (gold), stem 2 (blue), loop 2 (green). (*C*) Southern analysis of telomere restriction fragments from *K. lactis* strains harboring WT or ΔU959 TER1 with BclI template mutation to mark telomerase action, shown in *D*. Genomic DNA samples were digested with EcoRI (-) or EcoRI + BclI (+) restriction endonucleases, electrophoresed in a 1% agarose gel, blotted, and hybridized first with a BclI-specific telomere probe (*Upper*) and then with a WT telomere probe (*Lower*). (*D*) Schematic representation showing a telomere containing WT (blue) and BclI (green) repeats.

Southern blotting and hybridization with a BclI-specific and WT telomeric probes (Fig. 1 *C* and *D* and Fig. S1). 1D imino spectra of kPKDU in different salt conditions reveal that pseudoknot formation is stabilized by either $MgCl_2$ or KCl (Fig. S2*B*). Further NMR structural studies were done with added Mg^{+2} (*Materials and Methods*).

Fig. 2*A* shows HNN- correlation spectroscopy (COSY) spectra of ^{13}C-,^{15}N-AU-, and ^{13}C,^{15}N-GC-labeled kPKDU aligned with the imino proton region of the 2D H_2O NOESY. The stems and loop 1 imino protons were assigned by standard methods using sequential NOE connectivities in the NOESY, along with HNN-COSY, which directly detects hydrogen bonds to differentiate A-U WC, A-U Hoogsteen, and G-C WC base pairs (36). Although not directly detected by HNN-COSY, an additional resonance is identified as a hydrogen bonded protonated C(861) imino by its distinct downfield chemical shift. Six sequential imino NOE connectivities were observed between the Hoogsteen base pairs of loop 1–stem 2 (C861 to U866), indicating a continuous triple helix (Fig. 2*A*). A Hoogsteen base paired triplex NOE pattern is also observed in the nonexchangeable NOESY spectra for the six nucleotides of loop 1 (22) (*SI Materials and Methods*).

At the helical junction between stem 1 and stem 2, a surprising difference from the predicted secondary structure is observed. There are two consecutive G-C base pairs at the top of stem 2, which could only form if an additional two nucleotides (U955 and C957) are bulged out (Fig. 2*A*, green). Sequential NOEs are observed through the stems and junction from U860–G954–G956–U876, which indicates that the pseudoknot is continuously base paired and stacked through the two stems. Furthermore, one of these G-C base pairs forms a base triple with the first C in loop 1 ($C861^{+}$•G956-C877). Thus, analysis of these spectra indicates that an extended pyrimidine motif triple helix forms between loop 1 and stem 2 that includes one C^{+}•G-C triple and five U•A-U triples, stem 2 extends by two G-C bases pairs, including a stem 2–loop 2 junction interaction, and there are two bulge bases at the top of stem 2 (Fig. 2*B*).

Solution Structure of kPKDU. The solution structure of kPKDU was solved using 1,027 NOE distance restraints and 549 dihedral restraints and refined with 98 residual dipolar couplings (RDCs) (Table S1). Superposition of the lowest 20 energy structures shows that the pseudoknot is well defined, with an overall rmsd to the mean of 1.1 ± 0.2 Å (Fig. 3*A* and Fig. S3*A*). A continuous A-RNA helix is formed by 17 consecutive WC stacked base pairs: 5 in stem 1, 1 in the stem 2–loop 2 junction (C878-G954), and 11 in stem 2. There are no significant bends in the helical axis. Near the junction, U955 and C957 are bulged out in the major and minor grooves, respectively (Fig. 3*C* and Fig. S3*B*) to allow for base pairing in the junction and top of stem 2. Loop 1 forms six successive Hoogsteen base pairs in the major groove of stem 2, creating a pyrimidine motif triple helix, whereas loop 2 lies in the minor groove of stem 1 on the same face of the structure. Stem 1 U860 and junction G954 are at stem/loop interfaces and adopt C2′ endo sugar pucker conformations. This increases the phosphate to phosphate distance, allowing placement of the adjoining residues in the grooves and stacking of U860 and G954. There is a sharp turn in the backbone at the end of the triple-helix (loop 1/stem 2) due to the consecutive residues U866 and A867 having tertiary/secondary interactions that separate them by five base pairs (Fig. 3*D* and Fig. S3*C*). These two residues are ~10 Å apart.

Because Mg^{+2} stabilized the folded pseudoknot, the sites of divalent metal ion localization were investigated by titrating $MnCl_2$ into kPKDU and observing changes in ^{1}H-^{13}C heteronuclear single-quantum coherence (HSQC) spectra (Fig. S4). Resonances of residues close to the bound Mn^{+2} will broaden and disappear owing to paramagnetic relaxation (37). Two sites of divalent cation localization are observed, in the major groove near residues at the sharp loop 1/stem 2 turn and in the minor groove of stem 1 where loop 2 binds. These ions likely stabilize the close

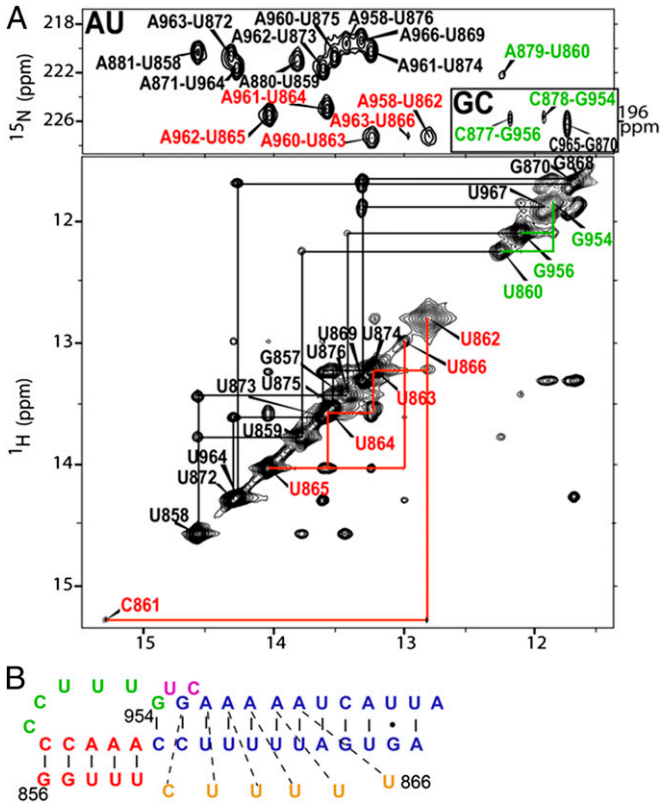


Fig. 2. (*A*) (*Upper*) AU and GC (*Inset*) *JNN*-HNN-COSY spectra aligned with (*Lower*) imino proton region of 2D H_2O NOESY at 278K and 800 MHz. NOE cross-peaks of stem 1/stem 2 WC pairs are connected above the diagonal in black and loop1–stem 2 Hoogsteen pairs below the diagonal in red. Iminos from the stem junction are connected below the diagonal and colored green. (*B*) Determined secondary structure of kPKDU (colored as in Fig. 1) with bulge nucleotides (U955, C957) in magenta.

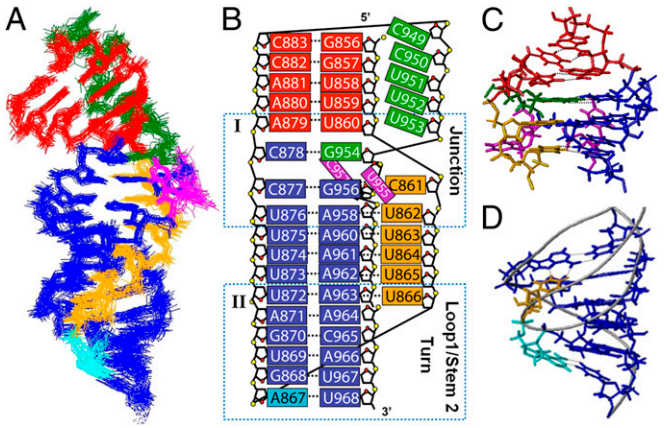


Fig. 3. (*A*) Twenty lowest energy structures from NMR solution structure determination of kPKDU superpositioned over all heavy atoms. (*B*) Schematic representation of the kPKDU tertiary structure. (*C*) Junction and (*D*) loop 1/stem 2 turn tertiary structures. Stereo views are presented in Fig. S3.

approach of the negatively charged phosphate backbone at the loop 1/stem 2 turn and the minor groove interactions of loop 2, respectively.

To determine whether the bulges and/or stem junction induce flexibility in the pseudoknot, we analyzed the RDC data to determine the generalized degree of order (GDO, or ϑ) parameter for the individual stems and assess the degree of interhelical motion. The internal GDO (ϑint = ϑstem1/ϑstem2) is equal to 1 for rigidly connected helices and decreases as interhelical motions increase dependent on motional amplitude and direction (38). kPKDU has $\vartheta_{int} = 0.96$, indicative of a rigid structure. Thus the junction stem 2-loop 2 base pair (C878-G954) and adjacent bulge nucleotides do not induce significant flexibility in the structure. Analysis of the RDCs from the WT minimal human pseudoknot, hPKWT (23), also gave a $\vartheta_{int} = 0.96$. The two pseudoknots have equally rigid structures despite differences in the stem junctions and number of bulge bases.

kPKWT Has the Same Tertiary Structure as kPKDU. After determining that two additional nucleotides were bulged out near the junction in close proximity to the deleted U959, a WT pseudoknot construct was made that included U959 (kPKWT) to further confirm that deletion of U959 did not cause significant structural differences. Analysis of the imino region of the 2D H_2O NOESY spectra shows the same NOE pattern for the two pseudoknots, with the only chemical shift differences for residues near the bulge U959 in kPKWT (Fig. S5). kPKWT has a base paired junction and stacked stems, with U955 and C957 bulged out. The NOE between the base paired iminos of U876 and U875 indicates that U959 is bulged out as predicted. The triple helix NOE pattern is also present for all six nucleotides of loop 1. We conclude that there are no significant tertiary structure differences between kPKDU and kPKWT.

Extended Triple Helix Stabilizes TER Pseudoknot Formation. To investigate the contribution of the tertiary interactions to the structure and folding of the pseudoknot, RNA constructs were made with nucleotide substitutions in the triple helix and examined by NMR. First, kPKDU, which has a truncated stem 1, was extended to a full-length stem 1 (kPKFL). Stem 2–loop 2 hairpin (S2L2) constructs were made, and their 1D imino spectra were compared to identify the conformation of the pseudoknot constructs in solution (Fig. 4, green). Under low salt conditions, kPKDU is primarily in the S2L2 conformation, whereas kPKFL primarily forms the pseudoknot, indicating the kPKFL pseudoknot is more stable (Fig. 4 *A* and *B*, blue). Addition of Mg^{+2} to kPKDU and kPKFL (*Materials and Methods*) shifts the equilibrium of both constructs completely to the pseudoknot conformation (Fig. 4 *A* and *B*, red).

Previous mutational data showed that loop 1 substitution 864-6UUU:CCC, which should disrupt three of the six base triples, abolished telomerase activity in vivo (33). NMR studies of this substitution as well as 861-3CUU:UCC in kPKFL show that these RNAs do not form a pseudoknot, even when Mg^{+2} is present; instead only an elongated S2L2 forms (Fig. 4 *C* and *D*). Additional imino peaks in these hairpins arise from formation of additional U-A base pairs at the top of stem 2 (diagrammed in Fig. 4 *C* and *D*). Thus, base triples are needed to stabilize the formation of stem 1 and therefore folding of the pseudoknot as it is transcribed, revealing an important role for the extended triple helix in *K. lactis* TER. In the human pseudoknot, changing loop 1 residues (99-100UU:CC), which greatly decreases activity in vitro (20), abolished formation of stem 1 and stem 2 base pairs around the junction and all tertiary triplex interactions (39). This nucleotide substitution also stabilized an alternate conformation in high salt conditions, as shown by FRET analysis (39).

Mutational Studies Indicate the Determined Structure is Important for Function in Vivo. To test whether the determined tertiary interactions of kPKDU are important for function in vivo, a series of TER mutations were made, and their effects on telomerase activity, telomere length, and colony phenotype were investigated. Our structure revealed a triple helix with five U•A-U base triples in a distinct register (register 1), starting with the first stem 2 A-U base pair (relative to the junction) (Fig. 5*A*). However, modeling studies indicated the possibility of an alternate register (register 2) in which only four U•A-U triples would form, starting with the second stem 2 A-U base pair (Fig. 5*B*). To determine which register formed in vivo, mutations were designed to affix each of the conformations separately (Fig. 5 *C* and *D*). First, alternate As in the top strand (A) of stem 2 were replaced with Gs (S2A). Second, alternate Us in the bottom strand (B) of stem 2 were replaced with Cs, and alternate Us in loop 1 were replaced with Cs starting with the first U (862) or the second U (863) (S2B+L1 and SB2+L1′,

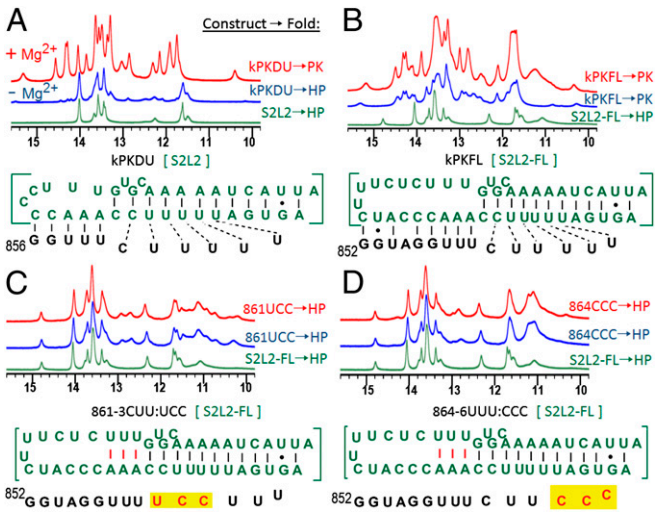


Fig. 4. One-dimensional imino proton spectra at 283K and 500 MHz of RNA constructs of the *K. lactis* pseudoknot and stem 2–loop 2 hairpin in 10 mM Tris·D (pH 6.3) (green and blue) and with added Mg^{+2} (red) (*Materials and Methods*). Hairpin (green, bracketed) and pseudoknot sequences are shown below the spectra. (*A*) kPKDU and S2L2, (*B*) kPKFL and corresponding S2L2-FL. kPKFL and S2L2-FL have the full-length stem 1. (*C*) kPKFL(861-3CUU:UCC) and S2L2-FL, and (*D*) kPKFL(864-866UUU:CCC) and S2L2-FL. Dashed lines on the sequences indicated observed tertiary base pairs. Nucleotide substitutions in *C* and *D* were designed to prevent formation of the first three or last three triples. Spectra are labeled with the RNA construct and primary tertiary fold, hairpin (HP), or pseudoknot (PK), under the given conditions.

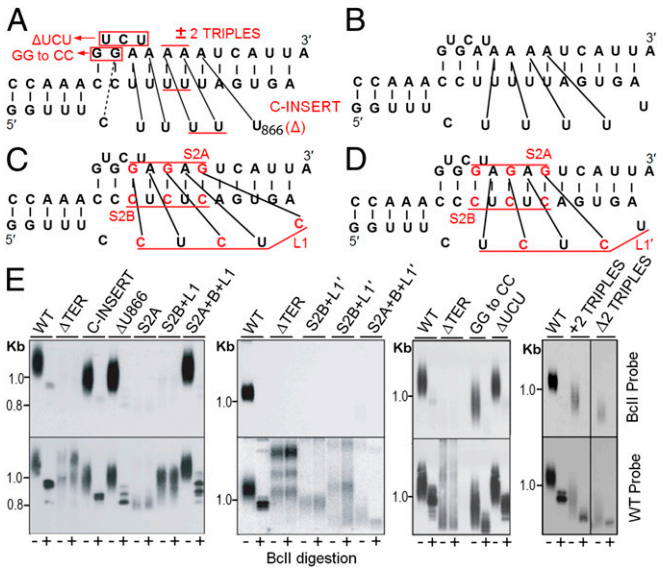


Fig. 5. (*A* and *B*) Two alternative base triple registers, 1 and 2, predicted for the *K. lactis* pseudoknot. Register 1 (*A*) is observed in the solution structure. (*C* and *D*) Compensatory mutations designed to affix (*C*) register 1 and (*D*) register 2 in vivo. Mutations were made in the top strand (S2A), bottom strand (S2B), and loop (L1, L1′). (*E*) Representative Southern analysis of telomere restriction fragments prepared from WT, ΔTER1, and pseudoknot mutants (shown in red in *A*, *C*, and *D*) (*Materials and Methods* and Fig. S1).

respectively). The individual S2A, S2B+L1, and S2B+L1′ mutations, which are expected to disrupt stem 2 formation and potential tertiary interactions, abolished telomerase activity in vivo, as apparent by the lack of BclI repeat incorporation, the smeared pattern of telomere restriction fragments typical of an alternative, recombination-dependent, telomere elongation mechanism (Fig. 5*E*, assay schematic in Fig. S1), and a rough colony phenotype indicative of impaired telomere maintenance (Fig. S6). These results demonstrate that stem 2 and triplex formation is essential for telomerase function in vivo. The combined S2A+S2B+L1 or S2A+S2B+L1′ mutations were designed such that alternating U•A-U and C^+•G-C triples would form along the helix, because adjacent C^+•G-C triples are very unstable (owing to the requirement for protonation of the Hoogsteen paired C) and fail to support normal telomerase function (34). Strikingly, the yeasts with TER S2A+S2B+L1 (register 1) substitutions have normal telomerase activity, telomere length, and colony morphology, whereas S2A+S2B+L1′ (register 2) results in barely detectable telomerase activity, severely short telomeres, and rough colonies (Fig. 5*E* and Fig. S5). These results indicate that the particular register is essential for telomerase function in vivo. The importance of the length of the triple helix was also investigated by making mutations that extend or shorten it by two base triples. These sets of mutations caused moderate and severe telomere shortening, respectively, but did not completely abolish telomerase activity, as indicated by the incorporation of BclI repeats (Fig. 5*E*). Although the shortening of the triplex may destabilize the structure, the effect of its extension indicates that the length of the triplex or the position of nucleotides at the apical part of loop 1 or adjacent stem 2 are crucial for telomerase function. The importance of the length of loop 1 was tested by deleting the apical nucleotide U866 or inserting a C downstream to it. Both of these mutations also caused moderate but significant shortening of telomeres (Fig. 5*E*). Deletion of U866 would abolish the last triple and likely disrupt the terminal A867-U986 base pair, and its effect is consistent with the length of the triple helix being important for telomerase activity in vivo. The extra C in the loop could affect the positioning of the third strand or could interfere with a TERT specific interaction. In *Saccharomyces cerevisiae*, TERT is predicted to interact in this region adjacent to the triple helix (35).

An unexpected feature of the *K. lactis* WT pseudoknot was the presence of three bulge nucleotides (C955, U957, and U959) in the purine strand of the triplex adjacent to the top three triples. A bulged U177 in the human telomerase pseudoknot below the bottom triple significantly destabilizes stem 2 and therefore the loop 2–stem 2 hairpin (23, 36, 40), and deletion of U177 decreases activity in vitro (22, 40). On the basis of the kPKDU structure, we predicted that deletion of the three bulge nucleotides from the *K. lactis* pseudoknot would stabilize the pseudoknot but otherwise not affect its conformation. To test the importance of the bulge nucleotides in the *K. lactis* pseudoknot on telomerase function in vivo, we deleted them simultaneously. This triple deletion had no significant effect on telomerase activity or telomere length (Fig. 5*E*), indicating that the bulge residues in the *K. lactis* pseudoknot are not important for telomerase function in vivo. Finally, G954 and G956 were substituted with Cs, disrupting the junction base pair and C^+•G-C base triple. This substitution caused significant telomere shortening, demonstrating the importance of this end of stem 2 and the triplex (Fig. 5*E*).

The previous mutational analysis correctly identified a loop 1–stem 2 triple helix but failed to define the junction base pairs (Fig. 1*A*) (34). Deleting the bulge (Δ876-8) in the initially predicted structure or making a more definitive bulge (876-8UCC:AUG + 957-9CAU:GGA) abolished activity, whereas base pairing the bulge (876-8UCC:AUG, or 957-9CAU:GGA) maintained WT activity, indicating that the junction must base pair to function (34). The solution structure explains these data, revealing that three residues bulge out to allow for a base paired junction and formation of two additional base triples, including a C^+•G-C triple. Taken together, the effects of the pseudoknot substitutions on telomere length in vivo confirm that the secondary and tertiary interactions in kPKDU determined by NMR are present in the context of the full-length TER in the telomerase holoenzyme and that formation of the extended triple helix is critical for telomerase activity in vivo.

***Kluyveromyces* Species TER Pseudoknots Have a Conserved Tertiary Structure.** Previous analysis of TER pseudoknot sequences from the *Kluyveromyces marxianus* cluster of species revealed conservation of a potential triple helix (34). We have cloned the TER gene from a newly identified species, *Kluyveromyces siamensis* (41). Six of the 40 nucleotides forming the *K. siamensis* pseudoknot differ from the *K. lactis* sequence, and it lacks all three of the nucleotides that are bulged out in the *K. lactis* pseudoknot. We predicted the secondary structure on the basis of the *K. lactis* structure and found that it could form an identical secondary structure except for the absence of the three bulge nucleotides (Fig. S7). We reanalyzed the five other known *Kluyveromyces sp.* pseudoknot sequences and found that by bulging out none, one, two, or three residues, as necessary, each could form a similar triple helix with five consecutive U•A-U triples and in all but one case an additional C^+•G-C or U•G-C triple (Fig. S7). The absence of conservation of the bulge nucleotides along with the results of the mutagenesis data indicate that the bulge nucleotides are dispensable for function, in contrast to the case for the bulge U177 in the human pseudoknot.

Comparison with *S. cerevisiae* Telomerase Pseudoknot. *S. cerevisiae* TER (TLC1) was shown to form a pseudoknot with base triple interactions involving stem 2 (A803-806), which are important for telomerase activity (35). However, the precise secondary and tertiary structure of this pseudoknot has yet to be defined. A recent thermodynamics study predicts that stem 2 is extended by 6 base pairs beyond a 6-nt bulge loop and that loop 1 forms additional triples with this stem, evidentially supported by sequence conservation of the U•A-U residues (Fig. S8) (42). We noticed that two additional C^+•G-C triples could form with this part of stem 2 by protonating the Cs (741-742) in loop 1. Preliminary NMR data of an *S. cerevisiae* minimal pseudoknot shows two distinct downfield shifted iminos that correspond to protonated

Cs (Fig. 6*B*), supporting the hypothesis of an extended bipartite triple helix in the *S. cerevisiae* pseudoknot. Pseudoknots with a bipartite triple helical region have been predicted to form in *Candida* yeast TER as well (43). At the helical junction, the *S. cerevisiae* pseudoknot has an apparent 3-nt bulge that would preclude the formation of the continuous stem 1–stem 2 stacking interactions seen in the *K. lactis* pseudoknot. However, we noticed that by bulging out residues A733-G734 the three As in the predicted bulge could base pair to extend stem 1 (Fig. 6*A*). This would allow for a stacked junction proximal to an extended triple helix, resulting in a pseudoknot structurally similar to that of *K. lactis*.

Human and *K. lactis* Pseudoknot Structures Are Remarkably Similar. Comparison of the tertiary structures of the pseudoknots from *K. lactis* (kPK) and human (hPK) (22, 23) reveals that despite the lack of significant sequence identity the 3D structures are strikingly similar (Fig. 7). Both *K. lactis* and human PKs have continuous stacking interactions between the two stems across the junction. kPK has WC base pairing in the junction allowing the stems to coaxially stack, while the hPK stems are separated by a loop–loop Hoogsteen interaction. This causes underwinding and a small bend in the helical axis at the junction of the two stems in hPK. However, a second bend at the bulge U177 results in an overall straight stem similar to kPK. Both pseudoknots have a sharp loop 1/stem 2 turn where the phosphates in the backbone come very close together, and it is likely that cation binding in this region is important for stabilizing the tertiary structure (39). Although the extended triple helix has a different arrangement of base triples around the junction for kPK vs. hPK, the triple-helical region is the same length. hPK has three Hoogsteen base triples in the major groove of stem 2, one loop–loop Hoogsteen base pair in the junction, and two base triples in the minor groove of stem 1, which stack consecutively (nucleotides 99–102, 172–173; Fig. 7*B*). kPK forms a triple helix with the six sequential residues of loop 1, including a C^+•G-C triple, and has no minor groove or loop–loop base interactions (Fig. 7*A*).

Although both human and *K. lactis* TER pseudoknots have bulged residues in the purine-rich strand of stem 2, examination of the structures reveals they are at different locations (Fig. 7). The bulged residues in kPK are near the junction at the 5′ end of the stem 2, and deletion of one or all three of these bulge residues has little effect on telomerase activity or telomere length in vivo. hPK has a bulged residue (U177) in stem 2 just below the triple helix, deletion of which decreases activity by 50% in vitro (22, 40). Backbone superposition of the two structures (Fig. 7*C*) reveals that the bulge residues are on opposite faces of the helix. In the model of the human TER core domain (44), the bulge U177 is near the template and facing toward the modeled TERT and putative active site. In contrast, bulges near the junction like those found in the *K. lactis* pseudoknot would apparently face away from the active site, potentially explaining their lack of effect on telomerase function.

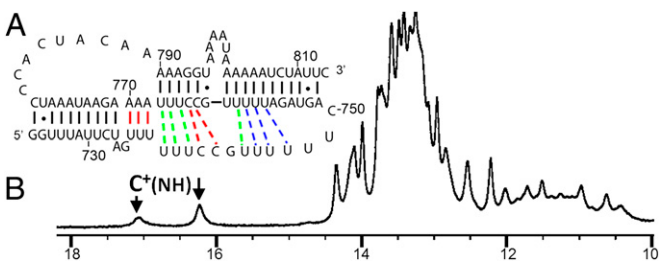


Fig. 6. (*A*) Minimal *S. cerevisiae* pseudoknot with predicted secondary structure and tertiary interactions. Interactions are validated by assays (blue), sequence conservation (green), or unconfirmed (red). (*B*) One-dimensional imino spectra of *S. cerevisiae* pseudoknot with protonated C^+ iminos (NH) indicated by arrows, recorded at 283K on 600-MHz NMR. Sample buffer is 10 mM Na phosphate (pH 6.3) and 50 mM KCl.

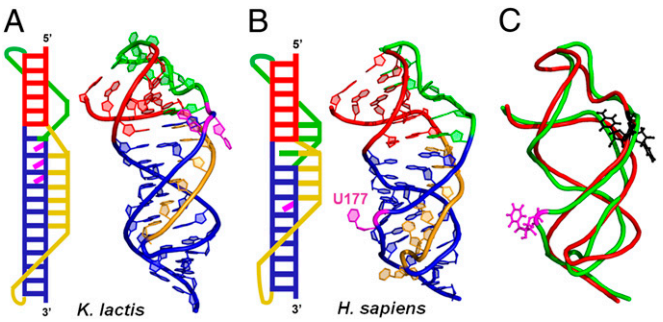


Fig. 7. Solution structures of minimal TER pseudoknots from (*A*) *K. lactis* (kPK) and (*B*) *Homo sapiens* (hPK), colored as in Fig. 1. Tertiary structure schematics (*Left*) and lowest energy structures (*Right*) are shown. (*C*) Superposition of the backbones of the kPK (red) and hPK (green) tertiary structures, with bulges as stick/ball for kPK (black) and hPK (magenta).

Pyrimidine Motif Triplexes Stabilize Other Noncoding RNAs. Pseudoknots are a prevalent RNA motif, and a variety of major groove triples have been found in riboswitch aptamers that fold into pseudoknots for ligand capture [e.g., SAM-II riboswitch (45)], as well as in the group II intron near the catalytic site (46). However, pyrimidine motif triplexes (i.e., three or more consecutive U•A-U triples) have only been found in pseudoknots from human and yeast telomerase. Recently, pyrimidine motif triplexes were discovered in the viral polyadenylated nuclear (PAN) RNA of Kaposi's sarcoma-associated herpesvirus (47, 48) and the long noncoding RNAs (lncRNAs) metastasis-associated lung adenocarcinoma transcript 1 (MALAT1) and multiple endocrine neoplasia-β (MENβ), where they function to protect the 3′ A-rich ends from degradation (49, 50). Five U•A-U triples form in PAN RNA, whereas a bipartite triple helix containing two regions of three to five U•A-U triples separated by a bulge (with a possible C^+•G-C triple) form in MENβ and MALAT1, similar to *S. cerevisiae* and *Candida* pseudoknot predictions (42, 43). Although the secondary structure of the pseudoknot has two stems, it is likely that the viral RNAs, lncRNAs, and TER pseudoknots all form the triplex in the same way, with formation of an initial stem (stem 1 in pseudoknots) as the RNA is transcribed followed by insertion of a 3′ A-rich strand into a U-rich hairpin (36) or internal loop.

Conclusions

The triple helix has emerged as a conserved and essential element that stabilizes the pseudoknot fold in TERs. Folding of the TER pseudoknot is likely important for positioning the template and the TBE that regulates telomere repeat synthesis, at the catalytic site of TERT (33, 35). TERT does not seem to interact strongly with the triple helical region of the pseudoknot but rather with the adjacent base pairs near the core enclosing helix or TBE (33, 35). Previous studies of the human telomerase pseudoknot established the importance of the triple helix interactions for activity of telomerase reconstituted in vitro (22, 35). However, mutations that affect catalysis in vitro often have different effects in vivo (51). Here we have directly correlated the secondary and tertiary interactions observed in a determined structure of a pseudoknot with the effects of mutations that disrupt those interactions in vivo on telomerase activity, telomere length, and cell viability.

Studies of telomere-associated proteins have revealed strong structural similarities by divergent polypeptide sequences from distant species (e.g., TPP1 in human, Est3 in yeast, and TEBP-β in ciliates) (52). Similarly, the results presented here also indicate that although TER is highly divergent and presents no apparent sequence similarity across yeast and vertebrates, human and *K. lactis* TERs form remarkably similar structures around a conserved extended triple helix. Maintaining such a high structural conservation by RNA molecules that evolve almost as rapidly as

intergenic regions reflects evolutionary constraints to conserve an essential telomerase function.

Materials and Methods

NMR Sample Preparation, Data Acquisition, and Structure Calculations. RNAs were synthesized by in vitro transcription with T7 RNA polymerase on a synthetic DNA template and purified by PAGE as previously described (22). For structure determination of kPKDU, unlabeled, (^{13}C,^{15}N)-AU and GC labeled, and fully labeled RNA samples were made. RNA concentrations were ~1 mM in a sample buffer of 10 mM Tris-D (98% deuterated) (pH 6.3), 0.5 mM $MgCl_2$. NMR spectra were recorded on Bruker DRX 500- and 600- and Avance 800-MHz spectrometers equipped with HCN cryoprobes. Spectra were recorded at 298K in 100% D_2O, except for imino data, which were recorded at 278K and 283K in 10% D_2O/90% (vol/vol) H_2O. NMR experiments used and methods for obtaining assignments and restraints, RDC analysis, and structure calculations were done as previously described (22, 23). Details are given in the *SI Materials and Methods*.

