Structure and sequence elements of the CR4/5 domain of medaka telomerase RNA important for telomerase function

Nak-Kyoon Kim1,2, Qi Zhang1,3 and Juli Feigon1,*

1Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095-1569, USA, 2Advanced Analysis Center, Korea Institute of Science and Technology, Seoul, 136-791, Republic of Korea and 3Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Received August 11, 2013; Revised November 13, 2013; Accepted November 16, 2013

ABSTRACT
Telomerase is a unique reverse transcriptase that maintains the 3’ ends of eukaryotic chromosomes by adding tandem telomeric repeats. The RNA subunit (TR) of vertebrate telomerase provides a template for reverse transcription, contained within the conserved template/pseudoknot domain, and a conserved regions 4 and 5 (CR4/5) domain, all essential for catalytic activity. We report the nuclear magnetic resonance (NMR) solution structure of the full-length CR4/5 domain from the teleost fish medaka (Oryzias latipes). Three helices emanate from a structured internal loop, forming a Y-shaped structure, where helix P6 stacks on P5 and helix P6.1 points away from P6. The relative orientations of the three helices are Mg2+ dependent and dynamic. Although the three-way junction is structured and has unexpected base pairs, telomerase activity assays with nucleotide substitutions and deletions in CR4/5 indicate that none of these are essential for activity. The results suggest that the junction is likely to change conformation in complex with telomerase reverse transcriptase and that it provides a flexible scaffold that allows P6 and P6.1 to correctly fold and interact with telomerase reverse transcriptase.

INTRODUCTION
Telomeres are the 3’ ends of linear chromosomes and comprise short tandem double-stranded DNA repeats, ending in a single-stranded 3’ overhang, and telomere-specific DNA binding proteins (1,2). By capping the 3’ end single-stranded overhang, the associated telomere end-binding proteins stabilize genomic DNA and prevent end-to-end fusion and enzymatic degradation. In eukaryotes, incomplete replication of the 3’ ends during each cell cycle and chromosome end processing result in progressive attrition of telomeric DNA (3). Eventually, critically shortened telomeres lead to chromosome instability and replicative senescence (4,5). In most eukaryotes, to compensate for the telomeric DNA loss, telomerase, a specialized reverse transcriptase (RT), maintains the telomere by synthesizing tandem DNA repeats. In most somatic cells, telomerase activity is low or undetectable. Telomerase is active in proliferative cells such as epithelial, germline and hematopoietic cells and in adult stem cells, where it partially compensates for telomere loss associated with increasing age (6), and it is highly active in ~90% of cancer cells, likely enabling their immortal phenotype (7). Telomerase is a ribonucleoprotein (RNP) enzyme complex with two catalytically essential elements responsible for telomeric DNA synthesis, the telomerase reverse transcriptase (TERT) and the telomerase RNA (TR) (8,9). The telomerase holoenzyme also contains species-specific proteins important for in vivo telomerase activity, RNP assembly and localization (1,8). TERT, unlike other RTs, requires an internal RNA template, contained within the TR, for synthesis of the telomere repeat. Telomere repeat addition processivity requires a translocation step. TERT possesses four evolutionarily conserved domains: telomerase essential N-terminal (TEN), telomerase RNA binding domain (TRBD), RT and C-terminal extension (CTE) (10–13). The TEN domain contains an anchor site that binds single-stranded DNA product and is important for repeat addition processivity (12,14). The TRBD is required for specific high affinity interactions with TR for telomerase catalytic function (14,15). The RT domain includes the enzyme active site where the TR template is located. The CTE domain may promote telomerase processivity and enhance nucleic acid association (16,17).
In contrast to the conserved structure of TERT, TR is highly variable in sequence and size ranging from ~150 nt in ciliates to >2000 nt in Candida and Aspergilla species (18–20). However, phylogenetic comparisons of TR secondary structures and functional studies of wild-type and mutant TRs have revealed two catalytically essential domains almost universally shared among known species (1,18). These are the template/pseudo-knot (t/PK) (also called core) domain and stem terminus element (STE) (also called activation domain) (1,2), which is CR4/5 (conserved regions 4 and 5) in vertebrates. The t/PK domain includes the RNA template and a template-enclosing pseudoknot, with a conserved pyrimidine motif triple helix (21–24). The other essential domain, the STE, stimulates telomerase activity and binds to the TRBD (1,25). The STE generally contains a stem (in yeasts) or stem-loop (in filamentous fungi, vertebrates and ciliates) required for catalysis. This stem is connected to a three-way junction, except in ciliates, with a proposed conserved secondary structure (20,26). These two domains of TR can reconstitute activity in vitro when added to TERT expressed in rabbit reticulocyte lysate (RRL), even when added as separate domains (25).

In vertebrates, CR4/5 forms a three-way junction structure with flanking P5, P6 and P6.1 stem-loops. The human TR (hTR) P6 is extended by an additional helix P6b and forms a linear P6a-P6b stem-loop structure (Supplementary Figure S1A) (27). The P6b stem-loop is dispensable for activity, but the P6a helix and part of the internal loop sequence that connects P6a and P6b are required for telomerase activity (25). In the highly conserved P6.1 loop are essential for telomerase activity but not for TERT binding (28,29). Based on secondary structure prediction, the junction region of CR4/5 in hTR consists of 16 single-stranded nucleotides, including several conserved residues near P6a and P6.1, which form a large internal loop (18). There are several sets of complementary sequences in the internal loop that may form base pairs, including those predicted by comparison among different species (20,26), but the effect of mutations that would disrupt and compensatory mutations that would restore these base pairs as well as other nucleotide substitutions on TERT binding and activity provided little evidence for formation of base pairs within the internal loop in the holoenzyme context (29). The structural role of the three-way junction on telomerase assembly or catalytic activity remains elusive.

Medaka (Oryzias latipes), a teleost fish, TR (mdTR) is the smallest identified TR in vertebrates (Figure 1) (30). Although mdTR has only 312 nt, it includes the t/PK, CR4/5 and H/ACA scaRNA domains found in all vertebrate TRs. In general, the sizes of these domains are smaller than those in hTR. The medaka CR4/5 (mdCR4/5) differs from human CR4/5 (hCR4/5) primarily in having a shorter P6, lacking the P6b of hTR and possibly a smaller internal loop at the three-way junction. The highly conserved mdCR4/5 P6.1 is identical to the hCR4/5 P6.1 domain except for two non-conserved nucleotides in the apical loop.

Here, we have investigated the structural and sequence requirements of medaka TR for telomerase