In Vivo Telomere Length Assay. *K. lactis ter1Δ* strain yJR27 was used for the plasmid shuffling of WT with mutant TER genes encoded on a CEN-ARS plasmid, as described previously (53). Both WT and mutant *TER1* genes contained an additional *BclI* template mutation that is incorporated into the nascent telomeric repeats, introducing BclI restriction sites. Otherwise it does not affect telomerase activity or telomere length and is thus used to mark the action of the examined telomerase. Genomic DNA was prepared from the yeast strains at their sixth passage and analyzed by Southern blotting and hybridization first to a BclI-specific and then to a WT *K. lactis* telomeric probe, as described previously (34). A schematic of the assay is shown in Fig. S1. At least two clones were examined for each mutation.

ACKNOWLEDGMENTS. We thank Savitree Limtong, Kasetsart University, for the gift of *K. siamensis* and Jing Zhou for help to D.D.C. This work was supported by National Science Foundation Grant MCB1022379 and National Institutes of Health (NIH) Grant GM048123 (to J.F.) and United States-Israel Binational Science Foundation Grant 2009204 (to N.B.U. and Y.T.). D.D.C. is a member of the Cellular and Molecular Biology Training Grant program at University of California, Los Angeles supported by NIH Grant GM007185.

1. Blackburn EH (2001) Switching and signaling at the telomere. *Cell* 106(6):661–673.
2. Blackburn EH, Collins K (2011) Telomerase: An RNP enzyme synthesizes DNA. *Cold Spring Harb Perspect Biol* 3(5):3.
3. Londoño-Vallejo JA, Wellinger RJ (2012) Telomeres and telomerase dance to the rhythm of the cell cycle. *Trends Biochem Sci* 37(9):391–399.
4. Cech TR (2004) Beginning to understand the end of the chromosome. *Cell* 116(2): 273–279.
5. de Jesus BB, Blasco MA (2012) Potential of telomerase activation in extending health span and longevity. *Curr Opin Cell Biol* 24(6):739–743.
6. Blasco MA (2003) Telomeres and cancer: A tale with many endings. *Curr Opin Genet Dev* 13(1):70–76.
7. Collins K, Mitchell JR (2002) Telomerase in the human organism. *Oncogene* 21(4): 564–579.
8. Aubert G, Lansdorp PM (2008) Telomeres and aging. *Physiol Rev* 88(2):557–579.
9. Blasco MA (2007) Telomere length, stem cells and aging. *Nat Chem Biol* 3(10): 640–649.
10. Wong JM, Collins K (2003) Telomere maintenance and disease. *Lancet* 362(9388): 983–988.
11. Shay JW, Wright WE (2006) Telomerase therapeutics for cancer: Challenges and new directions. *Nat Rev Drug Discov* 5(7):577–584.
12. Armanios M, Blackburn EH (2012) The telomere syndromes. *Nat Rev Genet* 13(10): 693–704.
13. Chen JL, Greider CW (2004) Telomerase RNA structure and function: Implications for dyskeratosis congenita. *Trends Biochem Sci* 29(4):183–192.
14. Kirwan M, Dokal I (2008) Dyskeratosis congenita: A genetic disorder of many faces. *Clin Genet* 73(2):103–112.
15. Calado RT, Young NS (2008) Telomere maintenance and human bone marrow failure. *Blood* 111(9):4446–4455.
16. Collins K (2006) The biogenesis and regulation of telomerase holoenzymes. *Nat Rev Mol Cell Biol* 7(7):484–494.
17. Egan ED, Collins K (2012) Biogenesis of telomerase ribonucleoproteins. *RNA* 18(10): 1747–1759.
18. Chen JL, Greider CW (2003) Template boundary definition in mammalian telomerase. *Genes Dev* 17(22):2747–2752.
19. Chen JL, Greider CW (2004) An emerging consensus for telomerase RNA structure. *Proc Natl Acad Sci USA* 101(41):14683–14684.
20. Chen JL, Greider CW (2005) Functional analysis of the pseudoknot structure in human telomerase RNA. *Proc Natl Acad Sci USA* 102(23):8080–8085, discussion 8077–8079.
21. Ly H, Blackburn EH, Parslow TG (2003) Comprehensive structure-function analysis of the core domain of human telomerase RNA. *Mol Cell Biol* 23(19):6849–6856.
22. Theimer CA, Blois CA, Feigon J (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* 17 (5):671–682.
23. Kim NK, et al. (2008) Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J Mol Biol* 384(5):1249–1261.
24. Felsenfeld G, Rich A (1957) Studies on the formation of two- and three-stranded polyribonucleotides. *Biochim Biophys Acta* 26(3):457–468.
25. Arnott S, Bond PJ (1973) Structures for Poly(U)-poly(A)-poly(U)triple stranded polynucleotides. *Nat New Biol* 244(134):99–101.
26. Rajagopal P, Feigon J (1989) Triple-strand formation in the homopurine:homopyrimidine DNA oligonucleotides d(G-A)4 and d(T-C)4. *Nature* 339(6226):637–640.
27. Tzfati Y, Fulton TB, Roy J, Blackburn EH (2000) Template boundary in a yeast telomerase specified by RNA structure. *Science* 288(5467):863–867.
28. Dandjinou AT, et al. (2004) A phylogenetically based secondary structure for the yeast telomerase RNA. *Curr Biol* 14(13):1148–1158.
29. Chappell AS, Lundblad V (2004) Structural elements required for association of the Saccharomyces cerevisiae telomerase RNA with the Est2 reverse transcriptase. *Mol Cell Biol* 24(17):7720–7736.
30. Bertuch AA, Lundblad V (2006) The maintenance and masking of chromosome termini. *Curr Opin Cell Biol* 18(3):247–253.
31. Hug N, Lingner J (2006) Telomere length homeostasis. *Chromosoma* 115(6):413–425.
32. Zappulla DC, Cech TR (2004) Yeast telomerase RNA: A flexible scaffold for protein subunits. *Proc Natl Acad Sci USA* 101(27):10024–10029.
33. Lin J, et al. (2004) A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc Natl Acad Sci USA* 101(41):14713–14718.
34. Shefer K, et al. (2007) A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol Cell Biol* 27(6):2130–2143.
35. Qiao F, Cech TR (2008) Triple-helix structure in telomerase RNA contributes to catalysis. *Nat Struct Mol Biol* 15(6):634–640.
36. Theimer CA, Finger LD, Trantirek L, Feigon J (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc Natl Acad Sci USA* 100(2):449–454.
37. Butcher SE, Allain FH, Feigon J (2000) Determination of metal ion binding sites within the hairpin ribozyme domains by NMR. *Biochemistry* 39(9):2174–2182.
38. Al-Hashimi HM, et al. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings. *J Mol Biol* 315(2):95–102.
39. Hengesbach M, Kim NK, Feigon J, Stone MD (2012) Single-molecule FRET reveals the folding dynamics of the human telomerase RNA pseudoknot domain. *Angew Chem Int Ed Engl* 51(24):5876–5879.
40. Comolli LR, Smirnov I, Xu L, Blackburn EH, James TL (2002) A molecular switch underlies a human telomerase disease. *Proc Natl Acad Sci USA* 99(26):16998–17003.
41. Am-In S, Yongmanitchai W, Limtong S (2008) *Kluyveromyces siamensis* sp. nov., an ascomycetous yeast isolated from water in a mangrove forest in Ranong Province, Thailand. *FEMS Yeast Res* 8(5):823–828.
42. Liu F, Kim Y, Cruickshank C, Theimer CA (2012) Thermodynamic characterization of the *Saccharomyces cerevisiae* telomerase RNA pseudoknot domain in vitro. *RNA* 18 (5):973–991.
43. Gunisova S, et al. (2009) Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements. *RNA* 15(4):546–559.
44. Zhang Q, Kim NK, Feigon J (2011) Architecture of human telomerase RNA. *Proc Natl Acad Sci USA* 108(51):20325–20332.
45. Gilbert SD, Rambo RP, Van Tyne D, Batey RT (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol* 15(2):177–182.
46. Toor N, Keating KS, Taylor SD, Pyle AM (2008) Crystal structure of a self-spliced group II intron. *Science* 320(5872):77–82.
47. Mitton-Fry RM, DeGregorio SJ, Wang J, Steitz TA, Steitz JA (2010) Poly(A) tail recognition by a viral RNA element through assembly of a triple helix. *Science* 330(6008): 1244–1247.
48. Tycowski KT, Shu MD, Borah S, Shi M, Steitz JA (2012) Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell Rep* 2(1):26–32.
49. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA (2012) Formation of triple-helical structures by the 3′-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc Natl Acad Sci USA* 109(47):19202–19207.
50. Wilusz JE, et al. (2012) A triple helix stabilizes the 3′ ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* 26(21):2392–2407.
51. Robart AR, Collins K (2010) Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J Biol Chem* 285(7): 4375–4386.
52. Lewis KA, Wuttke DS (2012) Telomerase and telomere-associated proteins: Structural insights into mechanism and evolution. *Structure* 20(1):28–39.
53. Roy J, Fulton TB, Blackburn EH (1998) Specific telomerase RNA residues distant from the template are essential for telomerase function. *Genes Dev* 12(20):3286–3300.

Supporting Information

Cash et al. 10.1073/pnas.1309590110

SI Materials and Methods

NMR Sample Preparation. RNAs were synthesized by in vitro transcription on a synthetic DNA template. The components of the reaction are $MgCl_2$ (20–30 mM), DTT (2.5 mM), DNA template (1 μM), buffer [40 mM Tris (pH 8.1), 1 mM spermidine, and 0.01% Triton], NTPs (2–5 mM), and His_6-tagged P266L phage T7 mutant polymerase (1). Unlabeled NTPs and/or ^{13}C,^{15}N-labeled NTPs were used to prepare unlabeled, AU-labeled, GC-labeled, or fully labeled RNA samples. After transcription, the RNA was purified by PAGE, excising the band of interest, and eluting the RNA from the gel using an ELUTRAP device, as previously described (2, 3). The purified RNA was washed using the Amicon filtration system (three times with 1 M NaCl, four times with H_2O) before being exchanged into 10 mM Tris·D (98% deuterated) (pH 6.3). The sample was concentrated to ∼0.3 mM, and $MgCl_2$ was added to 15 mM at room temperature. Unbound Mg^{2+} was removed by exchanging the sample into the final NMR buffer [10 mM Tris·D (pH 6.3) and 0.5 mM $MgCl_2$], and the sample was concentrated to ∼1 mM. Optimal conditions for formation of the pseudoknot were determined by titration of samples of *Kluyveromyces lactis* pseudoknot ΔU959 construct (kPKDU), prepared in 10 mM Tris·D (pH 6.3) buffer as described above, with $MgCl_2$ and KCl and acquiring 1D imino proton spectra at each titration point (Fig. S2), as well as 2D NOESY spectra for conditions in which the sample appeared folded.

NMR spectra were recorded on Bruker DRX 500- and 600- and Avance 800-MHz spectrometers equipped with HCN cryoprobes. Nonexchangeable protons and their bound carbons were assigned by taking spectra in D2O at 293K. The assignments were obtained by analyzing a series of NMR experiments, including 2D NOESY, 2D total correlation spectroscopy (TOCSY), ^{1}H-^{13}C heteronuclear single-quantum coherence (HSQC), 2D HCCH-correlation spectroscopy (COSY) 3D HCCH-TOCSY, and a suite of 2D filtered/edited NOESY experiments (F1fF2f, F2f, F1fF2e, F1eF2e) (3–6). Exchangeable proton spectra were taken using samples in 95% H_2O/5% D2O at 278K and 283K. These proton and nitrogen resonances were assigned using 2D-NOESY, ^{1}H-^{15}N HMQC, and ^{15}N-correlated Carr-Purcell-Meiboom-Gill (CPMG)-NOESY experiments. Initial NOE sequential assignments and final NOEs for structure calculations were also obtained from these spectra as previously described (6–8). Hydrogen bonds for Watson–Crick and Hoogsteen base pairs were directly detected with HNN-COSY experiments (9, 10) for all base pairs except for terminal base pairs G856-C883 and U968-A866, and the $C861^{+}$-G956 Hoogsteen pair. However, observation of an imino proton resonance and NOE cross-peaks confirmed the presence of the $C861^{+}$-G956 Hoogsteen pair. All NMR spectra were processed using Bruker XWINNMR, TOPSPIN, and Sparky.

Residual dipolar couplings (RDCs) were measured for C-H (^{1}DC1′H1′, ^{1}DC2H2, ^{1}DC5H5, ^{1}DC6H6, ^{1}DC8H8) and N-H (^{1}DN1H1, ^{1}DN3H3) bonds using a uniformly ^{13}C,^{15}N-labeled kPKDU sample using an 800-MHz spectrometer (11). For RDC data collection, the RNA sample was dialyzed to buffer [10 mM PO_4 (pH 6.3) and 0.5 mM $MgCl_2$], which did not cause any significant shifts in the spectra. The spectra were processed with Bruker TOPSPIN and analyzed with NMRDraw. A total of 77 C-H and 21 N-H RDCs were determined by measuring the difference in splittings in the ^{1}H and ^{15}N dimensions between spectra taken in the absence and presence of 16.5 mg/mL Pf1 phage. For N-H bonds, the two RDC values obtained from ^{1}H and ^{15}N dimensions were averaged, unless there was a discrepancy, in which case only the ^{15}N dimension was used. The generalized degree of order (ϑ) (12) for the individual stems was derived from order matrix analysis (13).

NMR Assignments. Assignment of the nonexchangeable proton resonances and their bound carbons were obtained as previously described (3–6). All of the base protons and proton-attached carbons were assigned, but the sugars were only partially assigned (∼90%) owing to overlap in the spectra. Examination of the 2D D_2O NOESY shows NOEs that are consistent with formation of A-RNA helices in the two stems. The first G-C base pair (G856-C883) and the last A-U base pair (A867-U968) were not directly detected but have NOEs indicative of stacking in a helix. Therefore, these base pairs were included as restraints in the structure calculation. U955 and C957 have relatively broad line widths and do not have sequential NOEs to their neighboring nucleotides, supporting the conclusion that they are bulged out. Instead, junction residues G954-G956-A958 have sequential NOEs consistent with stacking and continuous base pairing within a helix. Standard sequential NOEs were also observed for U860 to G954. These data indicate that the two stems and junction form one continuous A-form helix.

A Hoogsteen base paired triplex NOE pattern is observed for the six nucleotides of loop 1, C861 to U866 (14). This includes a strong-intensity cross-strand NOE of the A/G H8 and the H1′ of the 5′ neighboring Hoogsteen base paired U/C, as well as a moderate-intensity NOE between the H1′s of the same respective neighboring residues in the 2D D2O NOESY. This pattern was observed for all six of the loop 1 residues. There was no evidence of loop 1/loop 2 base pairing interactions that are present in the human pseudoknot.

Structure Calculations. Interproton distances were determined from the integration of cross-peaks in 2D-NOESY and 2D filtered/edited NOESY experiments. The NOE distance restraints were classified as very strong (1.8–3.5 Å), strong (2.5–4.5 Å), medium (3.5–5.5 Å), weak (4.5–6.5 Å), and very weak (5.5–7.5 Å), normalized using the average stem pyrimidine H5-H6 cross-peak (2.45 Å) as a reference. For dihedral angles, stem residues were given A-form RNA restraints ($\alpha = -62 \pm 30$, $\beta = -179 \pm 30$, $\gamma = 47 \pm 30$, $\zeta = -73 \pm 30$, $\nu = 37 \pm 30$), whereas loop and junction residues were left unrestrained. The ribose sugar puckers and δ values were determined according to the intensities of H1′-H2′ cross-peaks in a double-quantum-filtered (DQF)-COSY: strong (C2′-endo, $\delta = 145° \pm 30°$), intermediate (C2′ endo, $\delta = 120° \pm 30°$), or no cross-peak (C3′ endo, $\delta = 82° \pm 30°$). There was no NOE evidence for any *syn* bases, thus χ angles were restrained to *anti* (-160 ± 30). Hydrogen bond distance restraints were used to restrain the 17 Watson–Crick base pairs in the stems and the 6 Hoogsteen base triple interactions consistent with the experimental data (from J*NN*-HNN-COSY and 2D H_2O NOESY). There are two restraints per hydrogen bond.

An initial 200 structures were calculated with X-PLOR-NIH 2.9.8, starting from an extended, unfolded RNA conformation using 1,027 NOE and 549 dihedral angle restraints, following standard X-PLOR protocols (8, 15). The 100 structures with the lowest energy were selected to undergo a second round of NOE structure refinement with a lower starting temperature (1,000K) and more cooling steps (40,000). The 100 structures were then further refined with 98 RDCs, which were normalized to a C-H bond length of 1.0 Å. A grid search produced the optimal values for the magnitude (Da) and the asymmetry of the alignment tensor (Dr), which were Da = 40 and Dr = 0.15 (16). Additional experimental restraints and structural statistics for the lowest 20 energy structures are in Table S1. The structures were viewed and analyzed with MOLMOL and PYMOL.

Determination of Divalent Metal Ion Binding Sites by Mn^{+2} Titration. Mn^{+2} was titrated into the 1-mM RNA sample by stepwise addition of 3 mM $MnCl_2$ to final concentrations of 10–100 μM Mn^{+2}. Spectra were taken at each titration point. Titration was stopped when no further changes in the spectra were seen. ^{1}H-^{13}C HSQC spectra were taken at 293K on an 800-MHz NMR spectrometer.

1. Guillerez J, Lopez PJ, Proux F, Launay H, Dreyfus M (2005) A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proc Natl Acad Sci USA* 102(17):5958–5963.
2. Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res* 20(17):4515–4523.
3. Dieckmann T, Feigon J (1997) Assignment methodology for larger RNA oligonucleotides: Application to an ATP-binding RNA aptamer. *J Biomol NMR* 9(3):259–272.
4. Cromsigt JA, Hilbers CW, Wijmenga SS (2001) Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation. *J Biomol NMR* 21(1):11–29.
5. Peterson RD, Theimer CA, Wu H, Feigon J (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J Biomol NMR* 28(1):59–67.
6. Wu H, et al. (2001) A novel family of RNA tetraloop structure forms the recognition site for *Saccharomyces cerevisiae* RNase III. *EMBO J* 20(24):7240–7249.
7. Theimer CA, Finger LD, Trantirek L, Feigon J (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc Natl Acad Sci USA* 100(2):449–454.
8. Kim NK, et al. (2008) Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J Mol Biol* 384(5):1249–1261.
9. Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *J Am Chem Soc* 120:8293–8297.
10. Dingley AJ, Nisius L, Cordier F, Grzesiek S (2008) Direct detection of N-H[...]N hydrogen bonds in biomolecules by NMR spectroscopy. *Nat Protoc* 3(2):242–248.
11. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5(12): 1065–1074.
12. Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. *J Am Chem Soc* 123(7): 1416–1424.
13. Saupe A (1968) Recent results in the field of liquid crystals. *Angew Chem Int Ed Engl* 7: 97–112.
14. Theimer CA, Blois CA, Feigon J (2005) Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* 17 (5):671–682.
15. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73.
16. de Alba E, Tjandra N (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. *Prog Nucl Magn Reson Spectrosc* 40:175–197.

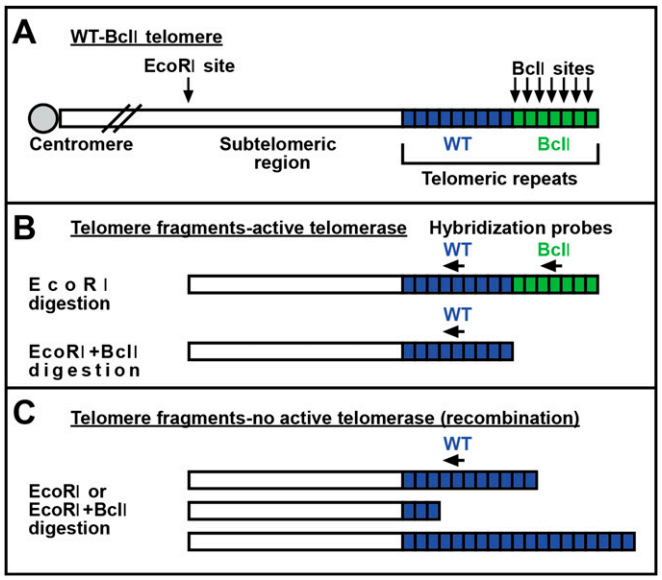


Fig. S1. A BclI silent mutation is used to mark the nascent products of an investigated telomerase. (*A*) A schematic representation showing a telomere containing WT (blue) and BclI (green) repeats. Telomeric restriction fragments are shown for WT-BclI (*B*) and telomerase null (*C*) strains, digested with either EcoRI or EcoRI+BclI. BclI-specific and WT hybridization probes for Southern analysis are shown as arrows above the fragments.

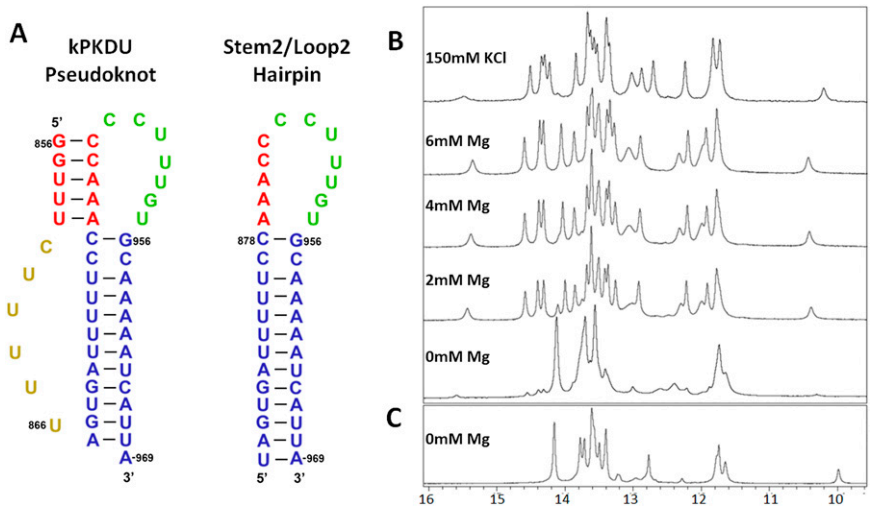


Fig. S2. (*A*) kPKDU and stem2–loop 2 hairpin constructs studied by NMR. (*B*) 1D imino proton spectra of kPKDU in 10 mM Tris·D (pH 6.3) titrated with Mg Cl_2 or KCl to the concentrations indicated. (*C*) 1D imino proton spectrum of the stem2–loop 2 hairpin in 10 mM Tris·D (pH 6.3). Spectra were taken at 283K on a 600-MHz NMR spectrometer.

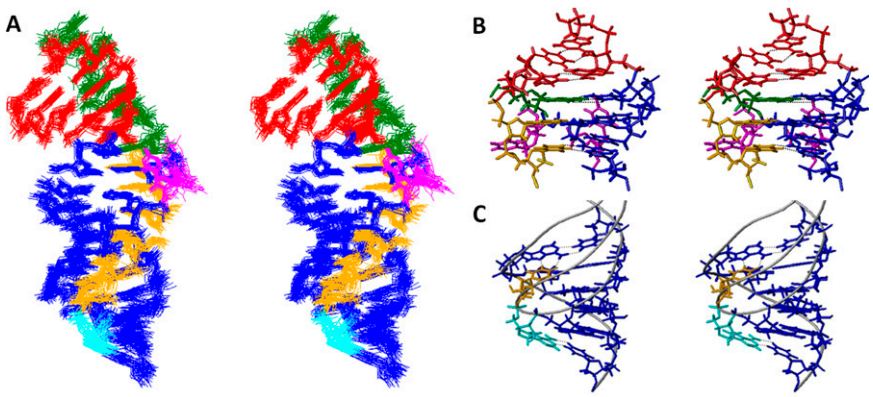


Fig. S3. Stereo views of kPKDU showing (*A*) superposition of 20 lowest energy structures, and lowest energy structure of (*B*) junction and (*C*) loop1/stem 2 turn.

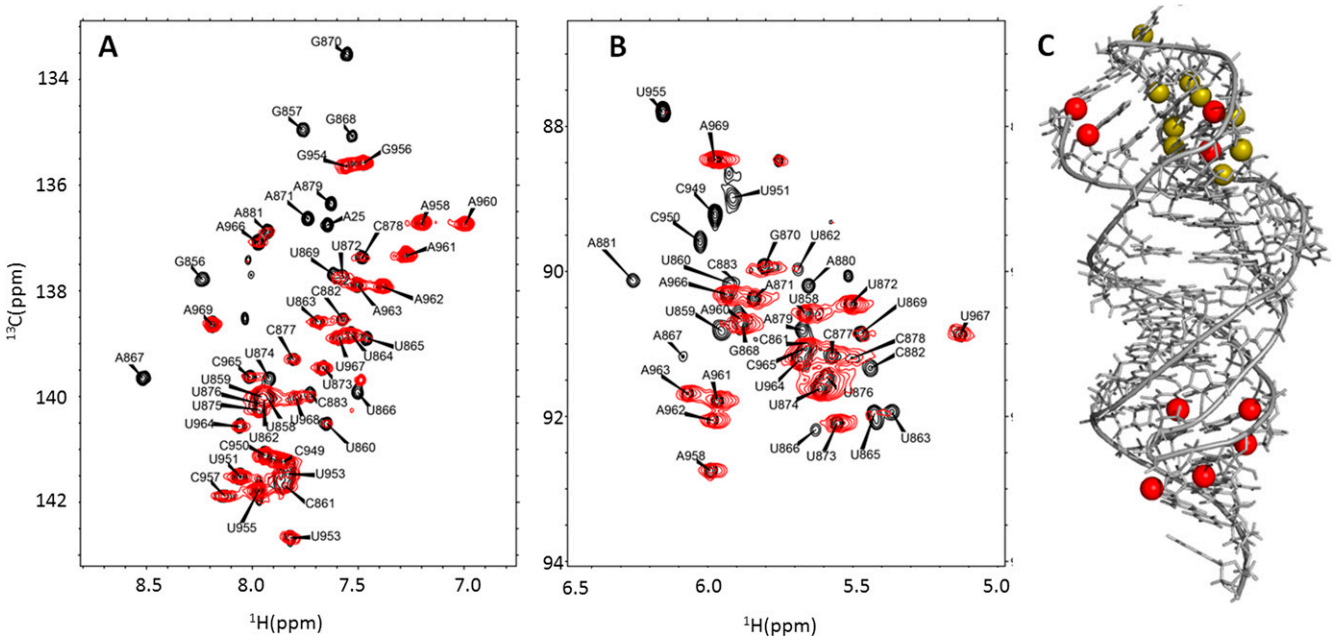


Fig. S4. ^{1}H-^{13}C HSQC spectra of kPKDU with Mn^{2+} (red) and without Mn^{2+} (black). Resonances of residues close to bound Mn disappear owing to paramagnetic relaxation. (*A*) C6H6/C8H8 HSQC overlay (red = 100 μM Mn^{2+}). (*B*) C1′H1′ HSQC overlay (red = 40 μM Mn^{2+}). (*C*) Atoms that disappear owing to broadening are indicated by red (C6H6/C8H8) or gold (C1′H1′) spheres on the pseudoknot tertiary structure. Spectra were taken at 298K on an 800-MHz NMR spectrometer.

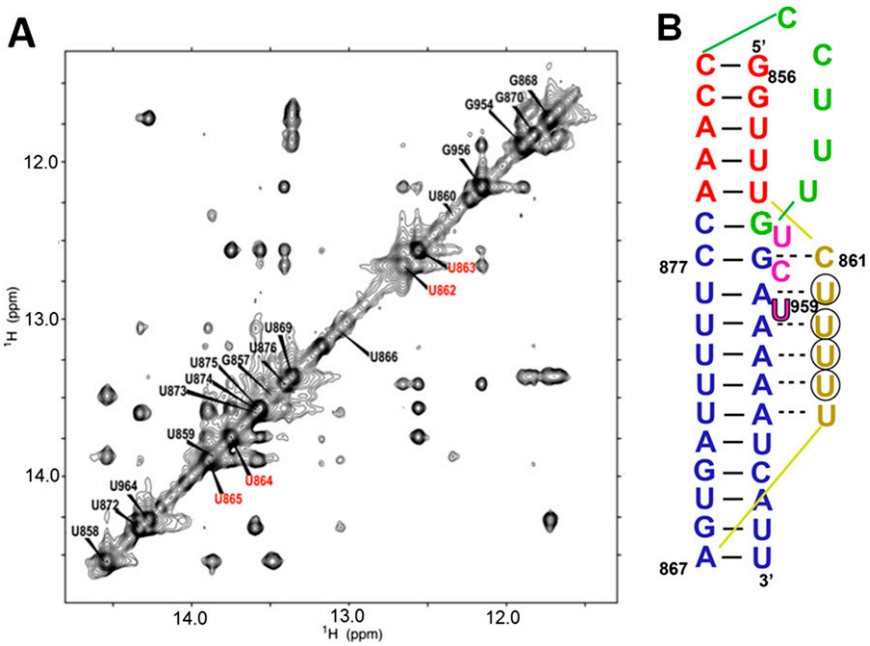


Fig. S5. (*A*) H_2O NOESY of imino region of kPKWT. Residues colored in red have chemical shift changes >0.1 ppm. Secondary structure is maintained compared with kPKDU as seen by similar cosspeak patterns and NOE connectivities. (*B*) kPKWT secondary structure schematic with residues circled that have chemical shift changes >0.1 ppm. Spectra were taken at 283K on an 800-MHz NMR spectrometer.

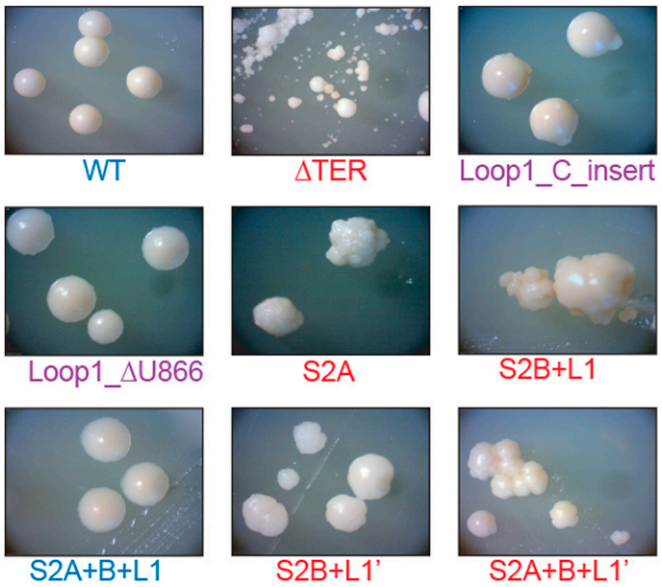


Fig. S6. Colony morphology of WT, ΔTER, and mutant pseudoknot strains of *K. lactis*. Phenontypes categorized as normal (blue), mild (purple), or rough (red).

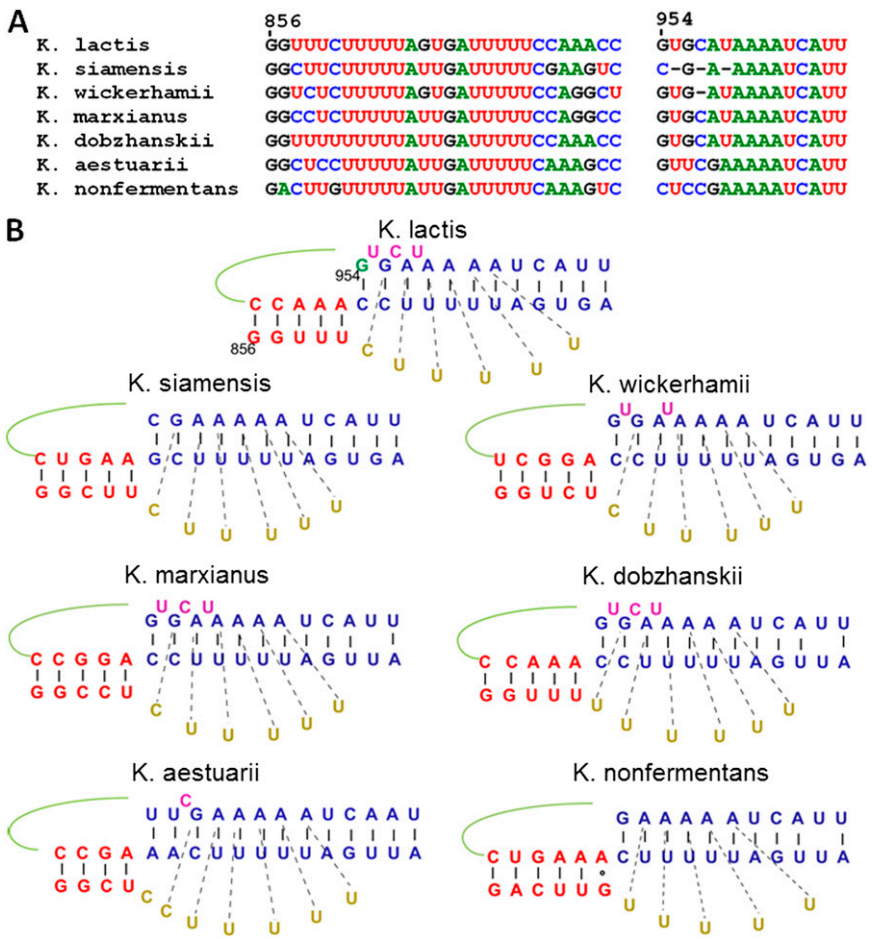


Fig. S7. (*A*) Sequence alignment of the pseudoknot domain of *Kluyveromyces* species telomerase RNA. (*B*) Secondary structure prediction of pseudoknots based on kPKDU structure.

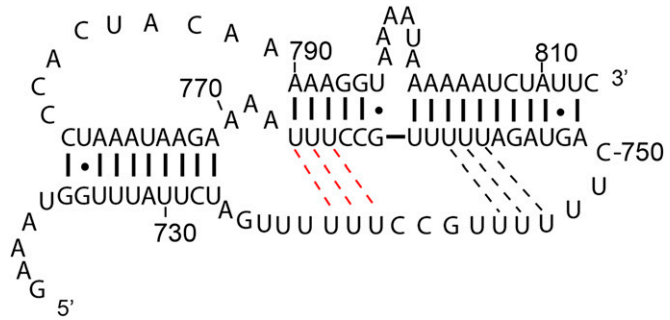


Fig. S8. Sequence and secondary structure of the *S. cerevisiae* telomerase RNA pseudoknot based on reference 1. The three U-A-U triples predicted by reference 2 are shown by black dashed lines, and additional predicted U-A-U triples are shown by red dashed line.

1. Liu F, Kim Y, Cruickshank C, Theimer CA (2012) Thermodynamic characterization of the *Saccharomyces cerevisiae* telomerase RNA pseudoknot domain in vitro. *RNA* 18(5):973–991.
2. Qiao F, Cech TR (2008) Triple-helix structure in telomerase RNA contributes to catalysis. *Nat Struct Mol Biol* 15(6):634–640.

Table S1. Restraints and structure statistics for kPKDU

Parameter	RNA
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	1,027
Intraresidue	380
Interresidue	647
Sequential ($\|i - j\| = 1$)	342
Medium-range ($\|i - j\| < 4$)	18
Long-range ($\|i - j\| > 5$)	287
Hydrogen bond restraints*	90
Total dihedral angle restraints	254
Sugar pucker	206
χ	48
RDCs	98
Structure statistics	
Violations (mean ± SD)	
Distance constraints (Å)	0.023 ± 0.001
Dihedral angle constraints (°)	0.11 ± 0.02
Maximum dihedral angle violation (°)	2.028
Maximum distance constraint violation (Å)	0.273
Deviations from idealized geometry	
Bond lengths (Å)	0.0044 ± 0.0001
Bond angles (°)	1.077 ± 0.005
Impropers (°)	0.526 ± 0.007
Average pairwise rmsd† (Å)	
Heavy	1.13 ± 0.24
Backbone	1.15 ± 0.25

*Two per hydrogen bond.
†Pairwise rmsd was calculated for 20 lowest energy structures.

CHAPTER 3

Structure and folding of the *Tetrahymena* telomerase RNA pseudoknot

Introduction

Telomerase is a ribonucleoprotein (RNP) complex composed of the catalytic telomerase reverse transcriptase (TERT), telomerase RNA (TER) and a number of species specific proteins (*1, 2*). Telomerase maintains the telomeric DNA at the 3′ ends of linear chromosomes by adding *de novo* telomeric DNA repeats (*3, 4*). Telomeres, composed of telomeric DNA and bound proteins, counteract DNA damage due to incomplete replication, degradation, chromosomal fusion and illicit DNA damage repair (*5-7*). TERT is generally well conserved ranging from ~900-1100 amino acids (aa) among the commonly studied vertebrates, yeast, and ciliates. It contains four conserved domains: the TERT essential N-terminus (TEN), RNA binding domain (RBD), reverse transcriptase (RT) domain, and C-terminal extension (CTE) (*8*). TER is more divergent, ranging in size from ~150 nucleotides (nt) in ciliates to over 3000 nt in yeasts (*9*). TER has a number of conserved elements, which are the template, pseudoknot, template boundary element (TBE), and stem terminus element (STE) (*1, 10*). The central region of TER, termed the core domain, contains the template that is used to copy the telomeric sequence and an adjacent pseudoknot (PK) that has been proposed to assist in template positioning within the active site (*11-13*). The TBE, also part of the core domain, is typically a stem or hairpin upstream of the template that helps prevent copying of nontemplate residues (*14*). Ciliates also have a template recognition element (TRE) 3′ of the template, which TERT uses to direct efficient use of the template for reverse transcription (*65*). The STE is distal from the core domain and stimulates telomerase activity typically through TERT (RBD) interactions (*10*). Along with TERT, the core domain and STE of TER are the minimal components necessary to reconstitute telomerase activity *in vitro* (*1*).

The structures of TER domains have been extensively studied in yeast, human and ciliates (*15-24*). The NMR solution structures of the human and yeast *K. lactis* TER pseudoknots revealed a number of conserved features (*17, 25*). Despite a divergent sequence and secondary structure between the two pseudoknots, the three-dimensional structures are very similar, indicating a conserved function. Both pseudoknots contain an extended triple helix where the

loop residues form base triples with the Watson-Crick (WC) paired stems. The base triples were shown to be essential for catalytic activity *in vitro* and *in vivo* (*13, 17, 25, 26*). Additional biochemical studies have also linked the pseudoknot to telomerase assembly and telomere repeat addition processivity (RAP) (*27, 28*).

A smaller pseudoknot has been predicted to form in ciliate TERs (*29*). The telomere-rich ciliate *Tetrahymena* has served as a model organism from which telomeres and telomerase was discovered, leading to a Nobel Prize in 2009 (*30*). Ciliate TERs are generally composed of four base paired regions (numbered I-IV and depicted in Figure 1) where Stems IIIa and IIIb form the pseudoknot (*29, 31, 32*). NMR structures of *Tetrahymena* stem-loop (SL) II and SLIV have been determined (*18, 21, 22, 33*). Stem IV is the ciliate STE, and was crystallized in the presence of telomerase accessory protein p65 (*37*). p65 binds Stem IV inducing a 105° bend at the GA bulge of the stem. This conformational change promotes hierarchical assembly of telomerase (*34-37*). Despite early discovery of the *Tetrahymena* pseudoknot in 1991 and its small size, its structure remained elusive presumably due to its "plasticity" (*38*). This refers to the instability and conformational dynamics of the pseudoknot, likely due its short Stem A and Loop A (Fig. 1). Interestingly, the short length of Stem A is conserved amongst ciliates (3-4 bp) suggesting a possible function for pseudoknot conformational flexibility (*39*).

A number of studies have examined pseudoknot formation in the *Tetrahymena* TER. Early studies using chemical probing showed that the isolated *Tetrahymena* pseudoknot (tetPK) can form a stable pseudoknot, while tetPK does not form in the context of full length TER (*38*). The authors concluded that in full length TER the pseudoknot region is in equilibrium with alternate conformations. More recent single-molecule FRET experiments verified that tetPK forms in isolation but not in full length TER (*40*). Furthermore, unexpected FRET values for the free TER suggested that the pseudoknot residues were involved in competing interactions with other regions of the RNA (*40*). Based on SHAPE data, a model for these competing interactions was proposed where residues from the pseudoknot region could base pair with residues from the Template-TRE (*41*). For example, they predicted that 81-UAGGUU from the pseudoknot stem

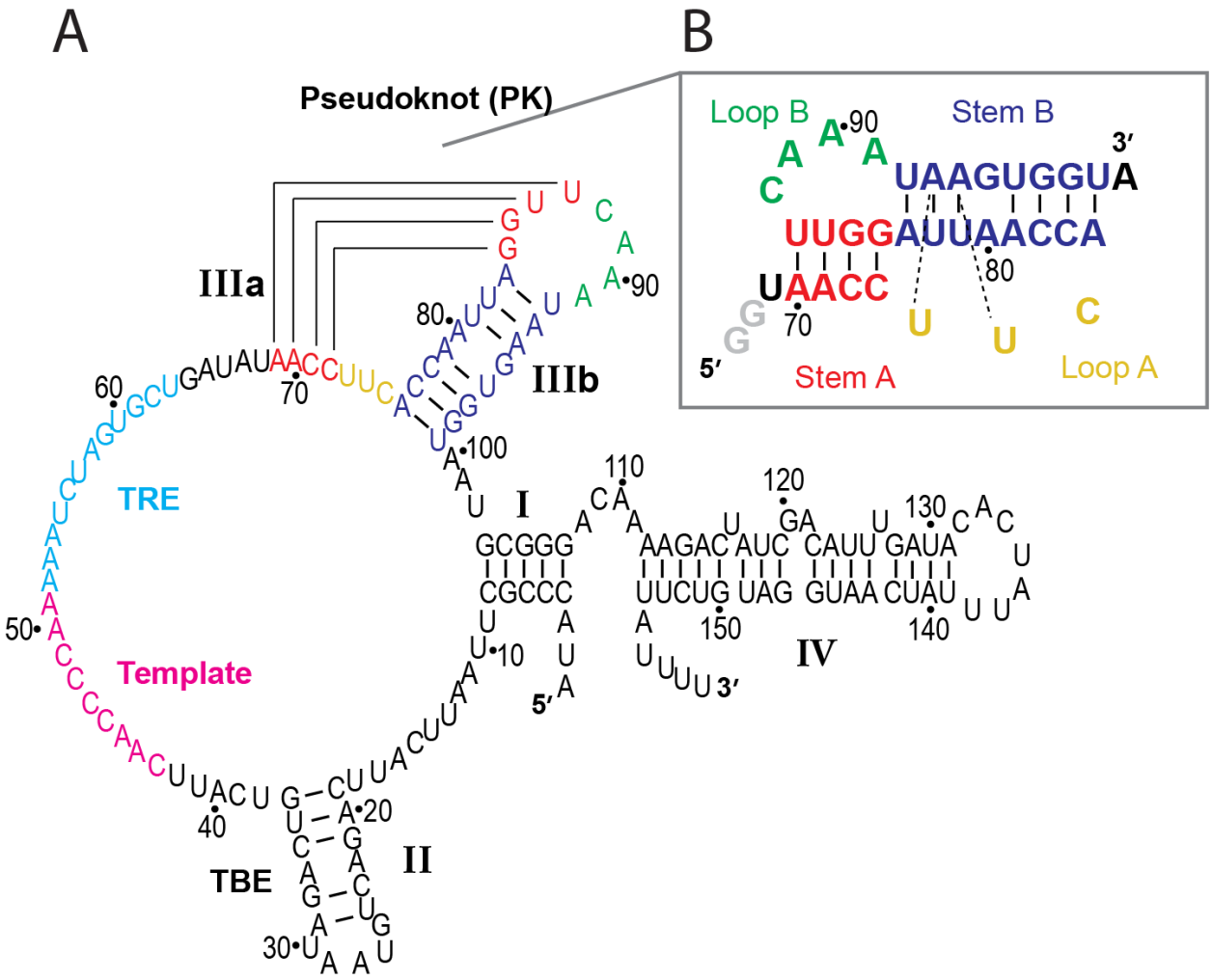


Figure 1: *Tetrahymena thermophila* telomerase RNA

(A) *Tetrahymena* telomerase RNA, with base paired helices labeled I-IV. Template (magenta), template boundary element (TBE), template recognition (TRE, cyan), and pseudoknot (PK) are also labeled. **(B)** *Tetrahymena* TER pseudoknot NMR construct. The secondary elements are colored as Stem A (red), Loop A (gold), Stem B (blue), and Loop B (green). Two G's (gray) were added to enhance *in vitro* transcription.

IIIb would pair with 53-AAUCUA from the TRE. Importantly, both SHAPE and FRET studies show that while tetPK does not form in free TER, it does form in the presence of TERT.

To further examine the folding of the *Tetrahymena* pseudoknot, in this work we used NMR to study *Tetrahymena* telomerase RNA structure. We report the NMR solution structure of tetPK. TetPK forms a pseudoknot with stacked stems and the loops interacting in the respective major and minor grooves. While there are a number of similarities to the human and yeast pseudoknots, the tetPK tertiary structure is significantly different with a more compact shape. The tertiary interactions include a novel A-G-C base triple that contributes to pseudoknot stability and is important for telomerase activity. NMR data confirms that tetPK does not form in full length TER. However, instead of pseudoknot residues base pairing with the Template-TRE as predicted by SHAPE data, we find that the Template-TRE forms a separate hairpin and Stem I becomes extended by three base pairs into the bottom of the pseudoknot stem IIIb. These competing interactions explain why the pseudoknot does not form in the full length TER core domain in the absence of TERT.

Materials and Methods

RNA preparation

RNA samples for NMR studies were synthesized by *in vitro* transcription, using T7 polymerase, from a synthetic DNA template as previously described (*42, 43*). Briefly, purified T7 (P266L mutant) polymerase (*44*) is added to a reaction containing 40 mM $MgCl_2$, 6 mM of each NTP, 1 μM DNA template, and 1x buffer containing 10 mM Tris-HCl pH 8, 1 mM spermidine, 2.5 mM DTT and .01% Triton. Uniformly ^{13}C,^{15}N labeled rNTPs were appropriately substituted in the reaction to make ^{13}C,^{15}N (A-U), ^{13}C,^{15}N (G-C), and ^{13}C,^{15}N (A-U-G-C) labeled samples. The template contains a minimal consensus hammerhead ribozyme sequence at the 3′ end to allow self-cleavage of each RNA molecule at a precise position (*45, 46*). The reaction was incubated at 37°C for 6-8hrs to allow for transcription and complete cleavage by the hammerhead ribozyme. The RNA was subsequently purified by 15% denaturing polyacrylamide

gel and electroelution (*47*). The RNA was washed, using Amicon centrifugal filters, with high salt (1.5 M KCl), 3 times with water, and then NMR buffer (10 mM $NaPO_4$ pH 6.3, 50 mM KCl). Under dilute conditions (10-50 μM) the RNA solution was heated at 95°C for 4 minutes and then snap cooled on ice. The RNA was then concentrated to ~1 mM for NMR studies.

RNAs for Core domain, PK-StemI-StemII, PK-StemI-StemIV, Template-TRE-PK, and StemI-Template-TRE-PK (diagrammed in Fig. 4) were cloned into pRSFDuet-1 (Novagen) plasmids with an upstream T7 promoter and downstream hammerhead ribozyme. The plasmids were amplified and purified from *E. coli* by Plasmid DNA Maxiprep kit (Invitrogen). The vectors were linearized at the template 3′ end by restriction enzyme digest (XhoI) and used as the template in the reaction described above (at 50 μg/mL), following the same transcription and purification procedure. For native purification of the Core domain, we loaded the RNA sample onto a DEAE (diethylaminoethyl) column. The column was then washed with 5 column volumes of wash buffer (10 mM sodium phosphate, pH 7.0, 200 mM KCl), then eluted with elution buffer (wash buffer with 1 M KCl). The RNA was then purified by size exclusion (Superdex200) with NMR buffer as the running buffer, followed by concentration to ~0.5 mM.

NMR spectroscopy

NMR spectra were recorded on Bruker DRX 500 MHz and 600 MHz and Avance 800MHz spectrometers equipped with HCN cryoprobes. Exchangeable proton resonances were assigned with samples in 90%H_2O/10%D_2O, using 2D NOESY, ^{1}H-^{15}N HSQC, and JNN-COSY spectra at 283K. Non-exchangeable proton resonances were assigned with samples in 100% D_2O, using 2D NOESY, 2D TOCSY, ^{1}H-^{13}C HSQC, 2D HCCH-COSY, 3D HCCH-TOCSY, and a suite of filtered/edited NOESY (F1fF2f, F2f, F1fF2e, F1eF2e) experiments at 283K (*43, 48-50*). Protonated ^{13}C and ^{15}N resonances were assigned with the HSQC experiments. The proton assignment strategy used sequential NOE connectivity as previously described (*16, 23, 50*). NOESY experiments were acquired with 100, 200, and 300 ms mixing times to determine which mixing time best approximated a linear relationship between NOE crosspeak volume and inter-

proton distance. Hydrogen bonds for WC and Hoogsteen base pairs were confirmed by JNN-COSY (*51, 52*). The imino proton resonances of the terminal base pairs (A69-U87, A76-U99) were not detected due to rapid exchange with water, but these base pairs were confirmed by NOEs (in D_2O-NOESY spectra). Non WC base pairs (A80-G95, A91-C72, A90-C71) were also confirmed by indicative NOEs (in D_2O-NOESY spectra).

RDCs were measured for C-H (^{1}DC1′H1′, ^{1}DC2H2, ^{1}DC5H5, ^{1}DC6H6, ^{1}DC8H8) and N-H (^{1}DN1H1, ^{1}DN3H3) using the ^{13}C,^{15}N (A-U-G-C) labeled sample on 800MHz spectrometer at 283K (*53*). The spectra were processed with Bruker TOPSPIN and analyzed with NMRDraw. A total of 69 C-H and 10 N-H RDCs were determined by measuring the difference in J-coupling in the absence and presence of 15mg/mL Pf1 phage.

Xplor structure calculations

For structure calculation, the NOEs were integrated to generate distance restraints using the pyrimidine H5-H6 crosspeak as an internal standard (2.45Å), using 200 ms mixing time NOESY spectra. NOEs were categorized as very strong (2.5Å), strong (3.5 Å), moderate (4.5 Å), weak (5.5 Å) or very weak (6.5 Å), with a range of ±1 Å (except "very strong" was limited to van der Waals lower bound, 1.8 Å). Loose A-form dihedral angle restraints were used for the stem residues (α = -62° ± 30°, β = -179° ± 30°, γ = 47° ± 30°, ζ = -73° ± 30°, ν = 37° ± 30°). The ribose sugar pucker, correlated to δ angle, was determined based on the H1′-H2′ crosspeak intensity in the DQF-COSY: strong (C2′-endo, δ= 145° ±30°), intermediate (C2′ endo, δ = 120° ± 30°), or no cross peak (C3′ endo, δ = 82° ± 30°). Syn/anti configuration is correlated to χ, and determined based on characteristic NOE pattern, where strong H8-H1′ crosspeak intensity indicates a syn conformation. Only G95 was syn (χ = 25° ± 30°), while all other residues were set as anti (χ = -160° ± 30°). Hydrogen bond restraints were used to restrain all experimentally determined base pairs, with two distance restraints per hydrogen bond.

Xplor-NIH 2.9.8 was used to calculate an initial 100 structures starting from a single strand, extended RNA molecule, using NOE, hydrogen bond, weak planarity (weight = 300.0 for

individual base and weight = 6.0 for base pairs), and dihedral restraints (*16, 54*). This was followed by a second round of NOE refinement with a lower starting temperature (1000K) and more cooling steps (40,000). In the last step of refinement, RDC refinement was done with 79 RDCs, in which a grid search produced optimal values for the axial (Da) and rhombic (Dr) components of the alignment tensor: Da = -39.0, Dr = .13 (*55*). The structural statistics for the lowest 10 (out of 100) energy structures are given in Table 1. The structures were viewed and analyzed with MOLMOL and PYMOL.

Telomerase activity assay

Telomerase activity was measured by primer extension assay as previously described (*56*). TERT was *in vitro* translated in rabbit reticulocyte lysate (RRL) using the TNT transcription/translation kit (Promega). The 10 μL translation reaction contained 0.4 μL of PCR enhancer (0.5 M KCl, 12.5 mM $Mg(OAc)_2$), 0.2 μL of 1 mM methionine, 8 μL RRL mix, and 150 ng TERT DNA plasmid (pCITE-TERT), and was incubated at 30°C for 1 hr. TER variants were encoded on a plasmid as a DNA template (with hammerhead ribozyme) for in vitro T7 transcription. The RNAs were transcribed and purified as described above, except the RNA was exchanged into water (instead of buffer). 0.5-1.0 μM of the purified RNA was added to the RRL translated TERT (1 μL volume RNA per 10 μL RRL reaction) and incubated at 30°C for 30 min for RNP reconstitution. p65 was purified as described (*57*), and was added to 1 μM during RNP reconstitution when necessary.

A 20 μL telomerase primer extension reaction contains 1x buffer (50 mM Tris-HCl pH 7.0 or 8.3, 1.25 mM $MgCl_2$, 1 mM TCEP), 1 μM primer [$(GT_2G_3)_3$], 100 μM TTP, 9 μM nonradioactive dGTP, 0.4 μL of [α-^{32}P]dGTP at 3000 Ci/mmol, and 6 μL of the reconstituted RNP. The reaction was incubated at 30°C for 60 minutes and terminated by addition of quench buffer (1 mM Tris-HCl pH 8.0, 0.5 M EDTA) containing 15-mer [α-^{32}P]5′-end labelled RNA recovery control (RC). The nucleic acid products and control were recovered by phenol/chloroform extraction and ethanol precipitation, and loaded on a 10% polyacrylamide

sequencing gel (19:1 polyacrylamide, 7 M urea, 1x TBE). The gel was electrophoresed at 50W for 1 hr, dried and exposed overnight to a phosphor imaging screen. The screen was scanned and analyzed with QuantumOne Software. The relative activity was determined by normalizing the integrated density of each lane relative to the RC, and comparing to the WT (as 100%).

Results

Folding of tetPK

For structural studies, the tetPK RNA construct was designed from the WT TER sequence (nt 69-100) with an additional two G's added at the 5′ end to enhance *in vitro* transcription (Fig. 1B). The *in vitro* transcribed RNA includes a 3′ hammerhead ribozyme sequence which self-cleaves the RNA product precisely at A100, ensuring 3′ end homogeneity. 2D imino NOESY spectra indicated that the pseudoknot was similarly folded in the presence of 1-5 mM Mg^{2+} and/or 50-200 mM KCl, exhibiting the same NOE pattern (data not shown). However, 50 mM KCl was used for further NMR studies due to the better spectra quality. The NMR buffer is 10 mM sodium phosphate pH 6.3 with 50 mM KCl.

TetPK has low stability compared to the TER pseudoknots of human and *K. lactis*, likely due to its short Stem A and Loop A. 1D imino and 2D TOCSY (H5-H6) data show that the pseudoknot begins to unfold at temperatures greater than 15°C or if pH is increased to 8.0. The destabilized pseudoknot is in equilibrium with alternate conformations, which can include the hairpins of each stem and single stranded RNA (*23*). The alternate conformations are indicated by the appearance of additional TOCSY (H5-H6) crosspeaks which are of similar intensity to those of the pseudoknot at increased temperature or pH (Fig. S1). The crosspeaks arise from separate unique structures in slow exchange with the pseudoknot. As temperature and pH increase, the alternate conformations increase in population. Additional Mg^{2+} or KCl does not have a significant effect on this equilibrium (data not shown). For these reasons, the tetPK structure was solved at 10°C and pH 6.3 where the pseudoknot conformation predominates.

The base pairing of tetPK was determined by analyzing 2D imino NOESY spectra in conjunction with A,U- and G,C-JNN COSY spectra (*51, 52*). The stems formed the predicted base pairs (shown in Figure 1B), and showed the expected NOE connectivities, including an NOE between stems (G84H1 to U92H3) indicating stacking of the two stems. Of the two predicted U-A-U base triples, only one (U81-A94-U74) could be directly confirmed by JNN-COSY. Surprisingly, a unique A-G-C base triple was identified (A80-G95-C75) (Figure 2E). The A80N1 is protonated so that it can hydrogen bond with the G95N7, since G95 is in the syn conformation. Although the A80 imino resonance was not observed (even at pH 5.0), there is substantial evidence that the A80N1 is protonated. N1 protonation causes the associated amino protons to shift significantly downfield and split (*58*). This is observed with the A80 amino protons of tetPK, and is similar to the aminos of the protonated cytosine of the *K. lactis* telomerase RNA pseudoknot's C-G-C^+ base triple (*25, 59*) (Fig. S2B). In addition, A80C2 is upfield shifted, another characteristic of adenines protonated at the N1 position (Fig. S2A) (*60*). G95 was identified as syn by the strong NOE between its H8 and H1′. In addition, the NOE between A80H2 and G95H8 is a strong indication of the A^+-synG base pair. This type of A^+-synG pair was observed in the crystal structure of an RNA 16-mer duplex with G-A mismatches (*61*). In tetPK the syn G is positioned with its Watson-Crick face in the major groove and forms a canonical WC base pair with the loop residue C75.

TetPK solution structure

The solution structure of tetPK was calculated by Xplor simulated annealing with 414 NOEs, 171 dihedral angles, and 79 RDCs as restraints. The structural statistics for the 10 lowest energy structures are in Table 1. The structure is well-defined with an RMSD of 0.8Å (Fig 2B). The two stems form a quasi-continuous A-form helix, stacked on top of each other without any significant bend. Loops A and B bind in the major groove (of Stem B) and minor groove (of Stem A) respectively. The pseudoknot is compact, with a major groove width of 11.0 Å (defined as the phosphate distance between i and i+6 cross-strand residues). Loop A and Stem B interact

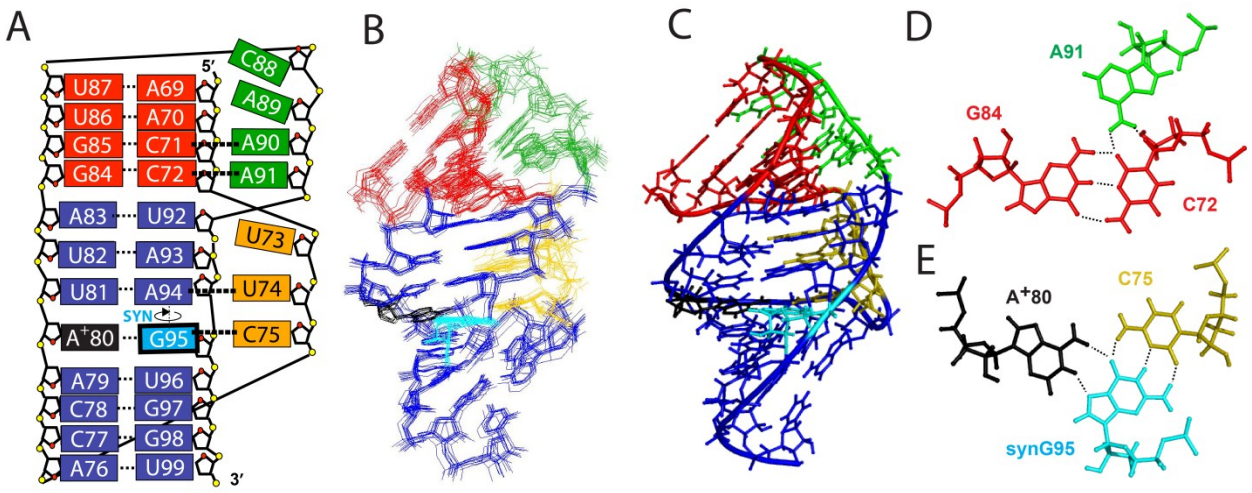


Figure 2: Structure of tetPK

(A) Diagram of determined secondary structure of tetPK. Solid lines represent backbone connectivity, and dotted lines represent hydrogen bonds. **(B)** Family of lowest 10 energy NMR structures of tetPK. **(C)** Lowest energy NMR structure of tetPK. **(D)** G84-C72-A91 minor groove base triple and **(E)** A80-G95-C75 major groove base triple, with hydrogen bonds shown as dotted lines. Pseudoknot elements are colored as in Figure 1, except A80 is black and G95 is cyan, for clarification when viewing structures.

Parameter	RNA
NMR distance and dihedral contraints	
Distance constraints	
Total NOE	414
Intraresidue	103
Interresidue	311
Sequential (|i-j|) = 1	176
Long range (|i-j|) > 1	135
Hydrogen bond restraints*	82
Total dihedral angle restraints	171
Sugar pucker	140
χ	31
RDCs	79
Structure statistics	
Violations (mean ± SD)	
Distance constrants (Å)	0.054 ± .001
Dihedral angle constraints (°)	.020 ± .02
Maximum dihedral angle violation (°)	2.034
Maximum distance violation (°)	0.292
Deviations from idealized geometry	
Bond lengths (Å)	.0057 ± .0001
Bond angles (°)	0.95 ± .02
Impropers (°)	0.61 ± .04
Average pairwise rmsd (Å) **	
Heavy	0.83 ± 0.23
Backbone	0.84 ± 0.24

Table 1: Restraint and structure statistics for tetPK

* Two hydrogen bond restraints were used for each hydrogen bond

** Pairwise RMSD was calculated for the lowest 10 energy structures

to form a small triplex, composed of a U-A-U and A-G-C base triple. While the second predicted U-A-U triple (U82-A93-U73) was not detected, a majority of the lowest energy structures have U73 positioned near the U82-A93 stem base pair in a geometry consistent with Hoogsteen base pairing. However the U73 imino resonance was not observed in any NMR spectra. In addition, the NOE pattern expected for this base triple, in which U73 would have crosspeaks to U92 and A93 (i.e. U73H1′ to U92H8, -H1′, -H2′ and U73H1′ to A93H8), was not present (*17*). Therefore we conclude that this base triple does not form (in the absence of other telomerase components).

Two minor groove interactions are formed between the adenine-rich Loop B and Stem A of tetPK, G84-C72-A91 and G85-C71-A90. The adenine-rich Loop B is common in vertebrate TER pseudoknots, as well as in ribosomal frameshifting viral pseudoknots (*62, 63*). In these RNAs, the loop adenines form minor groove triples involving 2′-OH and base protons (*63*). In tetPK, A90 and A91 are stacked on top of each other forming identical base triples with C71 and C72 respectively. The adenine amino protons form two hydrogen bonds with the cytosine carbonyl and 2′-OH (shown in Fig. 2D). This interaction is consistent with *in vivo* footprinting data which indicated that the Loop B - CAAA residues were protected (from dimethyl sulfate modification), presumably by minor groove binding (*64*).The human TER pseudoknot forms similar minor groove triples with its Loop B adenine residues, further signifying the importance of these types of minor groove interactions on pseudoknot structure.

Pseudoknot folding affects telomerase activity

The loop-stem tertiary interactions of TER pseudoknots have previously been shown to be important for telomerase activity in humans and yeast (*in vitro* and *in vivo*) (*13, 17, 25, 26*). Therefore, we tested the potential importance of the A80-G95-C75 base triple on telomerase activity by primer extension assay. Individual nucleotide substitutions, C75A, C75U, G95U and A80C, were made in full length TER. TER variants were then reconstituted with *in vitro* translated TERT (from RRL) for telomerase activity assays. Since tetPK folding and stability is pH dependent, activity assays were done at pH 7.0, rather than standard pH 8.0 (*56, 57*). Of the

substitutions tested, C75A has the biggest effect on telomerase activity, exhibiting ~50% WT activity levels. G95U and C75U have 75% activity, while A80C has near WT levels of 90%. Increasing pH to 8.0 reduces activity by ~80% for each TER variant, including WT, which is consistent with the importance of A80N1 protonation. The activity was also assayed in the presence of holoenzyme assembly protein p65. p65 has been shown to rescue a number of TER mutations that affect assembly, including mutations within the pseudoknot region (*57*). Likewise, for all of the A-G-C base triple substitutions presented here, p65 increased the activity to near WT levels (Fig. 3B).

To investigate whether pseudoknot structure correlated with telomerase activity, the A-G-C base triple substitutions were made in the context of the isolated tetPK, and 1D imino NMR spectra were acquired to provide secondary structure information (Fig. 3C). The isolated Stem IIIb hairpin was also made for comparison. The 1D imino spectrum of the Stem IIIb hairpin shows that the A^+-synG base pair does not form in this construct (indicated by the lack of A^+ amino peaks) (Fig. 3C). The 1D imino spectrum of tetPK(C75A) is similar to that of the Stem IIIb hairpin, indicating that tetPK(C75A) does not form a pseudoknot, but a hairpin instead. This signifies that the G-C tertiary interaction of the A-G-C triple is necessary to stabilize the unusual A^+-synG base pair and Stem IIIa (Fig. 3C). The C75A substitution disrupts pseudoknot folding resulting in a hairpin which explains the low activity of this TER variant. The substitutions with higher activity, tetPK(C75U, G95U, and A80C), each form a pseudoknot with the presence of Stem IIIa and Stem IIIb as shown in the 1D spectra (Fig. 3C). Furthermore, tetPK(C75U and A80C) can potentially form A-G-U and C-G-C^+ base triples respectively. There was evidence of an A-G-U base triple in tetPK(C75U), which folds into a pseudoknot with similar imino resonances and NOE pattern to that of the WT (Fig. 3C, 2D data not shown). However no evidence was found for a C-G-C^+ base triple in tetPK(A80C); although this triple could be stabilized in the holoenzyme by TERT binding. Taken together, this data provides a direct correlation between proper pseudoknot folding (formation of Stem IIIa and Stem IIIb) and telomerase activity.

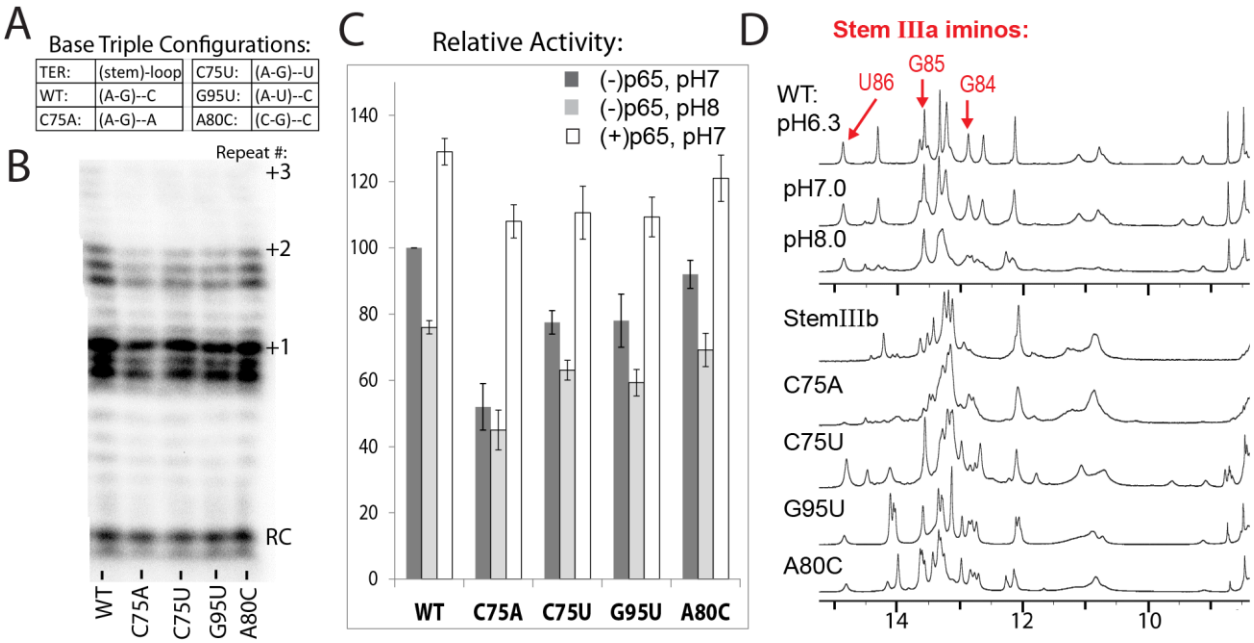


Figure 3: Telomerase activity and 1D imino spectra of pseudoknot variants

(A) TER substitutions with associated stem-loop base triple configuration. **(B)** Telomerase primer extension assay of WT and mutant TER constructs. The number of telomeric repeats and recovery control (RC) are indicated on the right. **(C)** Relative telomerase activity of WT and mutant TER constructs at pH 7 (dark gray), pH 8 (light gray), and with p65 at pH 7 (white). From the assay image (in **B**), the density of each lane was integrated and normalized to WT pH7 (as 100%). **(D)** (Top Panel) 1D imino spectra of WT tetPK at varying pH. (Bottom Panel) 1D imino spectra of Stem IIIb hairpin and tetPK variants. Imino resonances of Stem IIIa residues are shown in red. Stem IIIa forms in all constructs except Stem IIIb and C75A.

TER core domain folding

As discussed earlier, previous studies have shown that tetPK requires TERT to stably fold in the context of full length TER. Based on SHAPE data, a model of an alternate hairpin forming with the Template-TRE and pseudoknot residues was proposed for free TER (Fig. 5A) (*38, 40, 41*). Therefore we investigated the folding of free TER by NMR. Full length TER (nt 1-159, data not shown) and the Core domain (nt 1-107) had similar NOE patterns, so we focused on the Core domain due to better spectra quality. Stem I and Stem II RNA constructs were also made for comparison to the Core domain (Fig 3B). Comparison of the 2D imino NOESY spectra of Core domain, Stem I, and Stem II conclusively shows formation of Stem I and Stem II within the Core domain (Fig. 4A). In addition to iminos from Stem I and Stem II base pairs, there are additional imino resonances which must arise from base pairing from the Template, TRE and pseudoknot residues. These remaining iminos do not exhibit the tetPK NOE pattern (compare Fig. 4A green and Fig. 4C black), confirming that tetPK does not form within Core domain RNA, in agreement with previous studies (*38, 40, 41*). Therefore we expected that tetPK formation was disrupted due to base pairing competition with other TER regions, and that the remaining iminos could correspond to the SHAPE model of alternate base pairing. To test this hypothesis, we made the Template-TRE-PK (T-T-PK) RNA construct (nt 37-100). Comparison of the 2D imino NOESY spectra of tetPK and T-T-PK indicates that tetPK forms in the presence of the Template-TRE residues (Fig. 4C). In addition, the Template-TRE region forms a separate hairpin structure, indicated by the additional iminos in the T-T-PK spectra which have sequential connectivities (Fig. 4C). To confirm the formation of this hairpin, we made a construct of the isolated Template-TRE (nt 37-63) and examined its NMR spectra. 1D (and 2D, not shown) imino spectra show that the Template-TRE residues form a small hairpin, similar to that seen in T-T-PK (Fig. 4E). There are three sequential iminos, with a central G-U base pair. With this data we tentatively assigned the Template-TRE hairpin to 39-CAUU paired with 58-AGUG.

This Template-TRE hairpin also forms in the Core domain RNA with the same NOE pattern (Fig. 4A, cyan box). In addition, Stem I is extended by three base pairs as indicated by

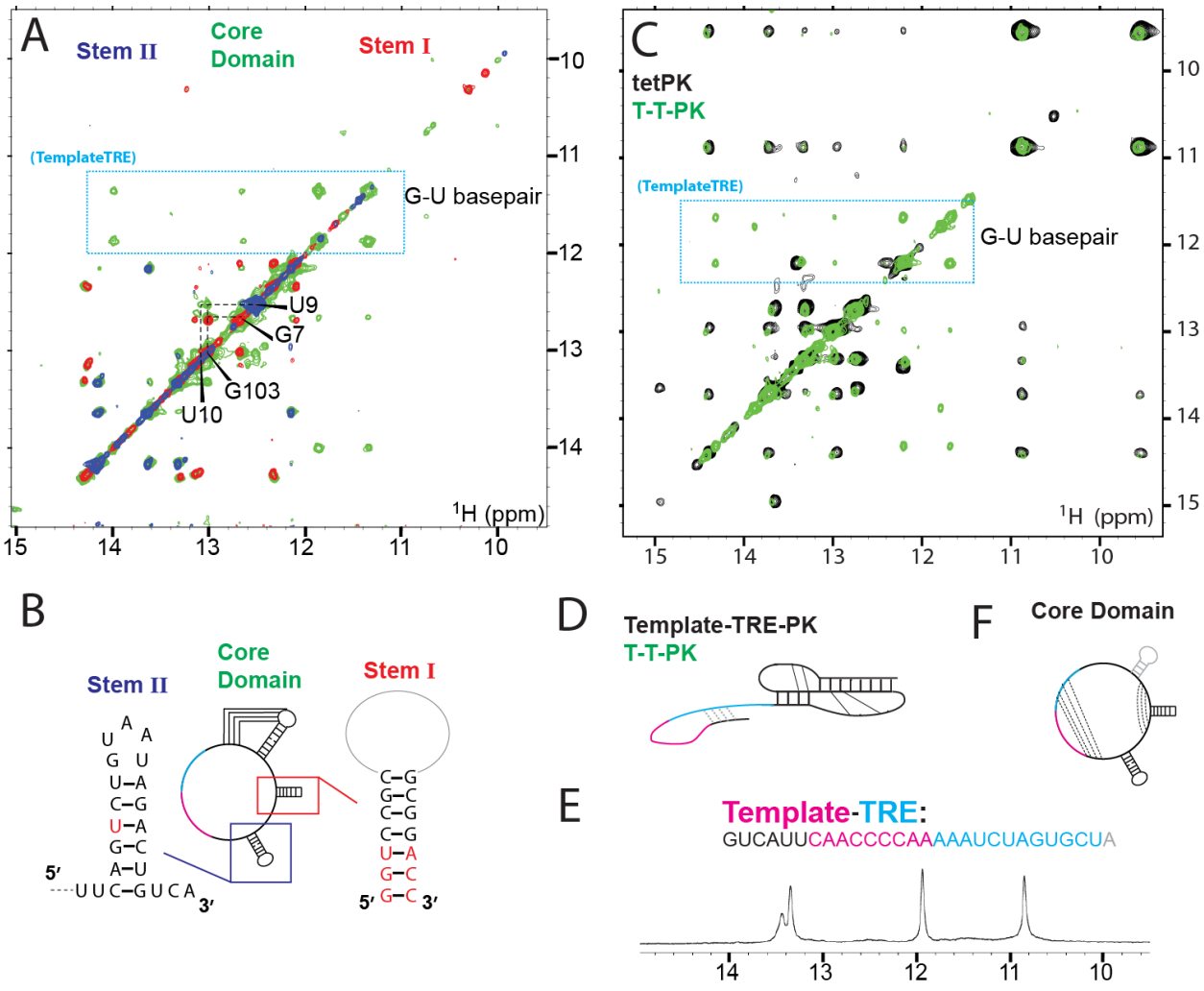


Figure 4: Folding of the TER Core domain

(A) Overlay of 2D imino NOESY spectra from Core domain (green), Stem I (red), and Stem II (blue). Template-TRE hairpin resonances are indicated by cyan box. Imino assignment and connectivity is diagrammed for Stem I elongation (black). **(B)** Stem I, Stem II, and Core domain NMR constructs. Red nucleotides indicate changes from WT sequence. **(C)** Overlay of 2D imino NOESY spectra from tetPK (black) and Template-TRE-PK (T-T-PK, green) with Template-TRE hairpin resonances indicated by cyan box. **(D)** Diagram of T-T-PK structure. **(E)** Sequence of Template-TRE (magenta-cyan) NMR construct and corresponding 1D imino spectra. **(F)** Secondary structure of the Core domain in absence of TERT. The potential Stem IIIa hairpin is gray.

imino-imino NOE connectivities. The base pairing extends from C8-G107 to U9-A101 and then U10-A100 (Fig. 4A). Since terminal iminos are rarely observed due to rapid exchange with water, it is likely that an additional A11-U101 base pair also forms. Aside from Stem I, Stem II and the Template-TRE hairpin, there are no additional crosspeaks in the Core domain 2D imino spectrum indicating that a stable structure does not form in the pseudoknot region. However there are a few unassigned broad peaks in the 1D imino spectrum of the Core domain, which indicates that transient structures may form (data not shown). This is consistent with previous diethylpyrocarbonate (DEP) footprinting data, which probes for single stranded adenines (*38*). In the pseudoknot region of free TER, only A76 and A83 were protected (while A69, A70 of Stem IIIa and A79, A80, A93, A94 of Stem IIIb had single-stranded reactivity) (*38*). The authors concluded that the pseudoknot region of free TER is largely unstructured, with the two pseudoknot stems in equilibrium, with Stem IIIa the major species. Our NMR data confirms that tetPK does not form in free TER. However the NMR data indicates that the Template-TRE forms an internal hairpin which is consistent with SHAPE data that suggests structure within this region. In addition, the extension of the Stem I to the bottom of pseudoknot Stem IIIb may explain why the pseudoknot is destabilized in the context of the Core domain. A model of the Core domain with the proposed base pairing scheme is shown in Figure 5.

To analyze the base pairing competition between the pseudoknot and Stem I extension, we made a construct spanning PK-StemI-StemII (PK-S1-SII), using a circularly permuted RNA (which joins the native 5′ and 3′ ends with a hairpin as previously described (*65*), diagrammed in Fig. 5). The 1D (and 2D, not shown) imino spectrum shows that tetPK forms in the PK-SI-SII RNA indicated by the distinct amino protons of $A^{+}80$ (and other corresponding iminos) (Fig. 6). Surprisingly, the competition between Stem I extension and pseudoknot formation favors the pseudoknot in this RNA construct. Since the pseudoknot is favored in this "open" construct as opposed to the "closed" Core domain, we formed the hypothesis that this difference in topological restraint impacts the structural conformation. This was tested by modifying the T-T-PK construct (which forms a pseudoknot), by adding an enclosing Stem I, making StemI-

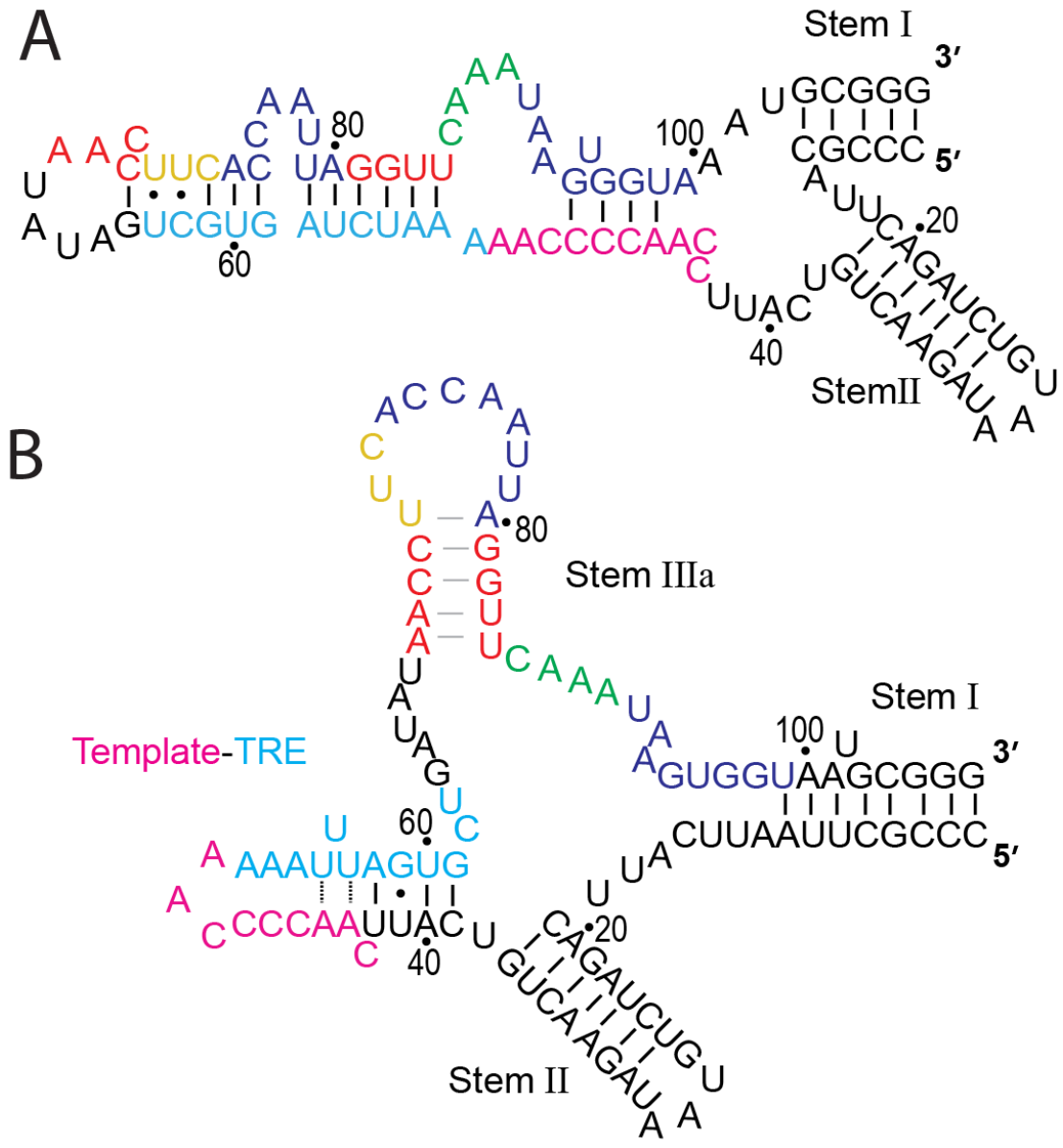


Figure 5: Secondary structure models of *Tetrahymena* TER

(A) TER model based on SHAPE data (*41*). **(B)** TER model based on NMR analysis. The potential Stem IIIa hairpin is shown with gray lines based on the evidence of transient structures forming in the pseudoknot region. In addition, previous footprinting analysis indicated that Stem IIIa and IIIb are in equilibrium with each other, with Stem IIIa being the major species (*38*). Coloring scheme is the same as Figure 1.

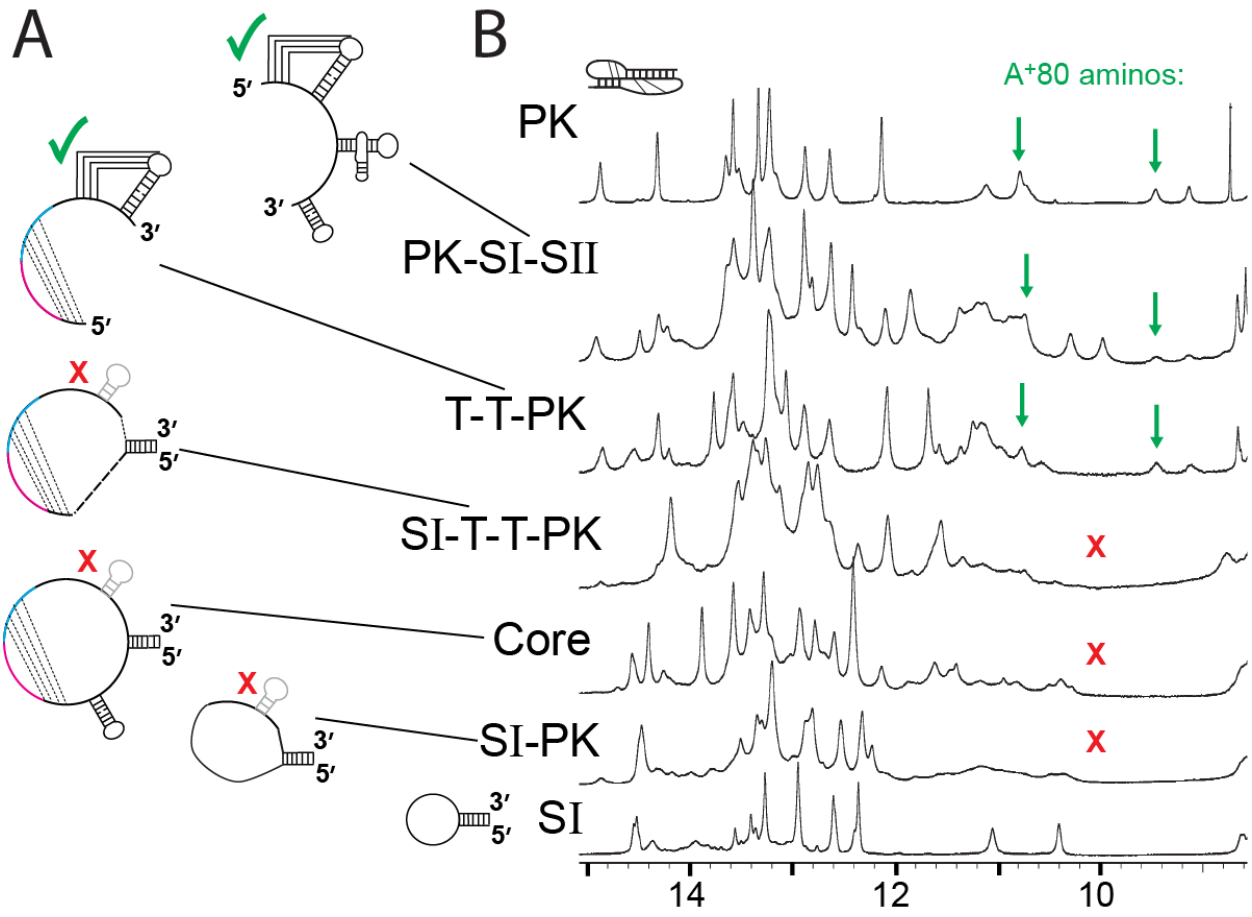


Figure 6: *Tetrahymena* TER pseudoknot formation is dependent on closed vs. open TER circle

(A) TER NMR constructs with different combinations of TER elements, including PK-StemI-StemII (PK-SI-SII), Template-TRE-PK (T-T-P-K), StemI-Template-TRE-PK (SI-T-T-P-K), StemI-PK (SI-PK), StemI (SI), and Core Domain. Whether each construct forms a pseudoknot (check mark) or not (X mark) is indicated. **(B)** 1D imino spectra of corresponding TER NMR constructs. A80 aminos are indicated by green arrows (or absence "X").

Template-TRE-PK (SI-T-T-PK) (Fig. 6). Addition of this enclosing stem abolished pseudoknot formation, while the Template-TRE hairpin is maintained. Furthermore, adding an enclosing stem around the pseudoknot residues (nt 63 – 102 with Stem I, making SI-PK), also abolishes pseudoknot formation (Fig. 6). If tetPK stability is affected by closing the TER circle, then co-transcriptional folding may also play a part in pseudoknot formation, since the pseudoknot may properly fold before Stem I forms and closes the circle. Therefore we purified the Core domain RNA under non-denaturing conditions, using ion exchange and size-exclusion columns. However there were no significant differences between native and denatured/re-folded Core domain (data not shown). Likely, the topological constraints within the closed circle of TER, combined with surrounding structures on its 5′ and 3′ ends, add to the intrinsic instability of tetPK, resulting in improper folding. The RBD of TERT binds the base of Stem II, which would disrupt the Template-TRE hairpin, and may be part of the mechanism which promotes tetPK formation as it causes the TER circle to become more open (*40, 41, 66*).

Discussion

Previous studies of ciliate, yeast, and human telomerase RNA have shown that formation of the pseudoknot and its tertiary interactions are essential for activity in the context of the minimal telomerase RNP (TERT + TER) (*13, 17, 28, 57, 67-69*). Ciliates have an assembly co-factor (p65) that can be purified and added to *in vitro* reactions to make a more robust telomerase enzyme (*34, 70*). Some early experiments purified telomerase holoenzyme from *Tetrahymena* cell extracts, which presumably contain p65 (*71, 72*). In these holoenzymes, MNase treatment of endogenous TER (followed by MNase deactivation) and supplementation with *in vitro* transcribed TER variants allowed for mutational studies of telomerase activity. p65 stimulates telomerase activity in ciliates, however the degree to which it can rescue pseudoknot mutations varies depending on the severity of the mutation (*57*). For example, one study which assayed holoenzyme with pseudoknot Stem IIIb deleted found telomerase activity reduced to 63% (*71*). Due to the presence of p65, the effect of such severe mutations results in a moderate decrease,

however the activity is consistently lower than WT (*57*). On the other hand, when less severe mutations are tested, p65 is able to recover activity to near WT levels (*57, 71, 72*). One study concluded that tetPK was not essential for telomerase function based on their *in vitro* (holoenzyme) and *in vivo* analysis (*72*). However the authors assumed that they disrupted tetPK formation, although the structural studies presented here and in the *K. lactis* TER pseudoknot study suggest that some pseudoknot structure remains, which would explain why their mutations exhibited WT activity. For example, deleting nt 86-89 was assumed to disrupt tetPK formation. Δ86-89 would disrupt the terminal A-U Stem IIIa base pairs of tetPK, although the stem IIIa G-C base pairs at the junction remain intact in addition to the tertiary interactions which support pseudoknot formation. The Δ86-89 pseudoknot could form continuous base pairing and stacking interactions through the stems and junction, a conserved pseudoknot attribute. Another mutation tested in this study was adding a 4nt bulge (CAAU) bulge at position 81/82 in Stem IIIb. Adding a bulge in Stem IIIb would not necessarily abolish tetPK formation, since all secondary and tertiary interactions remain intact. A number of bulge residues exist in the *K. lactis* and predicted *S. cerevisiae* pseudoknots which remain functional (*25, 73*). For these reasons, we conclude that the mutations tested in this study (Δ86-89 and 81/82-CAAU bulge) result in formation of an alternate, yet functional pseudoknot, which maintains WT activity in the presence of p65 (*72*). The collective data indicates that accessory protein p65 is able to rescue moderate tetPK mutations (in which an alternate pseudoknot likely forms) attaining WT activity levels, while more severe pseudoknot mutations (in which the pseudoknot does not form) lead to significant telomerase dysfunction, indicating an important pseudoknot role. This is supported by a single-molecule FRET study where tetPK equilibrium state was investigated in correlation with telomerase activity (*40*). They discovered that only RNP molecules containing a properly folded pseudoknot were catalytically active.

Comparison of ciliate, human, and yeast pseudoknots

The *Tetrahymena* TER pseudoknot has a number of similarities and differences to those of human (hPK) and yeast *K. lactis* (kPK), which were previously solved by NMR (Fig 6). All three pseudoknots have continuous stacking interactions through the stems and junction, appearing as continuous A-form RNA molecules. Only hPK has an intervening loop-loop interaction at the junction, while tetPK and kPK have stem-stem stacking. All three pseudoknots have a stabilizing triple helix that includes major groove Loop A-Stem B tertiary interactions. However, only tetPK and hPK have minor groove Loop B-Stem A base triples. kPK cannot form the typical minor groove interactions due to the lack of adenines in Loop B, and is likely compensated by its extended major groove triplex. The base triples were shown to be important for proper pseudoknot folding for all three pseudoknots. In both tetPK and kPK, mutations in the loop residues which disrupt base triples abolished formation of Stem A. In hPK, base triple mutations disrupted the stem base pairs near the junction. The stems of TER pseudoknots likely have varying degrees of stability based on sequence, bulges, G-C content and length, making them dependent on the base triple interactions to ensure proper folding.

TetPK has a significantly different tertiary structure when compared to hPK and kPK (Fig. 7B). The backbone of hPK and kPK overlay very well and the structures appear remarkably similar despite differences in sequence and tertiary interactions. TetPK on the other hand is more compact with a major groove width of 11Å, compared to ~15Å for both hPK and kPK. This may be partly due to the different ratios in the number of nucleotides in Loop A to Stem B. hPK has a ratio of ~1:1, kPK is 1:2, and tetPK is closer to ~1:3 (Loop A: Stem B nt ratio). This means that tetPK loop A must span a greater distance per nucleotide and is consistent with a more compact molecule. However, it is important to note that these pseudoknots belong to different organisms with different overall TER structure and holoenzyme components, and therefore their similarities may reflect a conserved function while their differences reflect their diverse environments. One study investigated a chimeric human TER, where hPK was replaced with tetPK, which could potentially compensate for function since the secondary structures of the two pseudoknots are

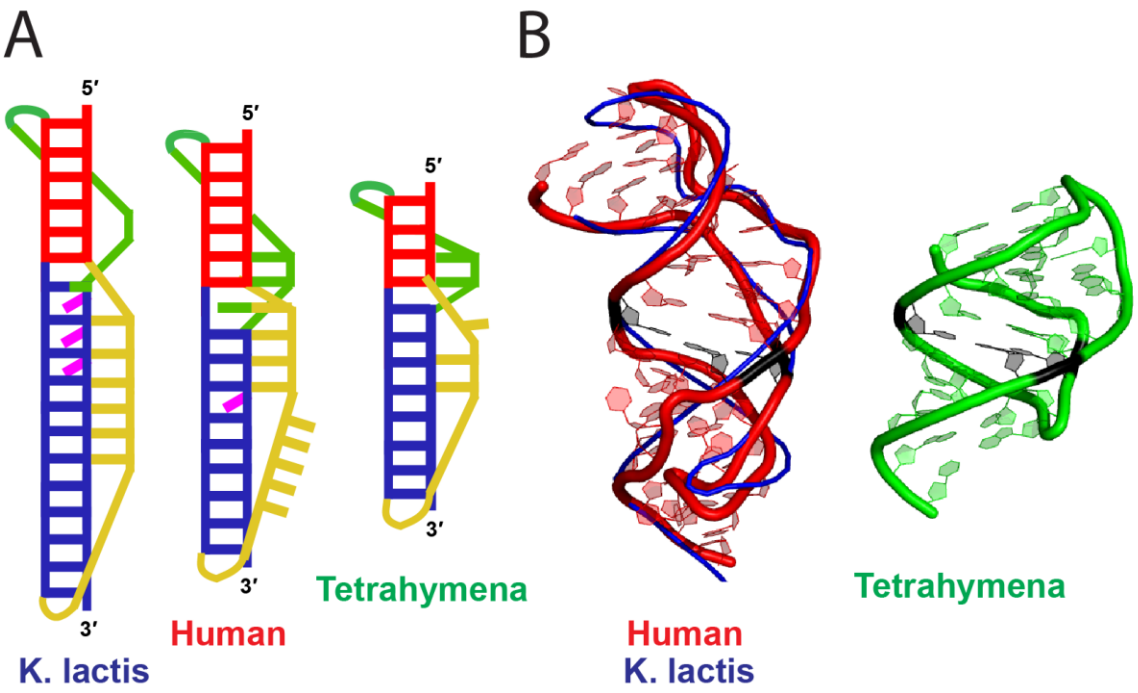


Figure 7: Comparison of ciliate, human, and yeast TER pseudoknots

(A) Diagram of TER pseudoknot secondary structure from *K. lactis*, human and *Tetrahymena*. **(B)** NMR solution structures of TER pseudoknots from human (red) and *K. lactis* (backbone only, blue) which are overlayed. TetPK is (green). Human and tetPK structures are aligned according to their similar base triples, with the corresponding U-A-U triple in each colored black.

similar (Fig. 7A) (*74*). However the chimeric TER was only "weakly active" when assayed *in vitro*, supporting the conclusion that the pseudoknot differences reflect species-specific requirements. The solution structures of the two pseudoknots explain their inability to be interchanged, since tetPK has a different shape and size than that of human (Fig. 7B).

This work highlights the important conserved features of TER pseudoknots, including their stabilizing tertiary interactions and stacked stems, which are important for telomerase activity. The solution structure of the *Tetrahymena* pseudoknot reveals its different shape and size, compared to the human and yeast pseudoknots which are remarkably similar. The similarities and differences between these pseudoknot structures will help delineate how the pseudoknot performs its function while giving insights into the differences between species.

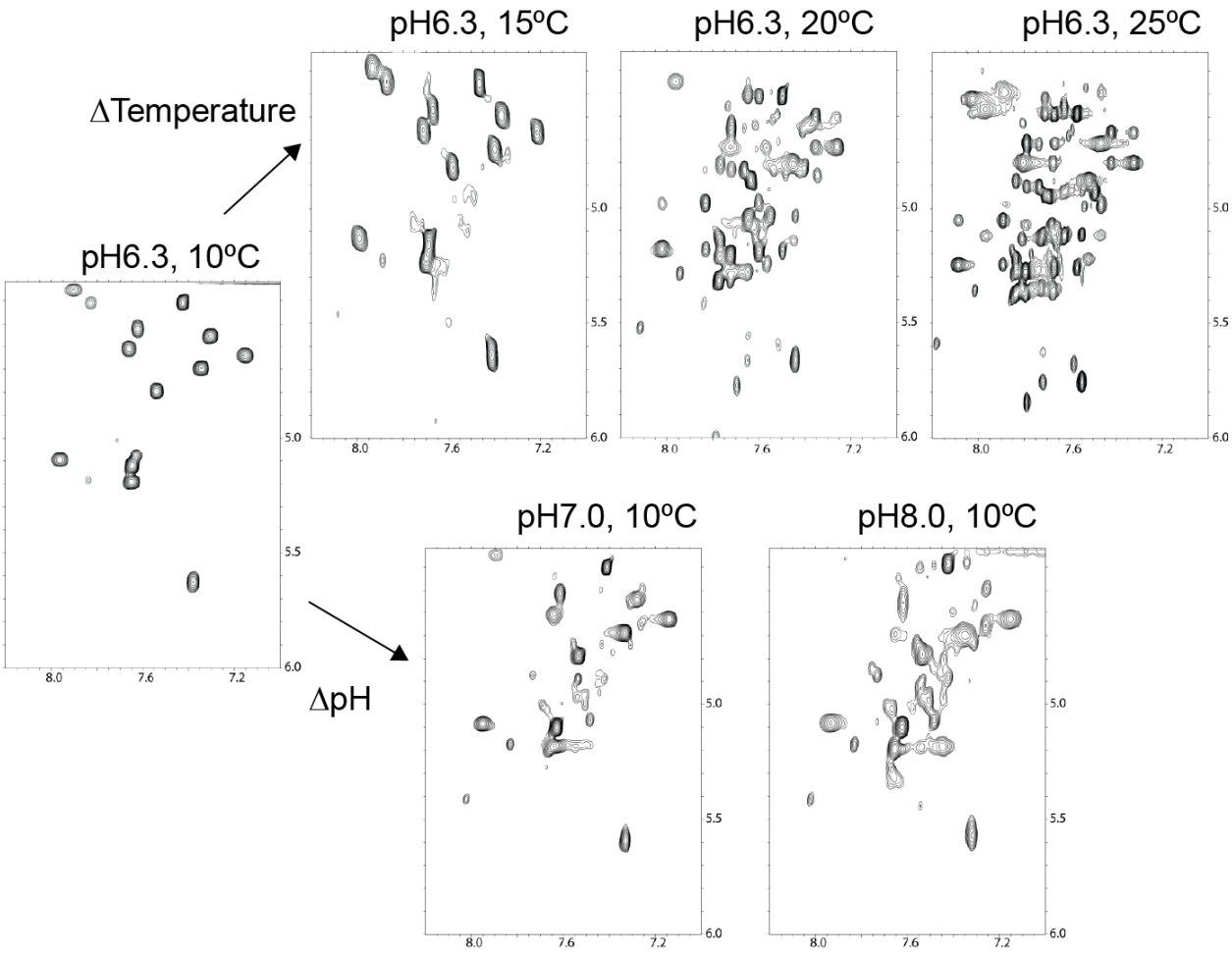


Figure S1: TetPK is unstable at higher temperature and pH

2D (H5-H6) TOCSY spectra of tetPK at varying temperature and pH. The increasing number and intensity of TOCSY crosspeaks indicates higher population of unfolded pseudoknot.

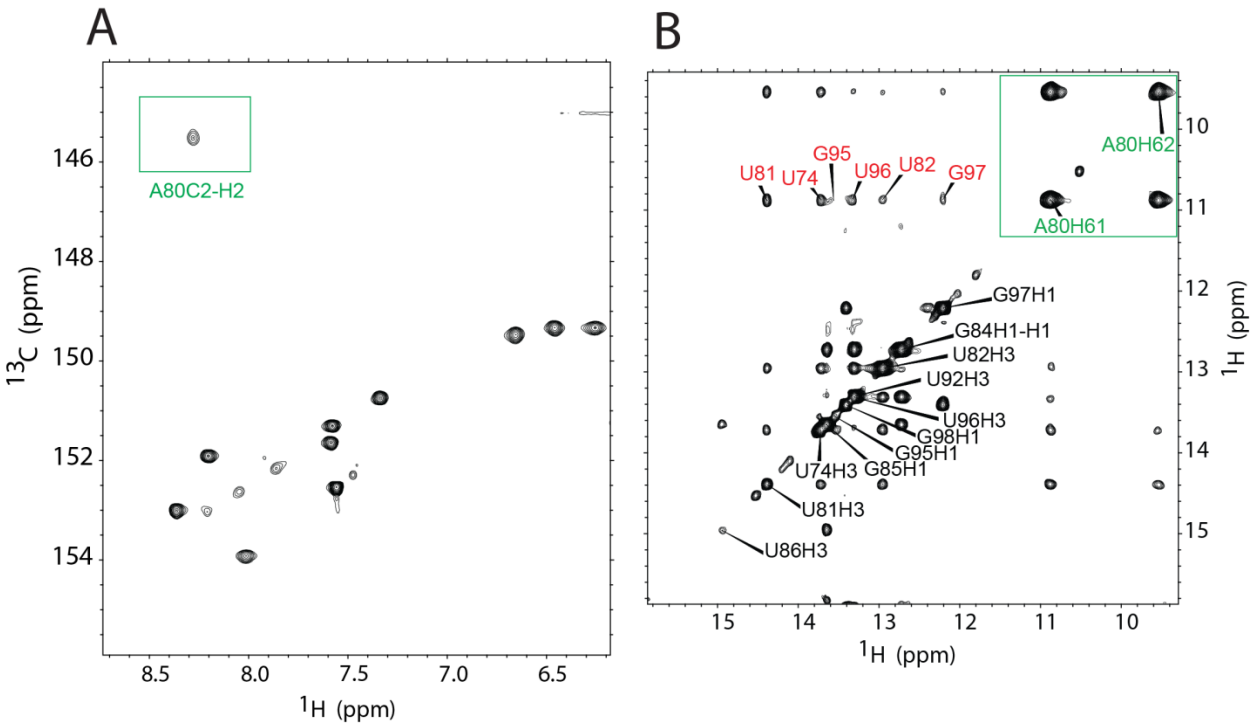


Figure S2: Evidence for N1 protonation of A80

(A) ^{13}C-^{1}H HSQC of C2-H2 region of tetPK. The C2 resonance of A80 is upfield shifted, boxed in green. **(B)** 2D imino NOESY of tetPK. The pseudoknot imino resonances are assigned and labeled on the diagonal in black. The A80 amino resonances are downfield shifted and split, boxed in green. The A80 amino resonances have NOE crosspeaks to various other pseudoknot imino resonances, indicated in red.

References

1. Collins, K. (2006) The biogenesis and regulation of telomerase holoenzymes, *Nat. Rev. Mol. Cell. Biol. 7*, 484-494.

2. Harrington, L. (2003) Biochemical aspects of telomerase function, *Cancer Lett 194*, 139-154.

3. Blackburn, E. H. (2000) The end of the (DNA) line, *Nat Struct Biol 7*, 847-850.

4. Nandakumar, J., and Cech, T. R. (2013) Finding the end: recruitment of telomerase to telomeres, *Nat Rev Mol Cell Biol 14*, 69-82.

5. Blackburn, E. H. (2001) Switching and signaling at the telomere, *Cell 106*, 661-673.

6. Blackburn, E. H., and Collins, K. (2011) Telomerase: an RNP enzyme synthesizes DNA, *Cold Spring Harb Perspect Biol 3*.

7. O'Sullivan, R. J., and Karlseder, J. (1994) Telomeres: protecting chromosomes against genome instability, *Nat Rev Mol Cell Biol 11*, 171-181.

8. Mason, M., Schuller, A., and Skordalakes, E. Telomerase structure function, *Curr Opin Struct Biol 21*, 92-100.

9. Kachouri-Lafond, R., Dujon, B., Gilson, E., Westhof, E., Fairhead, C., and Teixeira, M. T. (2009) Large telomerase RNA, telomere length heterogeneity and escape from senescence in Candida glabrata, *FEBS Lett 583*, 3605-3610.

10. Egan, E. D., and Collins, K. (2012) Biogenesis of telomerase ribonucleoproteins, *RNA 18*, 1747-1759.

11. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H., and James, T. L. (2002) A molecular switch underlies a human telomerase disease, *Proc Natl Acad Sci U S A 99*, 16998-17003.

12. Zhang, Q., Kim, N. K., and Feigon, J. (2011) Architecture of human telomerase RNA, *Proc Natl Acad Sci U S A 108*, 20325-20332.

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

14. Chen, J. L., and Greider, C. W. (2003) Template boundary definition in mammalian telomerase, *Genes & Development 17*, 2747-2752.

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

16. Kim, N. K., Zhang, Q., Zhou, J., Theimer, C. A., Peterson, R. D., and Feigon, J. (2008) Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA, *J Mol Biol 384*, 1249-1261.

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

18. Chen, Y., Fender, J., Legassie, J. D., Jarstfer, M. B., Bryan, T. M., and Varani, G. (2006) Structure of stem-loop IV of Tetrahymena telomerase RNA, *Embo Journal 25*, 3156-3166.

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

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

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

22. Richards, R. J., Wu, H., Trantirek, L., O'Connor, C. M., Collins, K., and Feigon, J. (2006) Structural study of elements of Tetrahymena telomerase RNA stem-loop IV domain important for function, *Rna-a Publication of the Rna Society 12*, 1475-1485.

23. Theimer, C. A., Finger, L. D., Trantirek, L., and Feigon, J. (2003) Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA, *Proc. Natl. Acad. Sci. USA 100*, 449-454.

24. Theimer, C. A., Jady, B. E., Chim, N., Richard, P., Breece, K. E., Kiss, T., and Feigon, J. (2007) Structural and functional characterization of human telomerase RNA processing and Cajal body localization signals, *Mol. Cell 27*, 869-881.

25. Cash, D. D., Cohen-Zontag, O., Kim, N. K., Shefer, K., Brown, Y., Ulyanov, N. B., Tzfati, Y., and Feigon, J. (2013) Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo, *Proc Natl Acad Sci U S A 110*, 10970-10975.

26. Tzfati, Y., Knight, Z., Roy, J., and Blackburn, E. H. (2003) A novel pseudoknot element is essential for the action of a yeast telomerase, *Genes Dev 17*, 1779-1788.

27. Gilley, D., and Blackburn, E. H. (1999) 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.

28. Lai, C. K., Miller, M. C., and Collins, K. (2003) Roles for RNA in telomerase nucleotide and repeat addition processivity, *Mol Cell 11*, 1673-1683.

29. ten Dam, E., van Belkum, A., and Pleij, K. (1991) A conserved pseudoknot in telomerase RNA, *Nucleic Acids Res 19*, 6951.

30. Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006) Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging, *Nature Medicine 12*, 1133-1138.

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

32. Ye, A. J., and Romero, D. P. (2002) Phylogenetic relationships amongst tetrahymenine ciliates inferred by a comparison of telomerase RNAs, *Int J Syst Evol Microbiol 52*, 2297-2302.

33. Jiang, J., Miracco, E. J., Hong, K., Eckert, B., Chan, H., Cash, D. D., Min, B., Zhou, Z. H., Collins, K., and Feigon, J. (2013) The architecture of Tetrahymena telomerase holoenzyme, *Nature 496*, 187-192.

34. O'Connor, C. M., and Collins, K. (2006) A novel RNA binding domain in tetrahymena telomerase p65 initiates hierarchical assembly of telomerase holoenzyme, *Mol Cell Biol 26*, 2029-2036.

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

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

37. Singh, M., Wang, Z., Koo, B. K., Patel, A., Cascio, D., Collins, K., and Feigon, J. (2012) Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65, *Mol Cell 47*, 16-26.

38. Bhattacharyya, A., and Blackburn, E. H. (1994) Architecture of telomerase RNA, *EMBO J 13*, 5721-5731.

39. Ulyanov, N. B., Shefer, K., James, T. L., and Tzfati, Y. (2007) Pseudoknot structures with conserved base triples in telomerase RNAs of ciliates, *Nucleic Acids Research 35*, 6150-6160.

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

41. Cole, D. I., Legassie, J. D., Bonifacio, L. N., Sekaran, V. G., Ding, F., Dokholyan, N. V., and Jarstfer, M. B. (2012) New models of Tetrahymena telomerase RNA from experimentally derived constraints and modeling, *J Am Chem Soc 134*, 20070-20080.

42. Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., and Williamson, J. R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA, *Nucleic Acids Res 20*, 4515-4523.

43. Dieckmann, T., and Feigon, J. (1997) Assignment methodology for larger RNA oligonucleotides: application to an ATP-binding RNA aptamer, *J. Biomol. NMR 9*, 259-272.

44. Gullerez, J., Lopez, P. J., Proux, F., Launay, H., and Dreyfus, M. (2005) A mutation in T7 RNA polymerase that facilitates promoter clearance, *Proc. Natl. Acad. Sci. USA 102*, 5958-5963.

45. Pley, H. W., Flaherty, K. M., and McKay, D. B. (1994) Three-dimensional structure of a hammerhead ribozyme, *Nature 372*, 68-74.

46. Scott, W. G., Finch, J. T., and Klug, A. (1995) The crystal structure of an all-RNA hammerhead ribozyme, *Nucleic Acids Symp Ser*, 214-216.

47. Petrov, A., Wu, T., Puglisi, E. V., and Puglisi, J. D. (2013) RNA purification by preparative polyacrylamide gel electrophoresis, *Methods Enzymol 530*, 315-330.

48. Cromsigt, J. A., Hilbers, C. W., and Wijmenga, S. S. (2001) Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation, *J Biomol NMR 21*, 11-29.

49. Peterson, R. D., Theimer, C. A., Wu, H., and Feigon, J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes, *J. Biomol. NMR 28*, 59-67.

50. Wu, H., Yang, P. K., Butcher, S. E., Kang, S., Chanfreau, G., and Feigon, J. (2001) A novel family of RNA tetraloop structure forms the recognition site for Saccharomyces cerevisiae RNase III, *EMBO J. 20*, 7240-7249.

51. Dingley, A. J., and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings, *J. Am. Chem. Soc. 120*, 8293-8297.

52. Dingley, A. J., Nisius, L., Cordier, F., and Grzesiek, S. (2008) Direct detection of N-H[...]N hydrogen bonds in biomolecules by NMR spectroscopy, *Nat Protoc 3*, 242-248.

53. Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol. 5*, 1065-1074.

54. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J Magn Reson 160*, 65-73.

55. de Alba, E., and Tjandra, N. (2002) NMR dipolar couplings for the structure determination of biopolymers in solution, *Progress in Nuclear Magnetic Resonance Spectroscopy 40*, 175-197.

56. Bryan, T. M., Goodrich, K. J., and Cech, T. R. (2000) A mutant of Tetrahymena telomerase reverse transcriptase with increased processivity, *J Biol Chem 275*, 24199-24207.

57. Berman, A. J., Gooding, A. R., and Cech, T. R. (2010) 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.

58. Guo J., Li N. (1977) An NMR study of the interaction of adenine with hydrochloride, *Soochow Journal of Math and Natural Science* Vol 3,143-150.

59. Sklenar, V., and Feigon, J. (1990) Formation of a stable triplex from a single DNA strand, *Nature 345*, 836-838.

60. Legault, P., and Pardi, A. (1994) 31P chemical shift as a probe of structural motifs in RNA, *J Magn Reson B 103*, 82-86.

61. Pan, B., Mitra, S. N., and Sundaralingam, M. (1999) Crystal structure of an RNA 16-mer duplex R(GCAGAGUUAAAUCUGC)2 with nonadjacent G(syn).A+(anti) mispairs, *Biochemistry 38*, 2826-2831.

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

63. Su, L., Chen, L., Egli, M., Berger, J. M., and Rich, A. (1999) Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot, *Nat Struct Biol 6*, 285-292.

64. Zaug, A. J., and Cech, T. R. (1995) Analysis of the structure of Tetrahymena nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA, *RNA 1*, 363-374.

65. Miller, M. C., and Collins, K. (2002) Telomerase recognizes its template by using an adjacent RNA motif, *Proc Natl Acad Sci U S A 99*, 6585-6590.

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

67. Licht, J. D., and Collins, K. (1999) Telomerase RNA function in recombinant Tetrahymena telomerase, *Genes Dev 13*, 1116-1125.

68. Sperger, J. M., and Cech, T. R. (2001) A stem-loop of Tetrahymena telomerase RNA distant from the template potentiates RNA folding and telomerase activity, *Biochemistry 40*, 7005-7016.

69. Lin, J., Ly, H., Hussain, A., Abraham, M., Pearl, S., Tzfati, Y., Parslow, T. G., and Blackburn, E. H. (2004) A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein, *Proc. Natl. Acad. Sci. USA 101*, 14713-14718.

70. Prathapam, R., Witkin, K. L., O'Connor, C. M., and Collins, K. (2005) A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase, *Nature Structural & Molecular Biology 12*, 252-257.

71. Autexier, C., and Greider, C. W. (1998) Mutational analysis of the Tetrahymena telomerase RNA: identification of residues affecting telomerase activity in vitro, *Nucleic Acids Res 26*, 787-795.

72. Cunningham, D. D., and Collins, K. (2005) Biological and biochemical functions of RNA in the tetrahymena telomerase holoenzyme, *Mol Cell Biol 25*, 4442-4454.

73. Liu, F., Kim, Y., Cruickshank, C., and Theimer, C. A. (2012) Thermodynamic characterization of the Saccharomyces cerevisiae telomerase RNA pseudoknot domain in vitro, *RNA 18*, 973-991.

74. Marie-Egyptienne, D. T., Cerone, M. A., Londono-Vallejo, J. A., and Autexier, C. (2005) A human-Tetrahymena pseudoknot chimeric telomerase RNA reconstitutes a nonprocessive enzyme in vitro that is defective in telomere elongation, *Nucleic Acids Research 33*, 5446-5457.

CHAPTER 4

Analysis and optimization of NMR structures of A-form RNA helices

Introduction

Three-dimensional nucleic acid structure has been thoroughly studied through a variety of methods, including X-ray crystallography and nuclear magnetic resonance (NMR). Crystal structures provide high resolution details of the oligonucleotides, however the molecule is subjected to crystal packing forces and local high ion concentrations which may affect the structure (*1, 2*). Such was the case of a 17 base pair RNA duplex which displayed four different conformations in a single crystal (*2*). As an alternative, NMR is a powerful tool for examining nucleic acid structure in solution, but it has disadvantages as well.

For example, protein structures are better characterized by NMR compared to RNA helices, due to their extensive long range proton interaction networks which can precisely define the global fold. Conversely, nucleic acids have a lower proton density and their nuclear Overhauser effects (NOEs) are mostly from intra-residue and sequential base pair interactions as opposed to long range (*3, 4*). In addition, since distance and angle restraints are given loose ranges, a variety of structures are potentially able to fit the data. Another disadvantage of NMR of RNA is the limited chemical shift dispersion due to typically having only four different nucleotides. This has been addressed by many methods, including ^{13}C and ^{15}N isotope labeling of the molecules for heteronuclear NMR experiments, as well as selective deuteration to eliminate specific resonances from the spectra (*5-9*). RNA sequential assignment can be made with multidimensional through-bond coherence transfer experiments and 2D filtered/edited NOESY experiments with increased resolution (*10, 11*). Furthermore, the incorporation of residual dipolar couplings (RDCs) has enhanced NMR structure determination by providing a means to orient distal elements of the molecule and refine the global structure (*4, 12-18*). RDCs are correlated to the orientation of individual bond vectors (^{1}H-^{15}N or ^{1}H-^{13}C) relative to a single axis system defined by the alignment tensor. RDCs are measured in a similar way as scalar couplings, adding on to the inherent coupling constant (~90Hz for $^{1}J_{HN}$ and ~160-200Hz for $^{1}J_{HC}$). RDCs only have substantial values when measured in anisotropic solutions, since they average to zero in isotropic conditions due to random molecular tumbling. To partially align the

molecules for RDC measurement different media have been employed, including bicelles, filamentous Pf1 bacteriophage, acrylamide gel, and DNA nanotubes (*18-20*).

While NMR methodology has significantly advanced, an analysis by Tolbert et al. in 2010 revealed a significant difference in the helical properties of X-ray versus NMR determined RNA structures (*21*). The most critical criteria for helix assessment used was the major groove width, measured as the distance between phosphate atoms of the (i) and (i+6) base paired residues (d_{P-P}). They analyzed seven crystal structures with Watson-Crick (WC) base pairs, compared to helical RNA structures deposited (in RCSB Structure Databank) between 2005 and 2009. They found that the major groove width of X-ray structures varied from 8.8 - 12.6Å, with an average of 11.1Å, while NMR structures varied from 7.4 - 28.6Å, with an average of 15.7Å. This is a surprising difference, especially considering that the upper limit of the major groove width for NMR structures is more than twice that of X-ray structures. To determine the source of this difference we analyzed the structures with large major grooves for apparent trends and/or errors. Two examples of such structures were the *Tetrahymena* telomerase RNA (TER) stem-loop 2 (SL2) and stem-loop 4 (SL4), which had major groove widths of 26.7Å and 28.6Å respectively (Fig. 1) (*22, 23*). These two structures contain mostly canonical WC base pairs, except SL2 has one A-A mismatch base pair. They are expected to have A-form geometry, although the determined structures have an extended shape (Fig. 1B).

Using SL4 as a model hairpin, we did a systematic analysis of the restraints necessary to calculate an accurate helix using Xplor-NIH (*24*). We found that the distance restraints derived from sequential base-to-base NOEs (H5, H6, H8 to neighboring H5, H6, H8) have a significant impact on determining the helical characteristics of RNA structures, with values assigned as too large leading to extended helices. We also found that incorporating more distance restraints, from difficult to obtain imino NOE crosspeaks, lead to an increase in structural precision and accuracy. Furthermore, the inclusion of a sufficient number of non-redundant RDCs is essential to properly refining the structure. With this strategy, we re-determined the structures of *Tetrahymena* telomerase RNA (TER) SL2 and SL4 using improved distance restraints and

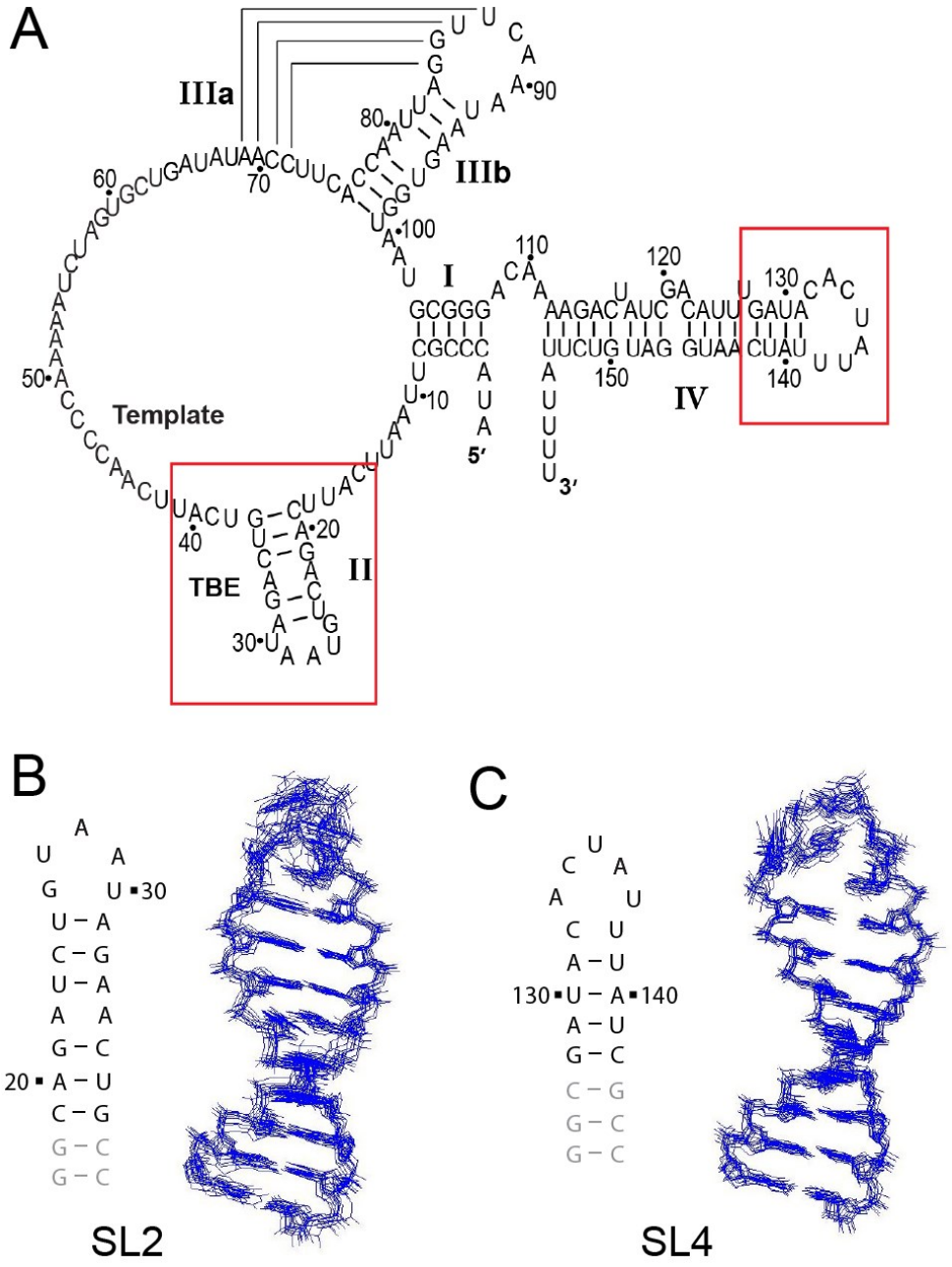


Figure 1: *Tetrahymena* telomerase RNA

(A) Diagram of *Tetrahymena* TER, with helical segments numbered I-IV. Stem-loops 2 and 4 (SL2 and SL4) are boxed in red. NMR constructs of **(B)** SL2 and **(C)** SL4 are shown, with corresponding lowest energy NMR structures (from original data). Non-native residues are colored gray.

RDCs. The resulting helices were more accurate with better A-form characteristics. These improved structures were subsequently fit into the *Tetrahymena* telomerase holoenzyme electron microscopy (EM) map for domain modelling (*25*).

Results and Discussion

Xplor-NIH RNA structure calculation method

The detailed methods for RNA sample preparation, NMR experiment acquisition, and resonance assignment have been described and reviewed elsewhere (*11, 26-31*). Here we focus on the structure determination process. Analysis of the NMR data is used for compiling restraint lists which are used to drive molecular dynamics (MD) simulations for structure calculation. A force field describes the potential energy function and parameters for interactions of the biomolecules during simulated annealing. The nonbonded interactions can be further defined by including torsion angle and base-base positioning database potentials derived from high resolution crystal structures, which bias sampling during simulated annealing to more reasonable conformations and has been shown to enhance NMR structures (*32*). The two major software packages for these types of calculations are CYANA and Xplor-NIH, which differ in their force field function, and can also be followed by further refinements such as AMBER (force field) (*24, 33, 34*). We will focus on usage of the Xplor-NIH software since it is widely used, and has been used successfully in our lab. Xplor starts from an extended single strand molecule which undergoes MD simulated annealing using the defined restraints. A second step refines the initial structures with "slow-cooling", allowing the energy to be further minimized and for structures trapped in local minima to escape. This refinement can be repeated iteratively and results in what we refer to as the "NOE structure" of the molecule, since it is based mostly on NOE data. In the last step, RDCs are applied to the NOE structure for refinement. RDCs orient the bond vectors relative to an alignment tensor axis, which is defined by its axial (Da) and rhombic (Dr) components. Da and Dr are determined by doing a grid search (changing Da and Dr

systematically) to find optimal values which best fit the data. RDCs are then used as restraints in refining the molecule to produce the final structure.

In addition to experimentally determined dihedral angles, it is common to use standard A-form dihedral angle restraints for helical segments of the RNA while leaving loops unrestrained (*35, 36*). Non-experimental dihedral angles are typically α, β, γ, ζ, and ν (given in Table 1). δ correlates to the sugar pucker, which is determined experimentally by analysis of the H1′-H2′ TOCSY (or DQF-COSY) crosspeaks (*37*). Strong crosspeaks indicate C2′-endo (145 ± 30°) and no cross-peak indicates C3′-endo (82 ± 30°), with intermediate intensities given an appropriate value. χ correlates to the glycosidic bond and is determined by the base-sugar NOE pattern of each residue, where strong H8-H1′ crosspeaks are indicators of syn nucleotides (anti = -160 and syn = 25 ± 30°). Hydrogen bonds are directly detected by JNN-COSY to be used as restraints (*38, 39*). Two distance restraints are used for each hydrogen bond. Weak base planarity and sugar chirality restraints are also used to enforce proper geometry of the base and sugar respectively. Distance restraints are derived from NOESY experiments, where NOE buildup experiments are used to determine the mixing time with a linear relationship between distance and peak intensity (*40, 41*). NOE crosspeak volume is related to distance by the following formula:

$$r_{ij} = r_{ref}\left(\frac{a_{ref}}{a_{ij}}\right)^{1/6}$$

where r_{ij} is the distance and a_{ij} is the NOE crosspeak volume between two protons (i and j). Pyrimidine H5-H6 crosspeaks are used for the reference (ref) with a distance equal to 2.45Å, and crosspeak volume obtained from the NOESY spectra. It is important to choose pyrimidines that are non-dynamic (in helices) and to average more than one crosspeak to generate the reference volume. Distances are categorized as very strong (2.5Å), strong (3.5Å), moderate (4.5Å), weak (5.5Å), and very weak (6.5Å). Distances are given ranges ± 1 Å, except very strong has a lower bound of 1.8Å corresponding to the van der Waals radius.

Distance restraint analysis

We generated mock datasets to determine the influence of the number of restraints on calculated structures. The summary of the datasets are listed in Table 1. Standard values for dihedral angles and distances are based on A-form helices. To analyze the importance of the distance restraints, we categorized them and added them systematically to the structure calculation. There are 5 datasets in order starting with the basic distance restraints which are generally first obtained (1) the sequential base-to-sugar and base-to-base correlations from D_2O NOESY spectra, and sequential imino-to-imino from H_2O NOESY, (2) base paired and sequential imino-to-amino. The more difficult to obtain (overlapped) NOEs are (3) sequential imino-to-base, and imino to its 5′ neighbor cross-strand amino protons (4) or base protons (5) (from H_2O NOESY spectra). RDCs were measured for SL4 RNA and were used to evaluate the NOE structures based on the correlation coefficient (RDC-R^2) and Q factor value, which is given by the formula,

$$Q = \frac{\sqrt{\sum_i (D_i^{exp} - D_i^{calc})^2}}{\sqrt{\sum_i (D_i^{exp})^2}}$$

where D^{exp} and D^{calc} are experimental and calculated RDC values (*42*). Lower Q values indicate more accurate structures, with 25% serving as a relative threshold (*42*).

The NOE structures and statistics for the Xplor calculations with different sets of data are shown in Figure 2. 100 structures were calculated for each dataset, and analysis was performed on the top 10 structures (average structure is shown). The NOE structure is an indicator of the quality of data, and is important to optimize before further refinement. Dataset 1 (base-sugar, base and imino sequentials) has the least number of distance restraints and represents the minimal data set obtained from basic sequential assignment. Xplor calculation with this data results in an NOE structure with a slightly extended major groove of 15.2Å, although it is near

Dataset 1

Dihedral angles

α = −62 ± 30, β = −179 ± 30, γ =47 ± 30, ζ = −73 ± 30, ν = 37 ± 30

δ = 82 ± 30, χ = -160 ± 30

Sequential base-to-sugar

	previous	own		previous	own
H6/H8	H1′ = 4.5	H1′ = 3.5	H1′	H2′ = 4.5	H2′ = 2.5
	H2′ = 2.5	H2′ = 3.5			H3′ = 3.5
	H3′ = 3.5	H3′ = 3.5			H4′ = 3.5
	H4′ = 5.5	H4′ = 3.5			H5′ = 4.5
		H5′ = 3.5			
		H5″ = 3.5			

	5′-Cross strand	next
H2	H1′ = 4.5	H1′ = 4.5

Sequential base-to-base

	previous	next
H6/H8	H6/H8 = 4.5	H6/H8 = 4.5
		H5 = 4.5 (or 3.5)

Sequential imino-to-imino

	previous	next
H1/H3	H1/H3 = 4.5	H1/H3 = 4.5

Dataset 2

Base paired imino-to-amino

	base paired
H1	H41 = 2.5
	H42 = 3.5
	N3 = 2.5
	N4 = 2.5
H3	H61 = 2.5
	H62 = 4.5
	H2 = 2.5
	N6 = 2.5

Sequential imino-to-amino

	previous	next
H1	H41 = 4.5	H41 = 3.5
	H42 = 6.5	H42 = 4.5
		H21 = 5.5
		H61 = 3.5
H3	H61 = 4.5	H41 = 3.5
	H62 = 4.5	H42 = 3.5

Dataset 3

Sequential imino-to-base

	next
H1	H6/H8 = 5.5
	H5 = 4.5
	H2 = 4.5
H3	H6/H8 = 4.5
	H5 = 3.5

Dataset 5

Dataset 4

5' cross-strand imino-to-amino

	5′-Cross strand
H1	H41 = 4.5
	H42 = 5.5
	H21 = 3.5
H3	H21 = 5.5

5′ cross strand imino-to-base

	5′-Cross strand
H1	H6/H8 = 5.5
	H5 = 4.5 (or 5.5)
H3	H5 = 5.5

Table 1: NMR restraints

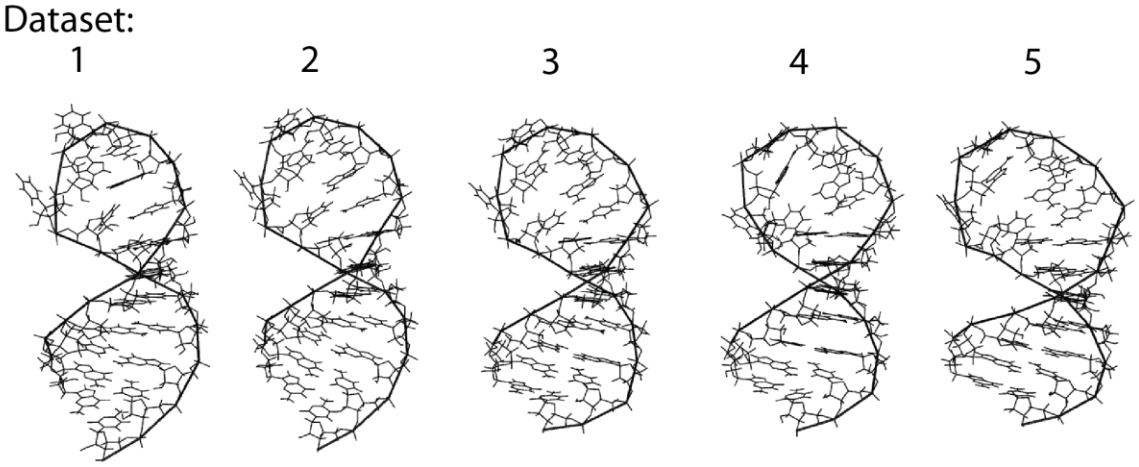
Ideal values used for dihedral angles and inter-proton distances are listed. Distance types were categorized (starting with a basic set of data) for use in structure calculations. For the mock SL4 calculations, each new dataset includes all of the data from the previous sets.

the average for NMR structures. Dataset 1 is the least accurate with R^2 value = 0.81 and Q value = 43.5%. Dataset 2 (adding base paired and sequential imino-to-amino), represents a typical dataset from a more thorough NOE analysis of the H_2O NOESY spectra, and yields a structure with major groove width = 14.5Å, R^2 = 0.87, and Q = 41.4%. Datasets 3, 4, and 5 represent a more rigorous analysis of NOESY spectra, since these distances are harder to obtain due to overlap. The addition of more distance restraints results in a progressive improvement of the structure precision and accuracy, with Dataset 5 having major groove width = 11.6Å, R^2 = 0.91, and Q = 29.1% (Fig. 2). For these datasets, the addition of RDCs improves each structure, so that they are all in the range of X-ray crystallography helices (major grooves ~10-12Å). This indicates that even with minimal distance restraints, RDC refinement can lead to structures of adequate accuracy, with similar helical parameters to that of the full dataset (Fig. 2, Table 2).

Impact of sequential base-to-base distance restraints on RNA structure

The data from Richards et al. which was used to solve the structures of SL2 and SL4 falls within the category of Dataset 2 (*22, 23*). However, the previously determined NOE structures of SL2 and SL4 had major groove widths of 25.4Å and 26.0Å respectively, compared to 14.5Å for the mock calculation with Dataset 2. This indicates an apparent error in the original distance restraints. For SL4, comparison of the mock (ideal) data and experimental data revealed that the most significant differences were in the restraints from sequential base-to-base distances (H5, H6, H8 to neighboring H5, H6, H8). In the A-form helix this distance is ~4.5Å, whereas the original data had a significant number of values set as 5.5Å, and some values were missing from the list which also could affect the structures, for both SL2 and SL4. We re-analyzed the NOESY spectra that were previously acquired and determined that the sequential base-to-base values should have been set to 4.5Å based on the proper crosspeak analysis, and were also able to assign additional crosspeaks. We then re-calculated the SL2 and SL4 structures using the original experimental distance restraints, modifying only the sequential base-to-base distances. For SL2,

A



B

NOE Dataset Number:

	1	2	3	4	5
NOE-structure:					
Major groove, D_{p-p} (Å)	15.2	14.5	14.1	13.8	11.6
10 structures - RMSD (Å)	1.12	1.04	1.30	0.91	0.72
RDC fit - R^2	0.81	0.86	0.87	0.89	0.91
RDC fit - Q	43.5	41.4	35.8	32.3	28.9
RDC-refined					
Major groove, D_{p-p} (Å)	11.9	11.6	9.6	9.9	10.5
10 structures - RMSD (Å)	0.71	0.72	0.77	0.64	0.56

Figure 2: Effect of including more distance restraints on SL4 structure

(A) The average structure, from the lowest 10 energy, is shown for each restraint dataset. **(B)** Statistics are given for each dataset for the NOE structure and RDC-refined structure.

NOE structures

Dataset	Shift	Slide	Rise	Tilt	Roll	Twist
1	-0.2±1.1	-1.2±0.9	3.6±0.6	0.1±4.1	9.2±6.2	30.3±4.6
2	-0.1±0.9	-1.2±0.9	3.7±0.9	0.5±4.9	9.3±4.7	31.1±5.2
3	0.0±1.1	-1.8±0.9	3.5±0.7	1.1±5.6	8.7±5.7	31.4±6.1
4	-0.2±0.9	-1.8±0.8	3.5±0.7	-1.2±1.7	9.7±4.3	32.1±7.5
5	-0.1±0.9	-2.0±0.7	3.3±0.6	0.5±4.5	7.8±5.9	32.2±5.6

RDC refined structures

Dataset	Shift	Slide	Rise	Tilt	Roll	Twist
1	-0.4±1.0	-1.1±0.8	3.1±0.4	-1.4±2.9	11.1±6.4	33.3±4.9
2	-0.5±0.8	-1.2±0.8	3.1±0.4	-1.0±2.9	11.3±8.0	32.5±6.8
3	-0.4±0.7	-1.5±0.7	2.9±0.5	-1.0±2.5	11.6±8.4	32.2±7.4
4	-0.4±0.6	-1.4±0.7	3.0±0.4	-0.8±3.5	11.6±7.1	32.3±7.5
5	-0.5±0.8	-1.5±0.6	2.8±0.5	-1.0±3.3	10.9±8.3	32.2±7.4

A-form RNA

A-form	0.0±0.6	-1.0±0.5	3.0±0.3	0.0±3.0	9.0±4.5	31.5+2.0

Table 2: Helical parameters determined using mock calculations of SL4

For the NOE and RDC refined structures, the helical parameters shift, slide, rise, tilt, roll and twist are indicated for the given dataset with average and standard deviation, compared to those of an A-form RNA helix.

eleven distance restraints were changed from 5.5Å to 4.5Å, and two distance restraints were added, in a total of 590 original distance restraints. For SL4, twelve distance restraints were changed from 5.5Å to 4.5Å, and three distance restraints were added, in a total of 490 original distance restraints. The structures and statistic results are depicted in Figure 3. Analysis of the re-determined structures shows SL2 major groove width decreased to 14.0Å from 25.4Å, and SL4 major groove width decreased to 14.5Å from 26.0Å. The large change in groove width from changing few distance restraints indicates the importance of including and accurately measuring these values. The appearance of the structures is indeed more A-form, as seen in Figure 3.

To determine if the structures re-calculated with modified distance restraints were more accurate, the NOE structures were analyzed to see how well they fit the RDCs. Back-calculated RDCs from the NOE structures were plotted against the experimentally measured RDCs (Fig. 4). For SL2, the original structures had an RDC-R^2 = 0.53 (Q = 83.3%), versus the re- calculated structures which have an RDC-R^2 = 0.93 (Q=24.5%). For SL4, the original data had RDC-R^2 = 0.60 (Q=75.5%), versus the re- calculated structures which have an RDC-R^2 = 0.87 (Q=31.5%). From the significantly improved RDC fitting and Q values, we conclude that the new structures with the modified sequential base-to-base distances more accurately reflect the solution structure and provide better starting models for the next step of the structure calculation, RDC refinement.

To further investigate the influence of sequential base-to-base distances we modified these values in the mock data to examine the effects on the resulting structures. First, we did a calculation with Dataset 3, modifying only the sequential base-to-base distances to those used by Richards et al., which contain inconsistent values (5 values = 4.5Å, 12 values = 5.5Å, 4 values are missing). With this data, the calculated structure became significantly elongated, with major groove of 20.8Å and RDC-R^2 = 0.71 (compared to 14.5Å and RDC-R^2 = 0.87 with the ideal data). To determine if more distance restraints could compensate for the missing and incorrect sequential base-to-base distances, we then performed a similar modification of Dataset 5, using the sequential base-to-base values of Richards et al. This resulted in a marked improvement, however the structure has an average major groove width of 16.7Å and RDC-R^2 = 0.79

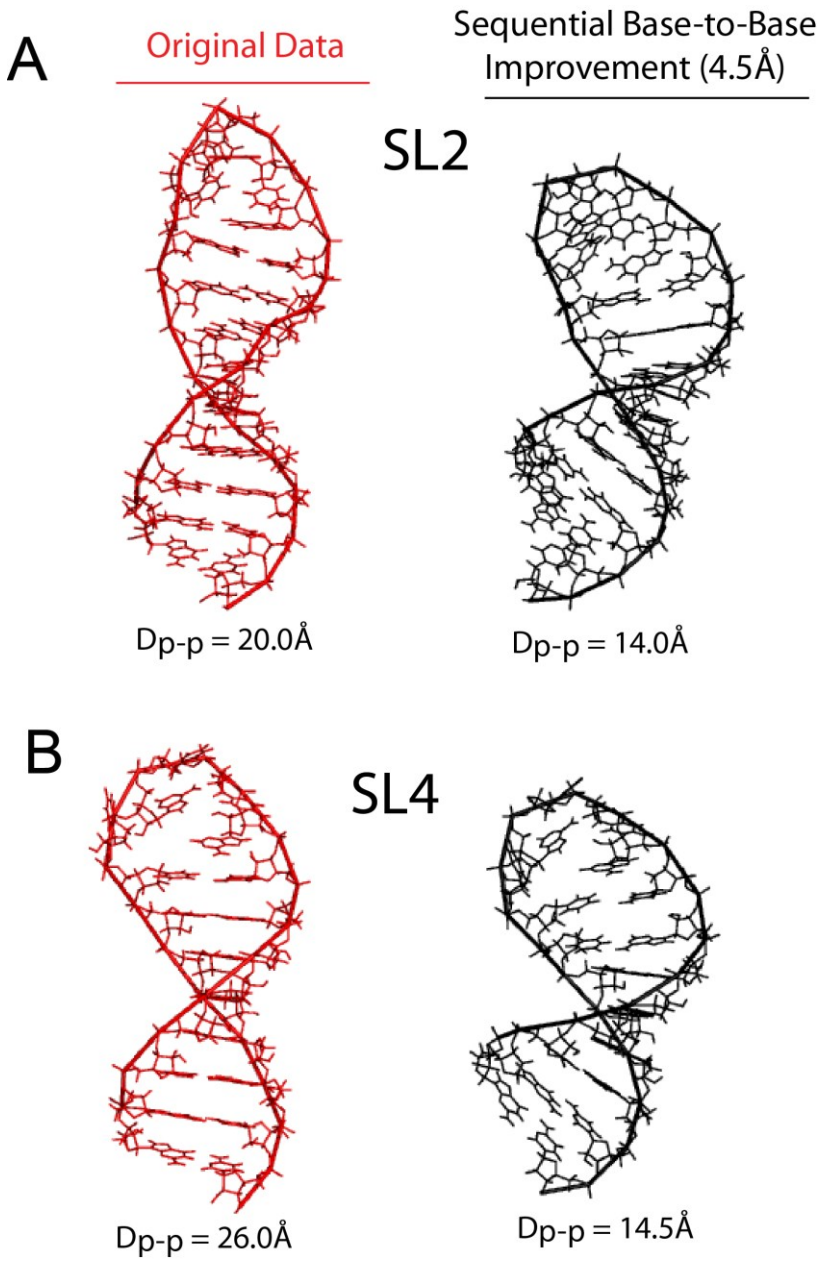


Figure 3: Influence of sequential base-to-base distance on structure

For these structure calculations, the original distance restraint list was used, except the sequential base-to-base distances were either left unchanged (Original data, red) or modified to the correct values **(**4.5Å, black). **(A)** SL2 and **(B)** SL4 average structures (from lowest 10 energy structures) are shown with corresponding major groove width ($D_{p\text{-}p}$).

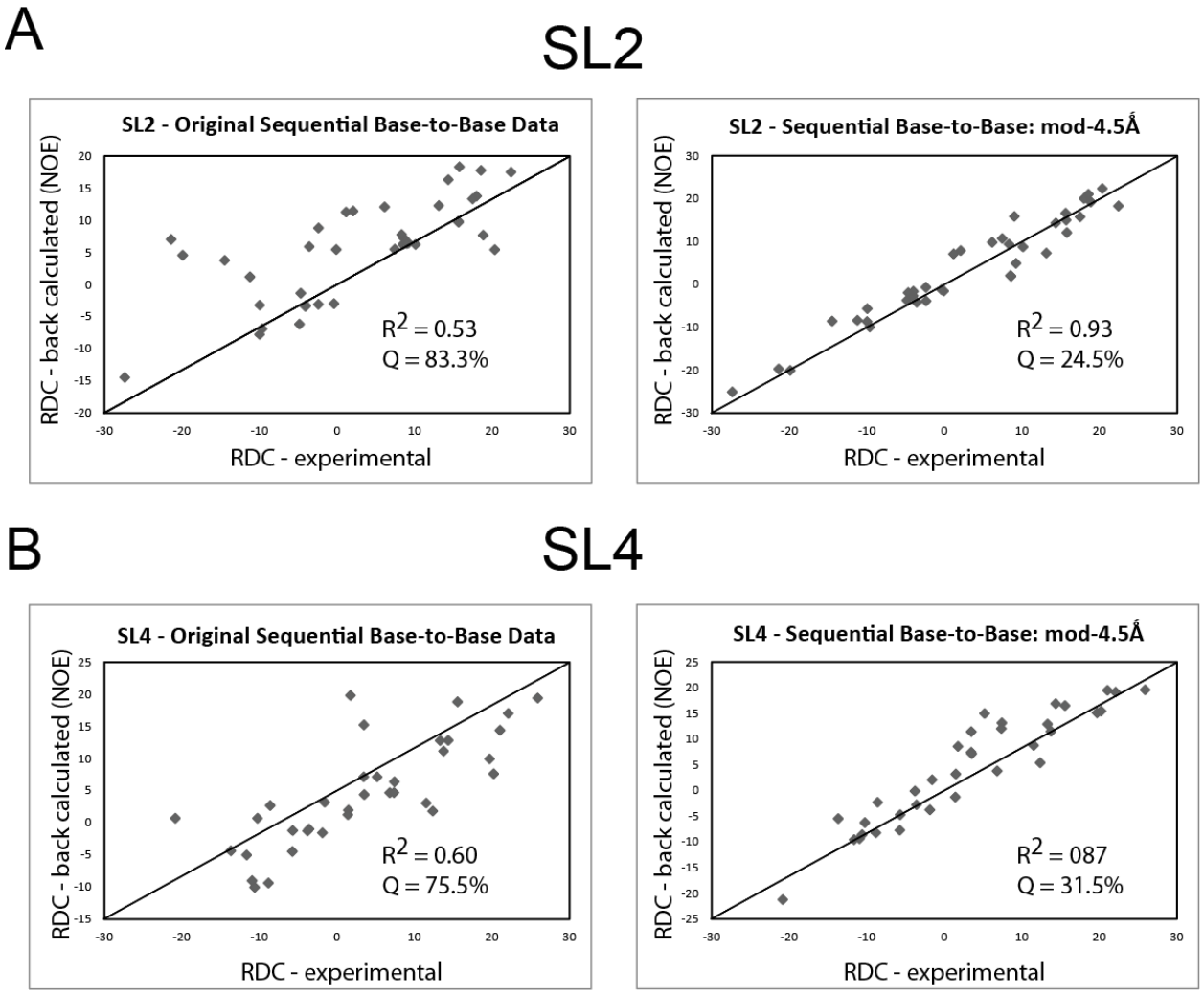


Figure 4. Fitting of NOE-based structures to RDCs

For **(A)** SL2 and **(B)** SL4, back calculated RDCs for NOE-based structures are plotted against experimental RDCs. The R^2 for RDC fitting, and Q values are indicated.

(compared to 11.6Å and RDC-R^2 = 0.91 with the ideal data). Taken together, this data reveals that the distance between base protons (H5, H6, H8) in sequential nucleotides provide essential information for NOE-based structure calculations. They serve as an indicator for A-form geometry (or lack thereof) and should be thoroughly assigned and quantified during the NOESY-distance analysis due to their impact on the structure.

RDC refinement

In the original calculations of the TER SL2 and SL4 structures, the RDC refinement step did not improve the structures as it should. Instead, RDC refinement increased the major groove width of SL2 from 20.0Å to 26.7Å, and of SL4 from 26.0Å to 28.6Å. We determined that this was due to inaccurate measurement of RDCs, as well as an inadequate number of RDCs. The alignment media used was an organic solvent (3% C12E6/hexanol) and was insufficient in aligning the RNA molecules resulting in RDC values in the range of ±2Hz. This is within the error of RDC measurement and thus does not allow for proper definition of the global bond vector orientations during refinement. The RDC analysis used only base bond vectors (C6-H6, C8-H8, C5-H5, and C2-H2) and did not include those of the sugar (C1′-H1′) and imino (N-H). The base bond vectors all lie in a similar plane with respect to the helical axis and thus give redundant information for defining the global fold. In contrast, C1′-H1′ bonds lie roughly perpendicular to the bases and provide a unique dataset for orientation analysis. A new set RDCs was measured with Pf1 bacteriophage at a concentration of 10-15 mg/mL (dissolved in NMR buffer) (*43*). These RDCs were in the range of ±30Hz and were measured for base, sugar, and imino bonds. Using the new more complete set of RDCs we first refined the original SL2 and SL4 NOE structures (using the original distance data). SL2 major groove width decreased to 14.1Å from 20.0Å and SL4 major groove width decreased to 15.8Å from 26.0Å (Fig. 5). Next, the new set of RDCs was used to refine the NOE structures from the calculations using the corrected sequential base-to-base distances. Together, the newly determined restraints resulted in

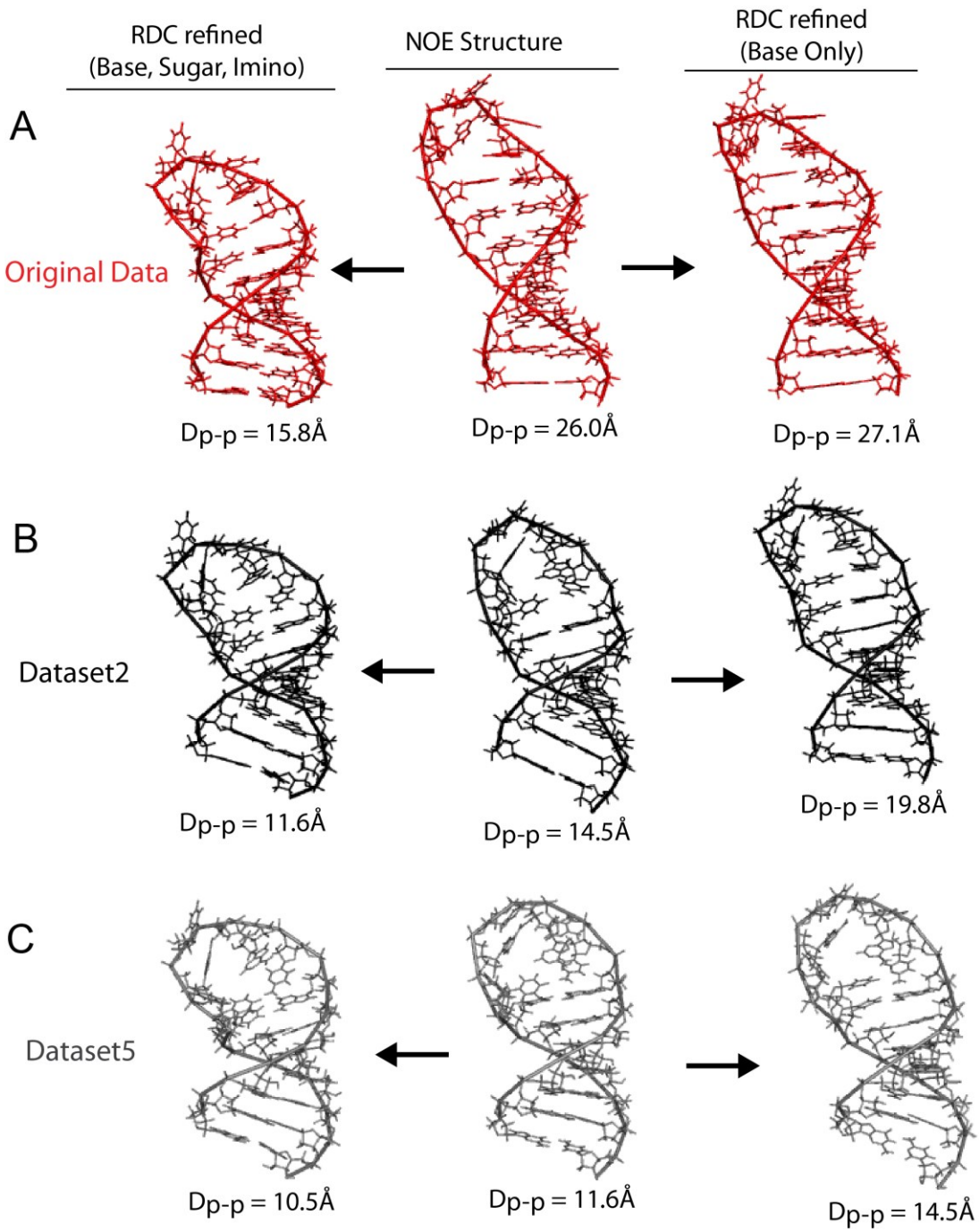


Figure 5. Non-redundant bond types are essential for RDC refinement

NOE structures were calculated using the **(A)** Original data, red, **(B)** Dataset 2, black, or **(C)** Dataset 5, gray. The NOE structures (middle) were then refined using the entire RDC data (left) or base RDC data only (right). Resulting major groove widths are indicated ($D_{p\text{-}p}$).

average major groove widths of 13.5Å and 12.4Å for SL2 and SL4 respectively. The accurate determination of RDC and distance restraints leads to a significant improvement in the RNA structures. Importantly, RDCs improve structures even with limited sets of distance data.

To confirm the importance of using non-redundant bond types (base, sugar, imino) in RDC analysis, we performed RDC refinement of SL4 using only the base RDC data from the newly measured RDCs. The new base RDC data has a range of -2 to +26Hz. With the original distance data, RDC refinement with only base RDCs increased the major groove width of the NOE structure from 26.0Å to 27.1 Å, compared to 15.8Å when refining with all RDCs (Fig. 5). To see if more distances restraints could improve the outcome we did similar calculations using the mock datasets with ideal values. Dataset 2 started with an NOE structure of average major groove width for NMR calculations, 14.5Å, but increased to 19.8Å after base RDC refinement, compared to 11.6Å with all RDCs. Furthermore, using the maximum amount of distance data, Dataset 5, resulted in a similar increase of the major groove width of the NOE structure which started at 11.6Å, then expanded to 14.5Å when refining with only base RDCs (compared to 10.5Å with all RDCs) (Fig. 5).

Investigation of the (RCSB) deposited NMR structures included in the Tolbert et al. (*21*) analysis revealed three additional structures which had been calculated using only base RDCs, PDB: 2QH2, 2QH3, and 2QH4 (*44*). The RDCs for these structures were determined from samples in C12E6/hexanol as alignment media, which gave values in the range of ±15Hz. However, only the base RDCs were included in the structure calculations. As a result, the RDC refined structures have major groove widths ranging from 18 - 21Å. With limited RDC information, a variety of structures are able to fit the RDC data, resulting in the proliferation of extended helices. Taken together this indicates the importance of not only properly measuring RDCs, but that non-redundant RDC data must be used to accurately define the global fold during refinement.

Re-analysis of deposited RNA structures

The analysis by Tolbert et al. was done in a general manner including all helical RNA structures with greater than 7 base pairs determined by NMR deposited between 2005 and 2009 (*21*). While this may give a broad view of how NMR performs in calculating structures, it is important to further analyze the data based on the different types of NMR methodology as well as to pay attention to which RNA structures are not expected to be A-form. If the erroneous structures (described above) and non A-form structures (subsequently described below) are separated from the analysis, and the RDC refined structures are evaluated, the NMR statistics improve greatly with a 10.1 - 17.1Å major groove width range, and 13.7Å average. The structures without RDCs are subject to having extended helices independent of the NMR calculation method. Without RDCs, AMBER refined structures have major groove widths of 9.3 - 21.9Å (avg = 14.9Å) and Xplor/CNS structures have major groove widths of 7.4 - 25.6Å (avg = 15.6Å). This further supports the conclusion that RDCs are critical in achieving accurate nucleic acid structures.

Finally, it is important to identify and analyze accordingly the RNA structures which are not expected to form A-form helices throughout. These are RNAs which contain numerous bulges, mismatches, and internal loops. This includes the classical swine fever virus (CSFV) internal ribosome entry site (IRES) RNA (PDB: 2HUA) which contains a 5 nt bulge, and the *Tetrahymena* TER Stem 4 (PDB: 2FEY) which contains 4 bulge residues, both of which were included in the analysis of Tolbert et al. (*45, 46*). The non-base paired elements of these RNAs are not expected to have A-form geometry, and thus should not have been used as an indicator of accuracy for NMR determined structures. Nevertheless, these structures contained RDCs and a simple way to observe the helical characteristics correctly is to align an A-form helix to the parts of the structures without bulges (Fig. 6). As can be seen with this alignment, the helical parts of these structures conform very well to A-form, despite having reported major groove widths of 20.0Å and 21.9Å for CSFV IRES and *Tetrahymena* TER Stem 4 respectively.

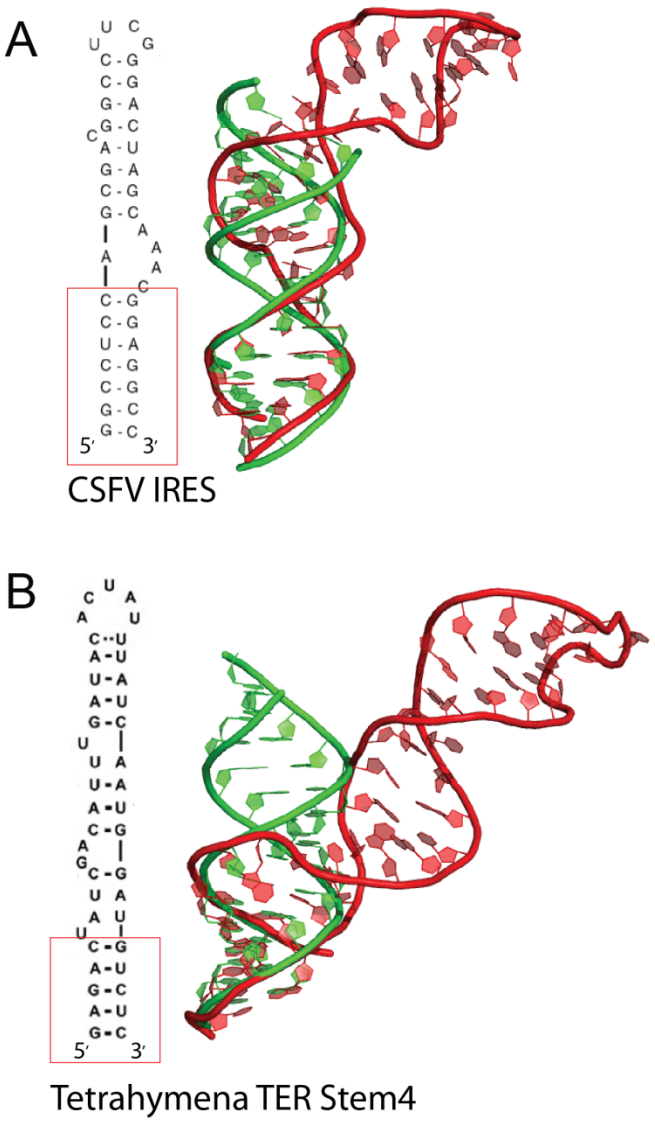


Figure 6: Fitting A-form helices to bulged RNA structures.

Secondary structure diagrams of **(A)** classical swine fever virus internal ribosome entry site (CSFV IRES) RNA and **(B)** *Tetrahymena* TER Stem 4. The Watson-Crick stems (red box) of the NMR structures of the RNAs (red) are fit to A-form helices (green).

Fitting of improved TER structures to EM structure

Jiang et al. determined the negative stain EM structure of the *Tetrahymena* telomerase holoenzyme at a resolution of ~25Å (Fig. 7) (*25*). The subunits of the particle were localized using affinity labelling techniques with Flag labelled proteins (telomerase components) recognized by anti-Flag antigen (Fab). TER SL 2 was localized by using a *Tetrahymena* strain in which SL2 is extended with a small hairpin tag (MS2hp), which can be recognized by the MS2 coat protein (MS2cp). The extra density in the EM map when compared to wildtype strains gives the approximate location of SL2. Upon locating all of the subunits, high-resolution structures were fit into the EM map to create a detailed model of the catalytic core (Fig. 7). The re-determined structures of SL2 and SL4 were used in this telomerase model. Both of these TER elements interact with the TRBD (RNA binding domain) of TERT. TRBD bridges the two stem-loops which are ~40Å apart, bringing SL4 close to the pseudoknot which is consistent with biochemical data (*47-50*). As higher resolution structures of the holoenzyme become available, more detailed interactions will be revealed, and we will be able to see how well the solution structures match the EM structure.

Conclusion

We have further validated the importance of RDCs in NMR structure determination of nucleic acids. We show that given a minimal set of distance restraints, a non-redundant and sufficient set of RDCs allows for accurate refinement of the structure. However, more distance data increases structural accuracy and precision. An important first step in the NMR structure calculation process is to achieve a sufficient NOE structure to which RDCs can be applied. To this end, we have ascertained the critical importance of sequential base-to-base distance restraints in defining helical structure. Importantly, we find that even RDC-refined NMR structures have slightly larger average major grooves (on average) compared to crystal structures (~13Å compared to 11Å). Further experiments and analyses could confirm whether or not this is an actual difference between solution and crystal structures of nucleic acids.

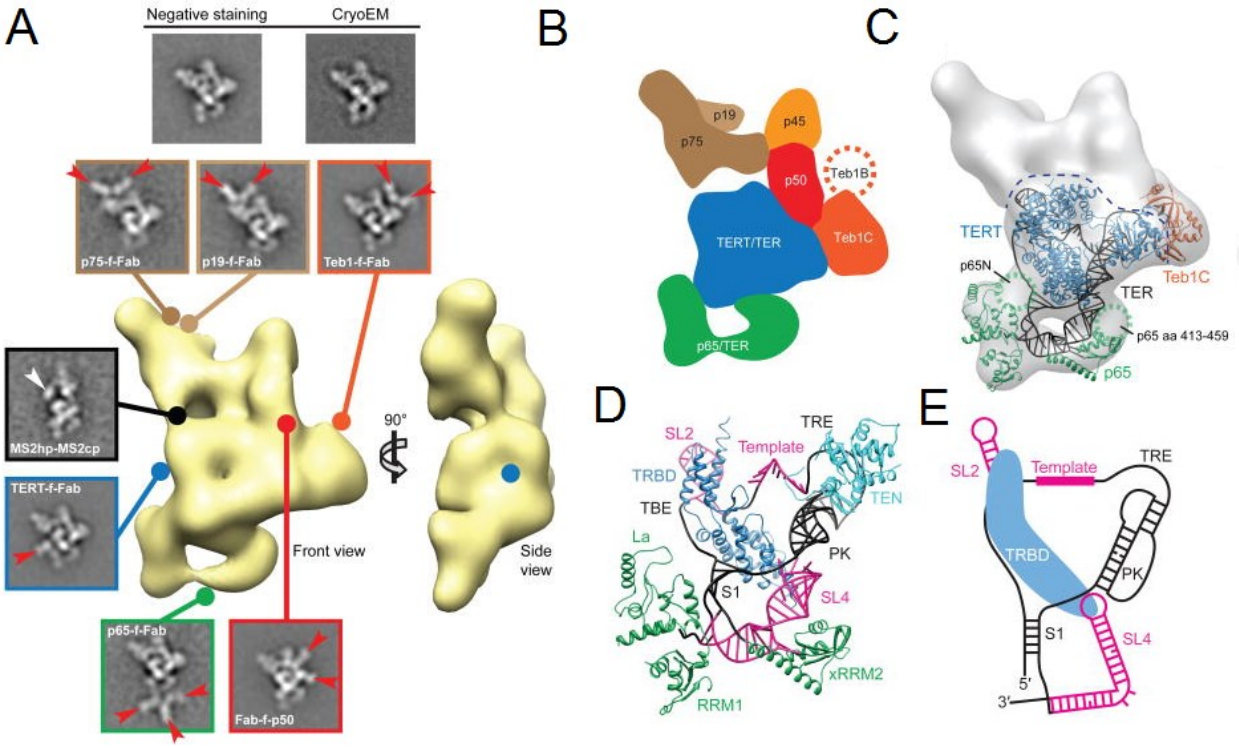


Figure 7: *Tetrahymena* telomerase holoenzyme, from Jiang et al. (*25*)

(A) (Top) Representative class averages of negative staining electron microscopy and cryoelectron microscopy images of TERT–Flag (F) telomerase. (Bottom) 3D reconstruction of Teb1–F telomerase (front and side views) and class averages of affinity-labelled telomerase particles. Lines with circle heads indicate attachment point of Fab (red arrows) and MS2cp (white arrow). **(B)** Telomerase subunit schematic (front view). **(C)** 3D reconstruction of Teb1–F telomerase with TERT (blue), p65 (green) and TER (black), plus Teb1C (orange) modelled into the electron microscopy density. **(D)** TER model structure (well determined = magenta; remaining = black) and interactions with TERT TRBD (blue) and TEN (cyan), and p65 La, RRM1, and xRRM2 domains (green). **(E)** Secondary structure schematic of TER with TRBD.

References

1. Crosio, M. P., Rodier, F., and Jullien, M. (1990) Packing forces in ribonuclease crystals, *FEBS Lett 271*, 152-156.

2. Shah, S. A., and Brunger, A. T. (1999) The 1.8 A crystal structure of a statically disordered 17 base-pair RNA duplex: principles of RNA crystal packing and its effect on nucleic acid structure, *J Mol Biol 285*, 1577-1588.

3. Allain, F. H., and Varani, G. (1997) How accurately and precisely can RNA structure be determined by NMR?, *J Mol Biol 267*, 338-351.

4. Lukavsky, P. J., and Puglisi, J. D. (2005) Structure determination of large biological RNAs, *Methods Enzymol 394*, 399-416.

5. Cromsigt, J. A., Schleucher, J., Kidd-Ljunggren, K., and Wijmenga, S. S. (2000) Synthesis of specifically deuterated nucleotides for NMR studies on RNA, *J Biomol Struct Dyn 17 Suppl 1*, 211-219.

6. Duss, O., Lukavsky, P. J., and Allain, F. H. (2012) Isotope labeling and segmental labeling of larger RNAs for NMR structural studies, *Adv Exp Med Biol 992*, 121-144.

7. Kim, I., Watanabe, S., Muto, Y., Hosono, K., Takai, K., Takaku, H., Kawai, G., Watanabe, K., and Yokoyama, S. (1995) Selective deuteration of RNA for NMR signal assignment, *Nucleic Acids Symp Ser*, 123-124.

8. Batey, R. T., Inada, M., Kujawinski, E., Puglisi, J. D., and Williamson, J. R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA, *Nucleic Acids Res 20*, 4515-4523.

9. Nikonowicz, E. P., and Pardi, A. (1992) Three-dimensional heteronuclear NMR studies of RNA, *Nature 355*, 184-186.

10. Nikonowicz, E. P., and Pardi, A. (1993) An efficient procedure for assignment of the proton, carbon and nitrogen resonances in 13C/15N labeled nucleic acids, *J Mol Biol 232*, 1141-1156.

11. Peterson, R. D., Theimer, C. A., Wu, H., and Feigon, J. (2004) New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes, *J. Biomol. NMR 28*, 59-67.

12. Al-Hashimi, H. M., Gorin, A., Majumdar, A., and Patel, D. J. (2001) Alignment of the HTLV-I Rex peptide bound to its target RNA aptamer from magnetic field-induced residual dipolar couplings and intermolecular hydrogen bonds, *J Am Chem Soc 123*, 3179-3180.

13. Al-Hashimi, H. M., Gosser, Y., Gorin, A., Hu, W., Majumdar, A., and Patel, D. J. (2002) Concerted motions in HIV-1 TAR RNA may allow access to bound state conformations: RNA dynamics from NMR residual dipolar couplings, *J Mol Biol 315*, 95-102.

14. Al-Hashimi, H. M., and Patel, D. J. (2002) Residual dipolar couplings: synergy between NMR and structural genomics, *J Biomol NMR 22*, 1-8.

15. D'Souza, V., Dey, A., Habib, D., and Summers, M. F. (2004) NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus, *J Mol Biol 337*, 427-442.

16. Kim, N. K., Zhang, Q., Zhou, J., Theimer, C. A., Peterson, R. D., and Feigon, J. (2008) Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA, *J Mol Biol 384*, 1249-1261.

17. Zhang, Q., Kim, N. K., and Feigon, J. (2011) Architecture of human telomerase RNA, *Proc Natl Acad Sci U S A 108*, 20325-20332.

18. Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science 278*, 1111-1114.

19. Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol. 5*, 1065-1074.

20. Douglas, S. M., Chou, J. J., and Shih, W. M. (2007) DNA-nanotube-induced alignment of membrane proteins for NMR structure determination, *Proc Natl Acad Sci U S A 104*, 6644-6648.

21. Tolbert, B. S., Miyazaki, Y., Barton, S., Kinde, B., Starck, P., Singh, R., Bax, A., Case, D. A., and Summers, M. F. (2010) Major groove width variations in RNA structures determined by NMR and impact of 13C residual chemical shift anisotropy and 1H-13C residual dipolar coupling on refinement, *J Biomol NMR 47*, 205-219.

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

23. Richards, R. J., Wu, H., Trantirek, L., O'Connor, C. M., Collins, K., and Feigon, J. (2006) Structural study of elements of Tetrahymena telomerase RNA stem-loop IV domain important for function, *Rna-a Publication of the Rna Society 12*, 1475-1485.

24. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package, *J Magn Reson 160*, 65-73.

25. Jiang, J., Miracco, E. J., Hong, K., Eckert, B., Chan, H., Cash, D. D., Min, B., Zhou, Z. H., Collins, K., and Feigon, J. (2013) The architecture of Tetrahymena telomerase holoenzyme, *Nature 496*, 187-192.

26. Schwalbe, H. (2003) Kurt Wuthrich, the ETH Zurich, and the development of NMR spectroscopy for the investigation of structure, dynamics, and folding of proteins, *Chembiochem 4*, 135-142.

27. Nozinovic, S., Furtig, B., Jonker, H. R., Richter, C., and Schwalbe, H. (2009) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA, *Nucleic Acids Res 38*, 683-694.

28. Wu, H., Finger, L. D., and Feigon, J. (2005) Structure determination of protein/RNA complexes by NMR, *Methods Enzymol 394*, 525-545.

29. Zidek, L., Stefl, R., and Sklenar, V. (2001) NMR methodology for the study of nucleic acids, *Curr Opin Struct Biol 11*, 275-281.

30. Furtig, B., Richter, C., Wohnert, J., and Schwalbe, H. (2003) NMR spectroscopy of RNA, *Chembiochem 4*, 936-962.

31. Scott, L. G., and Hennig, M. (2008) RNA structure determination by NMR, *Methods Mol Biol 452*, 29-61.

32. Clore, G. M., and Kuszewski, J. (2003) Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation, *J Am Chem Soc 125*, 1518-1525.

33. Guntert, P. (2004) Automated NMR structure calculation with CYANA, *Methods Mol Biol 278*, 353-378.

34. Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J Comput Chem 26*, 1668-1688.

35. Clore, G. M., Murphy, E. C., Gronenborn, A. M., and Bax, A. (1998) Determination of three-bond 1H3'-31P couplings in nucleic acids and protein-nucleic acid complexes by quantitative J correlation spectroscopy, *J Magn Reson 134*, 164-167.

36. Richter, C. A., Wright-Osment, M. K., Zajicek, J. L., Honeyfield, D. C., and Tillitt, D. E. (2009) Quantitative polymerase chain reaction (PCR) assays for a bacterial thiaminase I gene and the thiaminase-producing bacterium Paenibacillus thiaminolyticus, *J Aquat Anim Health 21*, 229-238.

37. Gochin, M., Zon, G., and James, T. L. (1990) Two-dimensional COSY and two-dimensional NOE spectroscopy of d(AC)4.d(GT)4: extraction of structural constraints, *Biochemistry 29*, 11161-11171.

38. Dingley, A. J., and Grzesiek, S. (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings, *J. Am. Chem. Soc. 120*, 8293-7.

39. Dingley, A. J., Nisius, L., Cordier, F., and Grzesiek, S. (2008) Direct detection of N-H[...]N hydrogen bonds in biomolecules by NMR spectroscopy, *Nat Protoc 3*, 242-248.

40. Chary, K. V., Modi, S., Hosur, R. V., Govil, G., Chen, C. Q., and Miles, H. T. (1989) Quantification of DNA structure from NMR data: conformation of d-ACATCGATGT, *Biochemistry 28*, 5240-5249.

41. Gawrisch, K., Eldho, N. V., and Polozov, I. V. (2002) Novel NMR tools to study structure and dynamics of biomembranes, *Chem Phys Lipids 116*, 135-151.

42. Bax, A. (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics, *Protein Sci 12*, 1-16.

43. Trempe, J. F., Morin, F. G., Xia, Z., Marchessault, R. H., and Gehring, K. (2002) Characterization of polyacrylamide-stabilized Pfl phage liquid crystals for protein NMR spectroscopy, *J Biomol NMR 22*, 83-87.

44. Theimer, C. A., Jady, B. E., Chim, N., Richard, P., Breece, K. E., Kiss, T., and Feigon, J. (2007) Structural and functional characterization of human telomerase RNA processing and Cajal body localization signals, *Mol. Cell 27*, 869-881.

45. Chen, Y., Fender, J., Legassie, J. D., Jarstfer, M. B., Bryan, T. M., and Varani, G. (2006) Structure of stem-loop IV of Tetrahymena telomerase RNA, *Embo Journal 25*, 3156.

46. Locker, N., Easton, L. E., and Lukavsky, P. J. (2007) HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly, *EMBO J 26*, 795-805.

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

48. Cole, D. I., Legassie, J. D., Bonifacio, L. N., Sekaran, V. G., Ding, F., Dokholyan, N. V., and Jarstfer, M. B. (2012) New models of Tetrahymena telomerase RNA from experimentally derived constraints and modeling, *J Am Chem Soc 134*, 20070-20080.

49. Egan, E. D., and Collins, K.(2012) Biogenesis of telomerase ribonucleoproteins, *RNA 18*,1747-1759.

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

CHAPTER 5

Diverse combination ligation independent cloning

Introduction

A significant bottleneck in the field of protein structural biology is obtaining suitable amounts of the desired protein for study (*1*). In the case of the more common structure analysis methods, X-ray crystallography and NMR, the amount of protein necessary is typically milligram quantities. Expression systems for recombinant proteins include bacteria (*E. coli*), yeast (*Pistia pastoris*), insect cells, mammalian cells, and cell-free (using whole cell extracts) (*2*). Bacterial expression systems are the most commonly used due to simplicity, speed of culture growth, low cost and high yields. Disadvantages include codon usage and lack of translational modifications and chaperones for molecular folding for eukaryotic proteins (*3*). Recombinant protein expression involves transforming cells with a plasmid carrying the gene of interest (*4*). Traditional methods for building expression vectors use PCR to amplify the gene, adding specific restriction enzyme sites to the 5′ and 3′ ends, followed by endonuclease digestion to allow sticky-end (or blunt-end) insertion into the desired plasmid (*5, 6*). The vector is digested as well, and commonly treated with alkaline phosphatase to remove the 5′ phosphate to prevent it from being able to re-ligate. The insert and vector can then be ligated together with T4 DNA ligase, transformed into *E. coli,* and screened for expression through SDS-PAGE analysis (*7*). In 1990, a new procedure was developed which simplified cloning by eliminating the reliance on restriction enzyme sites and ligation, termed ligation independent cloning (LIC) (*8*). LIC uses the 3′ > 5′ exonuclease activity of T4 DNA polymerase to create sticky ends of PCR products and vector (Fig. 1). Since the sticky ends of LIC are longer than those of restriction enzyme sites (10-15 nt compared to 2-4 nt for restriction enzymes), the re-annealed insert/vector is stable enough to be transformed into *E. coli* without ligation. The nicks in the plasmid are subsequently filled in by the *E. coli* DNA repair machinery. This approached provided a simple, fast, and efficient cloning method which was quickly commercialized (*9*).

To facilitate cloning of difficult to express or insoluble proteins I have developed vectors which consolidate the most advantageous features of fusion tag technology for enhancement of protein expression, solubility, screening and purification. This new system, termed diverse

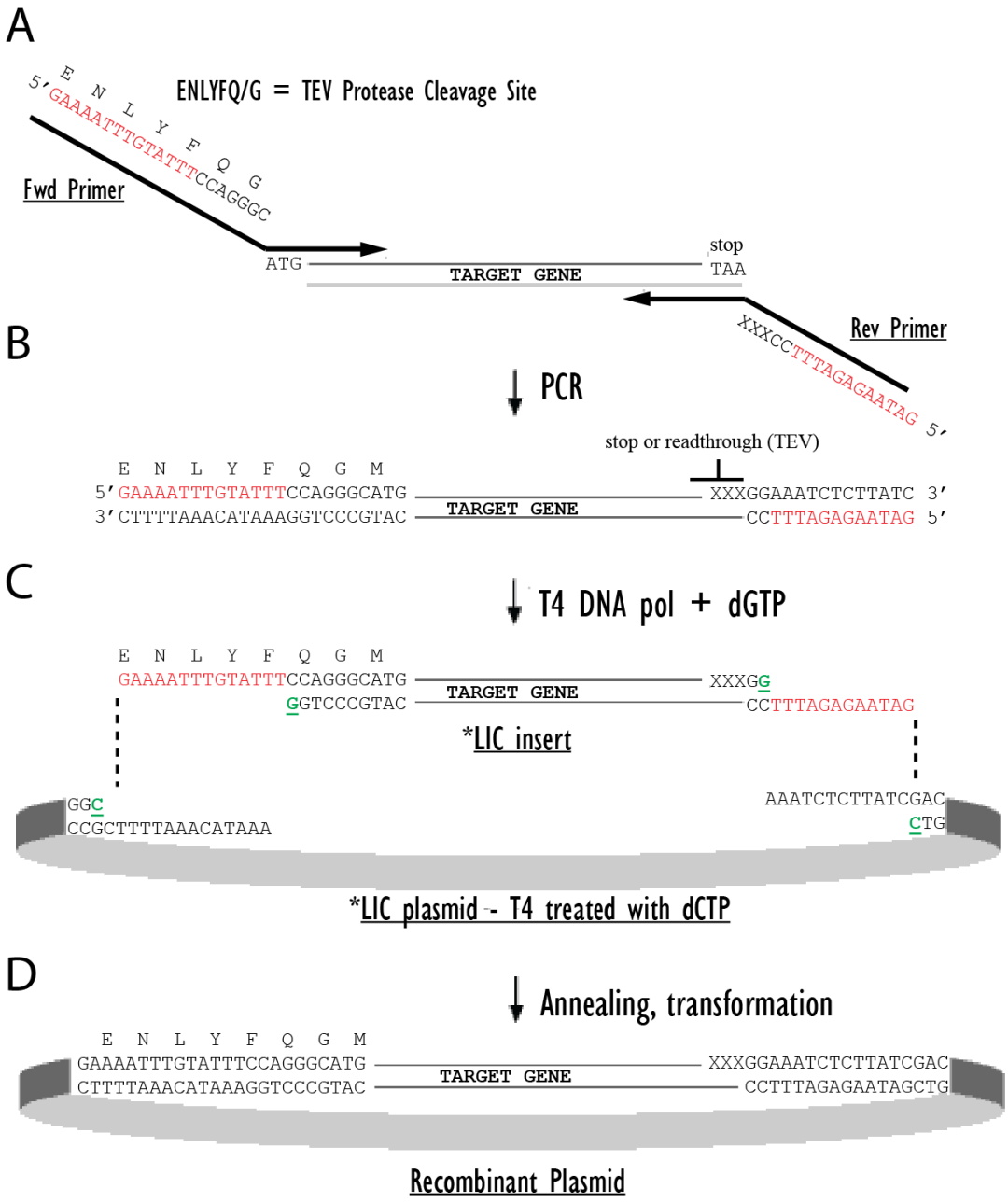


Figure 1: LIC cloning schematic

(A) Components of PCR, including forward (fwd) and reverse (rev) primers, and template (target gene). **(B)** PCR product with added LIC sequences. The 3′ end of the gene may encode a stop codon or readthrough sequence (i.e. TEV site) to add C-terminal tag. **(C)** T4 DNA polymerase treatment with corresponding dNTP (green) creates matching overhangs in the insert and vector. **(D)** Insert and vector are annealed to create the final recombinant plasmid for *E. coli* expression.

combination LIC (DC-LIC), allows for efficient assessment of multiple fusion partners with one PCR product. DC-LIC has been successfully applied to improve the expression of a variety of telomerase related proteins (from *Tetrahymena*, human and medaka fish), proving to be a cost-effective, high throughput and robust cloning system.

DC-LIC design and advantages

The most cost-effective LIC vectors are available as intact plasmids. This allows them to be easily reproduced through replication in *E. coli*, followed by simple plasmid purification protocols to supply an endless source of vector material. The circular plasmid is then treated with a blunt restriction enzyme, gel purified and T4 DNA polymerase treated for use in the LIC protocol. Plasmid libraries, such as AddGene, have similar LIC vectors. However they contain a low variety of vectors, utilizing variable LIC site sequences which require different PCR products.

The DC-LIC system aims to achieve a more complete cloning approach. First, DC-LIC utilizes one LIC site sequence, meaning that one PCR product can be used to test a variety of different vectors, reducing cloning costs and time. Second, the tobacco etch virus (TEV) protease cleavage site is incorporated into the LIC sequence so that the TEV site will be adjacent to the target protein. This minimizes the number of additional amino acids added to the target protein upon cleavage, which has advantages in subsequent biochemical experiments, such as protein crystallization. Furthermore, TEV protease is able to be easily produced in research labs, has highly specific cleavage, and functions under a wide range of conditions (*10*). If desired, different protease sites can be introduced via PCR primers, such as enterokinase or thrombin (*11*). Importantly, the LIC sequence is designed so that the N-terminal TEV cleavage site is optional, since the last three amino acids of the TEV site must be added through PCR primers (Fig. 1). This can be taken advantage of in situations where the researcher wants to cleave the C-terminal tag, while leaving the N-terminal tag intact (for solubility).

DC-LIC fusion tags

The 6x tandem histidine (His6) tag is a standard tag used in purifying recombinant proteins, utilizing immobilized metal-ion affinity chromatography (IMAC). IMAC is a robust purification method in which matrixed nickel gives high affinity and yields, while matrixed cobalt gives increased specificity and purity. The strong ionic interaction allows for purification in the presence of reagents (i.e. detergents) which may interfere with other affinity matrices. IMAC is also stable in protein under denaturing conditions. IMAC resin is relatively inexpensive compared to other chromatography media, and can be regenerated a large number of times. Furthermore, the small size of the His6 tag affords efficient purification and does not necessarily have to be cleaved off due to its minimal effects on protein folding, structure, and function. For these reasons, the His6 tag is a staple in the DC-LIC system. Most DC-LIC vectors contain a His6 N-terminal tag for a simple, high-throughput first purification step.

In addition to the His6 tag which assists purification, a number of fusion proteins are available to enhance protein expression and solubility. The DC-LIC system utilizes the most effective and thoroughly validated fusion tags with a spectrum of sizes, including N utilization substance A (NusA, 55kD), maltose binding protein (MBP, 40kD), glutathione S-transferase (GST, 26kD), and small ubiquitin-related modifier (SUMO, 11kD) (*12, 13*). MBP and GST also serve as affinity tags for further purification steps. In addition to the commonly used fusion tags, green fluorescent protein (GFP, 25kD) is also a part of the DC-LIC fusion library where it serves multiple roles. The variant of GFP used is known as "superfolder" (*14*). In contrast to wildtype GFP which can misfold if its fusion partner is poorly folded, superfolder GFP readily folds regardless of its fusion partner, and may even enhance folding of the passenger protein similar to other fusion tags. The diverse combination of fusion tags used in the DC-LIC system is depicted in Table 1.

GFP as a C-terminal fusion (C-GFP) also serves as an expression reporter. Explicit frameshifting and exclusion of a start codon from C-GFP ensures that only colonies expressing a properly cloned fusion protein will fluoresce (Fig. 2) (*15*). With this cloning technique, a library

N-Terminal Tags \ C-Terminal Tags	No C-tag	C-MBP	C-GST	C-GFP	C-GFP-His6
His6					n/a
His6-NUS					
His6-MBP		n/a			
His6-GST					n/a
His6-SUMO					n/a
NUS		n/a	n/a	n/a	
MBP		n/a	n/a	n/a	
GST		n/a	n/a	n/a	

Table 1: DC-LIC vectors

Table of DC-LIC vectors with different combinations of N-terminal and C-terminal fusion tags. N/a indicates vectors that were not made.

of clones can be quickly screened on LB-agar plates or in liquid culture for expression through GFP fluorescence analysis (Fig. 2). In a similar fashion, the GFP DC-LIC vectors can also be used for random PCR cloning to find soluble domains of a target protein, as utilized by the Cech lab to discover the N-terminal (TEN) domain of telomerase (*16*). In addition to positive/negative (green vs. not green) classification of C-GFP variants, the yield and solubility can be readily quantitated for comparison of expression levels by monitoring UV absorbance at 485nm. Furthermore, through UV detection C-GFP can help confirm that low-expressing constructs produce the desired product. Protein expression is typically visualized by intense bands on SDS-PAGE gels which clearly indicate the identity of the target protein. However if the protein of interest only expresses at low levels it may be hard to distinguish from native *E. coli* proteins. This could result in purification of the wrong protein of similar size or abandoning of a project due to the assumption that the protein is not expressed. In some cases, the small protein amounts are sufficient for subsequent experiments. C-GFP can assist in these cases since even small amounts of GFP appear green under (long-range) UV light and have a 485nm absorbance to confirm the presence of the target protein. Also, a non-denaturing SDS-PAGE analysis can be performed to visualize C-GFP proteins in the gel. This is done adding by adding SDS dye to the sample but not boiling it. Superfolder C-GFP remains folded through electrophoresis to allow for in-gel fluorescent visualization and detection of the target protein.

Protein and RNA co-expression

Another technique to improve protein expression is co-expression (*17*). Two proteins which potentially bind together are expressed at the same time in order to assist folding, increasing solubility and overall expression. The DC-LIC system facilitates this method by supplying vectors that are compatible with co-expression. The original vector library is all encoded on a pet-DUET (Novagen) backbone with ampicillin resistance. In addition, the basic fusion tags (His6, NusA, MBP, and GST) are available in the pRSF vector (Novagen), which has a different origin of replication and kanamycin resistance, making it suitable for co-

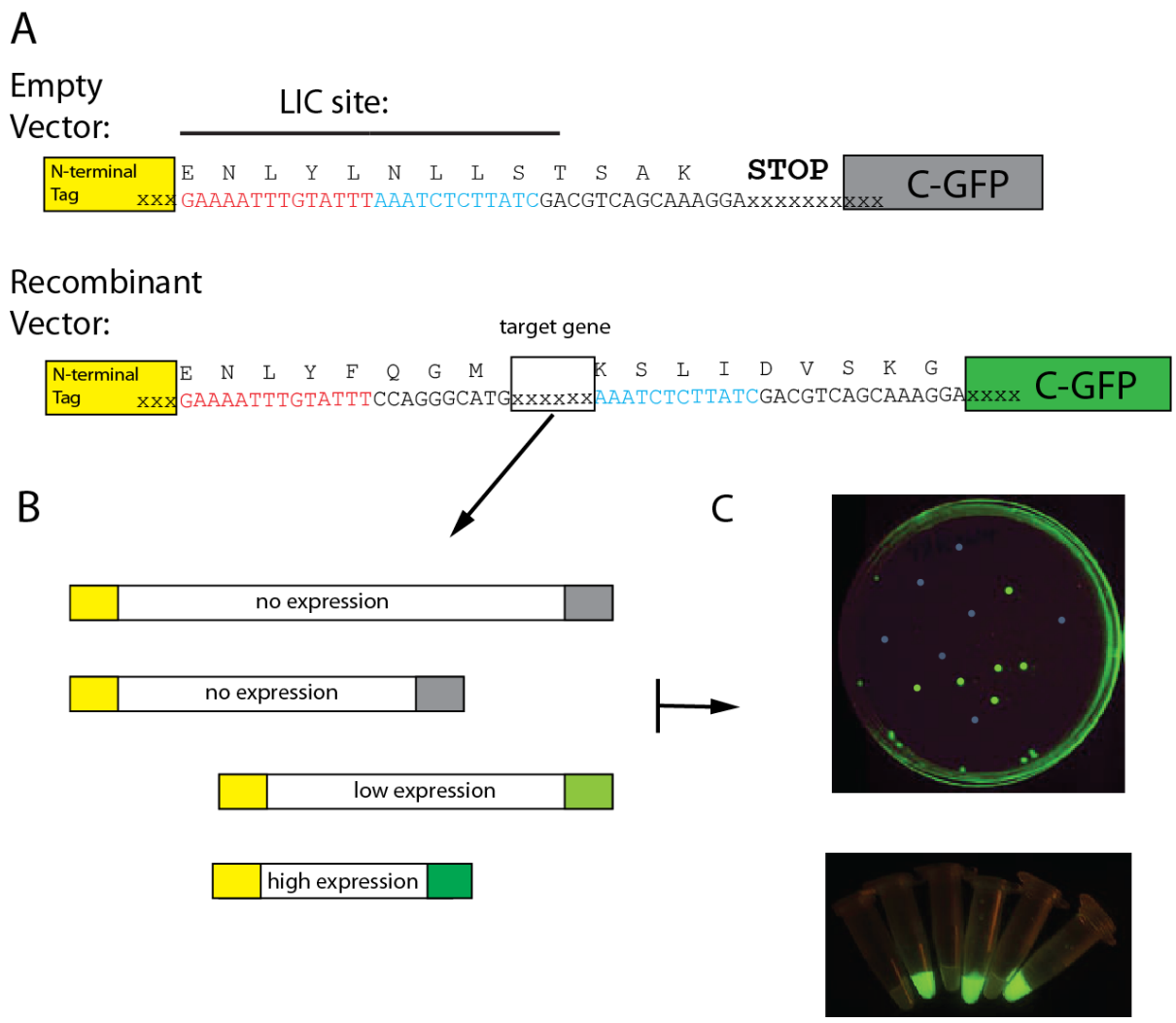


Figure 2: Strategy for C-GFP protein expression optimization

(A) Empty vector with LIC sequence displayed, where after cleavage results in 5′ (red) and 3′ (blue) halves. The cloning site is designed so that the C-terminal GFP is frame-shifted with a termination codon, so that empty vector does not express GFP. C-GFP is colored based on its expected expression/fluorescence levels, where gray indicates no expression/fluorescence and increasing green color indicates higher expression/fluorescence. **(B)** Recombinant vector with cloned insert. The cloning site is designed so that only correctly cloned genes will place C-GFP in the proper frame to express (green). **(C)** The target gene can be modified, randomly or systematically, and cloned into C-GFP vectors. The amount of GFP expression serves as an indicator for protein expression levels of each variant. **(D)** Example results for LB-agar plate (top) or liquid culture (bottom) C-GFP expression tests.

expression (*18*). The pRSF-DCLIC vectors also have the same LIC site sequence as pet-DCLIC, simplifying cloning.

An RNA co-expression vector was also made for use with RNA-binding proteins. The vector was made in the pCDF (Novagen) backbone plasmid, with streptamycin resistance, following design from published articles (*19, 20*). A tRNA scaffold is used to prevent RNA degradation by adding an enclosing tRNA motif that is processed by the cellular machinery producing indicators that protect the RNA terminal ends (*21*) (Fig. 3). The RNA expression vector employs restriction enzyme cloning methodology to allow for cloning into two possible sites, which include or exclude the optional streptavidin aptamer tag (Fig. 3) (*22*). The streptavidin aptamer is a 42 nt RNA stem-loop that has high affinity and specificity for the streptavidin protein. By incorporating this tag into the expressed RNA, the protein-RNA complex can be purified via immobilized streptavidin and eluted with biotin derivatives which have higher binding affinity than the RNA aptamer.

Examples of DC-LIC strategy

To enhance protein expression and solubility of a single protein construct the following protocol can be followed. Clone the target protein into DC-LIC vectors, NusA-(C-GFP), MBP-(C-GFP), GST-(C-GFP), and SUMO-(C-GFP). Perform an expression test by analyzing the soluble fraction by SDS-PAGE gel and quantitation by 485nm UV absorbance. After determining the N-terminal fusion partner that results in highest expression, optimize growth conditions, such as temperature, induction time, and IPTG amount with the same (SDS-PAGE, UV) analysis. Finally, optimize the C-terminus by testing C-GFP, C-GST, or no C-terminal tag. After C-GFP is used to monitor and optimize expression, replacement with C-GST provides a C-terminal affinity tag for isolation of full-length product while serving as an additional purification option.

Some proteins do not have detectable expression levels or are insoluble for a number of possible reasons (such as toxicity). In these cases it is necessary to truncate (or extend) the target

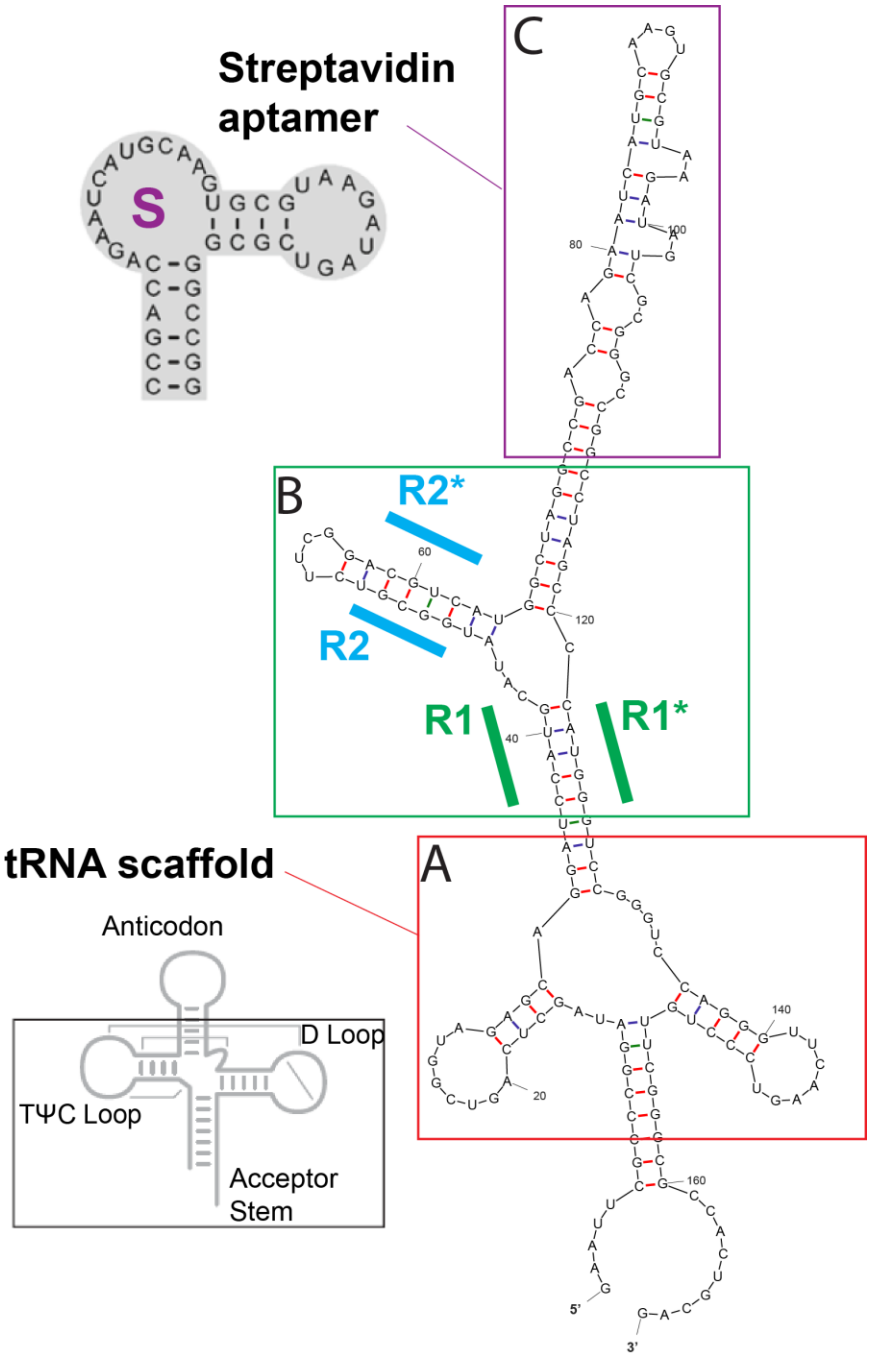


Figure 3: Vector for RNA expression in *E. coli*

(A) The tRNA scaffold assists RNA stability in the cell. The RNA of interest is cloned into the anticodon stem of the tRNA. **(B)** Two cloning sites with different restriction enzymes can be used to exclude (R1/R1*) or include (R2/R2*) the optional streptavidin aptamer when building RNA expression constructs. **(C)** The streptavidin RNA aptamer is a 42nt bulged stemloop that has high affinity for streptavidin (S), allowing for affinity purification.

protein to achieve a suitable construct. For this scenario, an optimal construct can be determined through the following protocol. Design appropriate variants of the target protein and clone them into an appropriate C-GFP vector (typically MBP-C-GFP works best in our lab). Optionally, plasmid sequence verification can be skipped to save time and cost, since C-GFP fluorescence reports properly cloned inserts (Fig. 2). If so, the annealed insert and vector mix can be transformed directly into BL21 *E. coli* cells (for protein expression). After transformation, add IPTG to the LB-agar plates. Dependent on the protein, IPTG can be added prior to colony growth or after colonies have grown. After incubation, examine the agar plates for fluorescent colonies (with UV light), which indicates that the target protein has expressed. After finding a construct that expresses on plates, perform a liquid culture expression test, separating the insoluble and soluble fractions to find the construct with best solubility through SDS-PAGE and UV analysis.

Preliminary Results of Current Research

Tetrahymena Telomerase

Tetrahymena thermophila TERT (ttTERT) has four domains, TEN-RBD-RT-CTE. Our attempts at making the full length protein (1112 amino acids) have failed due to lack of detectable expression. Using the secondary structure prediction for the C-terminus, truncations of ttTERT were made and cloned into the MBP-(C-GFP) vector (Figure 4). LB-Agar plate expression trials using C-GFP as an indicator (as described above) revealed that the largest expressable construct was amino acids 1-1015, which deletes the C-terminal half of the CTE. Follow-up expression tests in liquid culture confirmed that this construct solubly expresses and can be purified from *E. coli*. A C-terminal tag (GFP or GST) was necessary for expression, since no expression was detected from vectors lacking a C-tag. Preliminary studies with ttTERT(1-1015) indicate that it is able to bind in a ternary complex with TER and p65, however more studies must be done to further purify and characterize the complex.

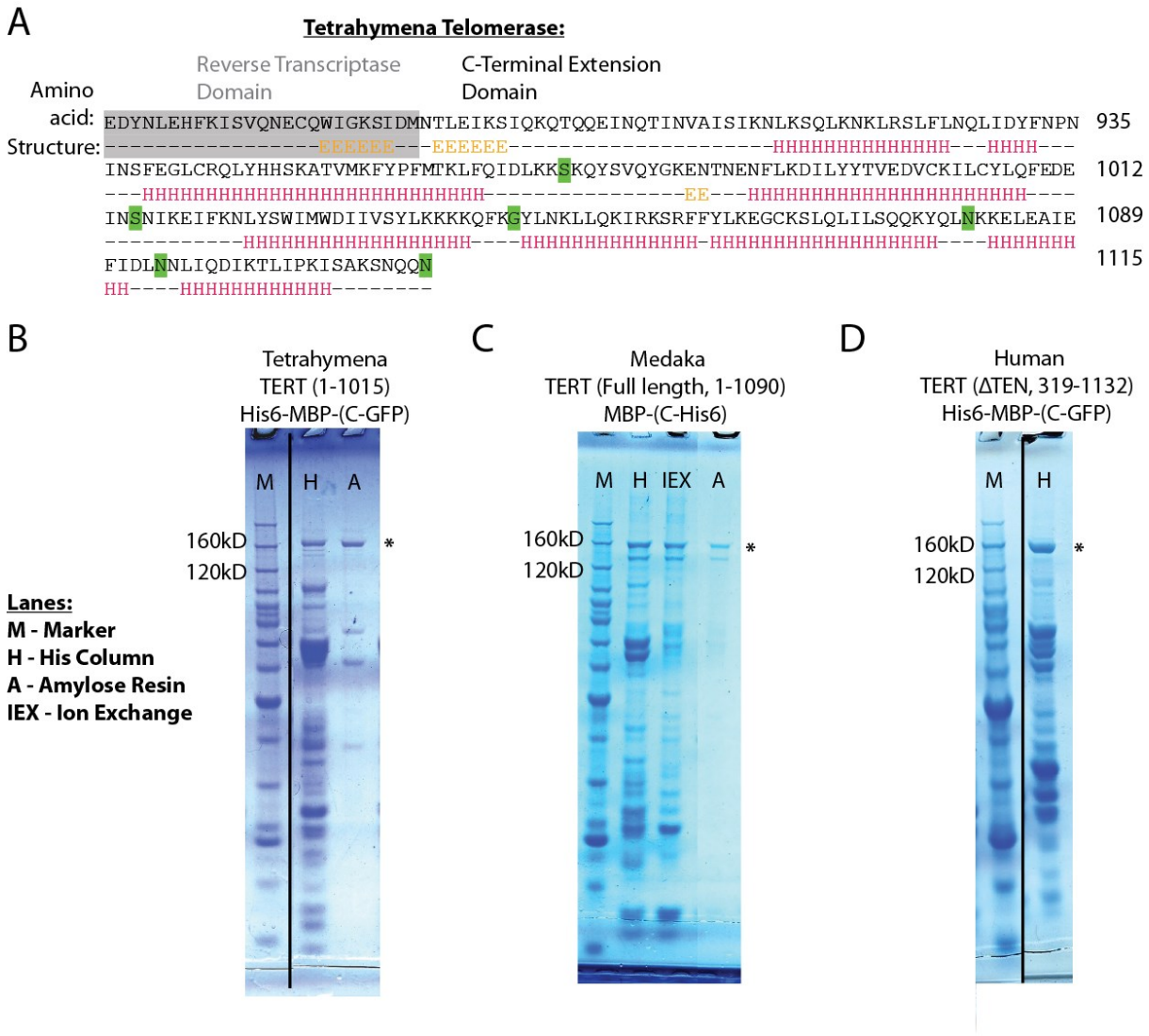


Figure 4: DC-LIC cloning system applied to telomerase proteins

(A) Amino acid sequence and corresponding secondary structure (E = beta sheet, H = alpha helix) for *Tetrahymena thermophila* (tt) TERT (C-terminus). Different TERT constructs were made according to the secondary structure, ending at the indicated residues (green highlight), and tested for expression using C-GFP. ttTERT (1-1015) was the largest expressing construct. SDS-PAGE gels show expression and purification of **(B)** *Tetrahymena* TERT (1-1015), His6-MBP-(C-GFP), **(C)** Medaka fish TERT (full length), MBP-(C-His6), and **(D)** Human TERT (319-1132), His6-MBP-(C-GFP). The lanes are labeled for the molecular weight marker (M), and purification step as His column (H), Amylose resin (A) or Ion exchange (IEX). The target protein location on the gel is indicated by (*).

Medaka Telomerase

Full length medaka fish TERT (mfTERT, 1-1090 amino acids) expresses in a variety of DC-LIC vectors. MBP-TERT-His6 is the most effective vector for this protein thus far as it allows for acceptable expression levels and a tandem (amylose, nickel) purification strategy which ensures full length product (Fig. 4). For this project, proper folding of mfTERT seems to be the most significant barrier. The purified mfTERT binds to *in vitro* transcribed TER, however telomerase activity has not been detected in these reconstitutions. Preliminary EM studies show potential telomerase particles, but the sample is not homogeneous. It is likely that chaperones are needed for proper folding and assembly, therefore future work will entail co-expression and co-purification with Dyskerin (a vertebrate telomerase co-factor) and possibly yeast expression systems which have the necessary homologs.

Human Telomerase

Human telomerase lacking the TEN domain, amino acids 319-1132, has been solubly expressed (Fig. 4). Additionally, the RBD of human TERT (319-601) has been co-expressed in *E. coli* with its RNA binding partner, the CR45 domain of human TER. NUS-RBD-(C-GFP) fusion protein was used to express the RBD and CR45-tRNA scaffold. The protein/RNA complex was purified intact using a nickel affinity column.

Conclusion

The DC-LIC cloning system, currently in use in the Feigon lab, is a useful tool for optimizing protein expression, while reducing cloning time and cost. It serves as a high throughput method for enhancing protein expression through efficient evaluation of multiple clones with different fusion tag partners. While protein expression is a significant hurdle, downstream obstacles include proper protein folding, solubility upon TEV cleavage of the fusion tag, and optimizing the purification process. The DC-LIC system begins to address the problems of protein folding and solubility by simplifying co-expression with protein and RNA

counterparts. The DC-LIC library can be expanded to include more cloning options, such as more fusion tags (and combinations), multiple insert cloning, and applications to the other expression systems (i.e. yeast, insect).

Materials and Methods

Building LIC vectors

N-terminal tags were cloned into the BamHI and AatII restriction sites of the PetDuet vector. The LIC site sequence was cloned at the same time as the N-terminal tag, and is included in the reverse primer with sequence 5′-AAATTT**GACGTC**GATAAGAGATTTAAATACAAATTTTCGXXX-3′, where **GACGTC** is the restriction site (AatII), the underlined is the LIC site, and XXX represents the N-terminal tag sequence. C-terminal tags were cloned into the AatII and XhoI restriction sites and are in frame with the AatII codons, GAC-GTC = Asp-Val.

LIC -Vector Preparation

1-2 µg of vector was digested with SwaI restriction enzyme from New England BioLabs (NEB) in 1x NEBuffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µg/ml BSA, pH 7.9) for 1hr at room temperature. The cut vector was purified by 1.2% agarose gel, and extracted from the gel using Invitrogen Gel Extraction Kit, eluted in 30 µL water. The vector was treated with T4 DNA polymerase by adding the following, 4µL of 10x NEB Cutsmart Buffer (500 mM potassium acetate, 200 mM tris-acetate, 100 mM magnesium acetate, 1 mg/ml BSA, pH 7.9), 4 µL of 25 mM CTP, 2 µL of 1 mM DTT, and 1 µL of NEB T4 DNA Polymerase. The treated vectors are stored at -20°C.

LIC-Insert preparation

For PCR, the forward primer has sequence 5′-GAAAATTTGTATTT(CCAGGGC)XXX-3′, where the underlined is the LIC complement, the parenthesis finishes encoding the optional TEV site, and XXX represents the target gene. The reverse primer has sequence 5′-GATAAGAGATTTCC(ZZZ)XXXX-3′, where the underlined is the LIC complement, (ZZZ) is either a stop codon or TEV sequence to cleave the C-terminal tag, and XXX represents the target gene complement. After PCR, the insert PCR product was agarose gel purified and extracted into 30 µL water. The insert was treated with T4 DNA polymerase by adding the following, 4 µL of 10x NEB Cutsmart Buffer, 4 µL of 25 mM GTP, 2 µL of 1 mM DTT, and 1 µL of NEB T4 DNA Polymerase. The treated insert is stored at -20°C.

LIC protocol

2 µL of insert and 1 µL of vector (both T4 DNA polymerase treated as described above) were added together and incubated for at least 5 minutes at room temperature. 1 µL of 25 mM EDTA was added to stop reaction. The entire reaction was added to an aliquot of competent cells for transformation. The DNA and cells were incubated on ice for 20-30 minutes, then heat shocked at 37°C for 1 minute, and incubated at 37°C for 1 hour. 100 µL of the cells are plated on LB-agar plates with appropriate antibiotic.

Protein expression and purification

In general, plasmids were transformed into BL21 (DE3) expression strain of *E. coli*. From the plate, a colony was picked for starter cultures, which were grown overnight in a volume that will be 1:100 dilution of the large culture. Large cultures (0.5-2 liters) were grown at 37°C to OD 0.6-0.7 and then changed to 18°C for 1 hour to allow cooling. IPTG was added to 0.1-1.0 mM, and then the cultures were grown overnight at 18°C. The cells were pelleted by centrifugation and resuspended in lysis buffer (20 mM Tris, 200 mM KCl, 20 mM imidazole, 1 mM TCEP, 10% glycerol, pH 7.5) and sonicated. The lysate was centrifuged, and supernatant

recovered and filtered. The filtered lysate was loaded onto an equilibrated 5mL HisTrap Column, washed with 10x column volumes of lysis buffer, and eluted with elution buffer (20 mM Tris, 200 mM KCl, 300 mM imidazole, 1 mM TCEP, 10% glycerol, pH 7.5). For TEV cleavage, the eluate was dialyzed against 1L of TEV cleavage buffer (20 mM Tris, 200 mM KCl, 2 mM β-mercaptoethanol, pH 7.5). The proteins were TEV cleaved for 1 hour at room temperature or 4°C overnight. This can be followed by size exclusion chromatography, ion exchange, or other affinity purification steps (i.e. amylose or GSTrap).

References

1. Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G., and Joachimiak, A. (2011) High-throughput protein purification and quality assessment for crystallization, *Methods 55*, 12-28.

2. Brondyk, W. H. (2009) Selecting an appropriate method for expressing a recombinant protein, *Methods Enzymol 463*, 131-147.

3. Hannig, G., and Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in Escherichia coli, *Trends Biotechnol 16*, 54-60.

4. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges, *Front Microbiol 5*, 172.

5. Betton, J. M. (2004) High throughput cloning and expression strategies for protein production, *Biochimie 86*, 601-605.

6. Festa, F., Steel, J., Bian, X., and Labaer, J. (2013) High-throughput cloning and expression library creation for functional proteomics, *Proteomics 13*, 1381-1399.

7. Cai, Y., Yan, W., Xu, W., Yin, Y., He, Y., Wang, H., and Zhang, X. (2013) Screening and identification of DnaJ interaction proteins in Streptococcus pneumoniae, *Curr Microbiol 67*, 732-741.

8. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic Acids Res 18*, 6069-6074.

9. Kwon, K., and Peterson, S. N. (2013) High-throughput cloning for biophysical applications, *Methods Mol Biol 1140*, 61-74.

10. Nallamsetty, S., Kapust, R. B., Tozser, J., Cherry, S., Tropea, J. E., Copeland, T. D., and Waugh, D. S. (2004) Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro, *Protein Expr Purif 38*, 108-115.

11. LaVallie, E. R., McCoy, J. M., Smith, D. B., and Riggs, P. (2001) Enzymatic and chemical cleavage of fusion proteins, *Curr Protoc Mol Biol Chapter 16*, Unit16 14B.

12. Young, C. L., Britton, Z. T., and Robinson, A. S. (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications, *Biotechnol J 7*, 620-634.

13. Kimple, M. E., and Sondek, J. (2004) Overview of affinity tags for protein purification, *Curr Protoc Protein Sci Chapter 9*, Unit 9 9.

14. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein, *Nat Biotechnol 24*, 79-88.

15. Nikolic, N., Barner, T., and Ackermann, M. (2013) Analysis of fluorescent reporters indicates heterogeneity in glucose uptake and utilization in clonal bacterial populations, *BMC Microbiol 13*, 258.

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

17. Tolia, N. H., and Joshua-Tor, L. (2006) Strategies for protein coexpression in Escherichia coli, *Nat Methods 3*, 55-64.

18. Ho, T. Q., Zhong, Z., Aung, S., and Pogliano, J. (2002) Compatible bacterial plasmids are targeted to independent cellular locations in Escherichia coli, *EMBO J 21*, 1864-1872.

19. Ponchon, L., Beauvais, G., Nonin-Lecomte, S., and Dardel, F. (2009) A generic protocol for the expression and purification of recombinant RNA in Escherichia coli using a tRNA scaffold, *Nat Protoc 4*, 947-959.

20. Ponchon, L., and Dardel, F. (2007) Recombinant RNA technology: the tRNA scaffold, *Nat Methods 4*, 571-576.

21. Deutscher, M. P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA, *Nucleic Acids Res 34*, 659-666.

22. Srisawat, C., and Engelke, D. R. (2001) Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins, *RNA 7*, 632-641.

APPENDIX

DC-LIC constructs

Construct*:	His Tag	N-Fusion	C-Fusion	Expression	Solubility	Folding	Notes
mfTERT(1-1090):							
mfTERT(1-1090):	N-term	Sumo		low	soluble	aggregate	
mfTERT(1-1090):	N-term	Sumo	GFP	low	soluble	aggregate	
mfTERT(1-1090):	N-term	MBP		low	soluble	aggregate	
mfTERT(1-1090):	C-term	MBP		low	soluble	folds with TER	binds to TERtrna in vivo
mfTERT(1-1090):	N-term	MBP	GFP	low	soluble	aggregate	
mfTERT(1-1090):	C-term	MBP	GFP	low	soluble	aggregate	
mfTERT(306-1090):							
mfTERT(306-1090):	N-term	Sumo	GFP	low	soluble	N/D	
mfTERT(306-1090):	N-term	MBP	GFP	low	soluble	N/D	
mfTERT(306-1090):	C-term	MBP	GFP	low	soluble	N/D	
mfTERT(306-860):							
mfTERT(306-860):	N-term	MBP	GFP	N/D	N/D	N/D	
mfTERT(306-860):	C-term	MBP	GFP	N/D	N/D	N/D	
mfTERT(306-579:RBD):							
mfTERT(306-579:RBD):	N-term	MBP		high	soluble	aggregate	Precipitate when cleave
mfTERT(306-579:RBD):	N-term	MBP	GFP	high	soluble	aggregate	Precipitate when cleave
hsTERT(1-1132):	N-term	MBP	GFP	moderate			
hsTERT(319-1132):							
hsTERT(319-1132):	N-term	MBP	GFP	low	soluble	aggregate	
hsTERT(319-1132):	C-term	MBP	GFP	low	soluble	aggregate	
hsTERT(319-1132):	N-term	Sumo	GFP	low	soluble	aggregate	
hsTERT(319-939):							
hsTERT(319-939):	N-term	MBP	GFP	N/D	N/D		
hsTERT(319-939):	C-term	MBP	GFP	N/D	N/D		
hsTERT(319-601:RBD):							
hsTERT(319-601:RBD):	N-term	MBP		moderate	soluble	N/D	
hsTERT(319-601:RBD):	N-term	GST		low	soluble	N/D	
hsTERT(319-601:RBD):	N-term	NUS		high	soluble	N/D	
hsTERT(319-601:RBD):	N-term	NUS	GFP	high	soluble	aggregate	
hsTERT(319-601:RBD):	C-term	NUS	GFP	high	soluble	aggregate	binds to CR45trna in vivo
hsTERT(319-601:RBD):	N-term	MBP	GFP	moderate	soluble	aggregate	
hsTERT(319-601:RBD):	C-term	MBP	GFP	moderate	soluble	aggregate	
hsDyskerin (21-418):							
hsDyskerin (21-418):	N-term	MBP	GFP	high	soluble		
hsDyskerin (21-418):	N-term	NUS	GFP	high	soluble		
hsDyskerin (21-418):	C-term	MBP	GFP	N/D	N/D		
ttTERT(1-1117):	N-term	MBP	GFP	no expression			
ttTERT(1-1094):	N-term	MBP	GFP	no expression			
ttTERT(1-1081):	N-term	MBP	GFP	no expression			
ttTERT(1-1045):	N-term	MBP	GFP	no expression			
ttTERT(1-938):	N-term	MBP	GFP	low	soluble	aggregate	
ttTERT(1-885):	N-term	MBP	GFP	moderate	soluble	aggregate	

***mf** = medaka fish, **hs** = homo sapien, **tt** = *Tetrahymena thermophila*

ttTERT(1-1015):							
ttTERT(1-1015):	N-term	MBP		no expression			
ttTERT(1-1015):	N-term	HIS		no expression			
ttTERT(1-1015):	N-term	MBP	GFP	low	soluble	aggregate	folds with P65/ter
ttTERT(1-1015):	C-term	MBP	GFP	very low	soluble	aggregate	
	N-term	MBP	GST	low	soluble	aggregate	
ttTERT(1-519):							
ttTERT(1-519):	N-term	MBP		high	soluble	monomer	
ttTERT(1-519):	N-term	MBP	GFP	high	soluble	monomer	
ttTERT(519-885:RT):	N-term	MBP		high	soluble/ most insolube	aggregate	
ttTERT-P65 chimeras:							
ttTERT(1-519)	N-term	HIS	P65C	high	soluble	monomer	binds to TERtrna in vivo
ttTERT(1-519)	N-term	HIS	P65-LRC	high	soluble	monomer	binds to TERtrna in vivo
ttTERT(1-885)	N-term	HIS	P65-LRC	N/D			
ttTERT(1-885)	N-term	MBP	P65-LRC	N/D			
ttTERT(1-1015)	N-term	MBP	P65-LRC	no expression			
ttTERT(883-1112)(CTE)	N-term	MBP	GFP	high	soluble		
ttTERT(979-1112)(CTE)	N-term	MBP	GFP	high	soluble		
ttTERT(979-1065)(CTE)	N-term	MBP	GFP	high	soluble		
ttP50N30:							
ttP50N30:	N-term	MBP		moderate	soluble	aggregate	
ttP50N30:	N-term	MBP	GFP	moderate	soluble	aggregate	
ttP75:	N-term	MBP	GFP	high	most insoluble	aggregate	
ttP65FL(1-543):	N-term	MBP	GFP	high	soluble	monomer	
ttP65(111-543dl5:LRC)	N-term	HIS		high	soluble	monomer	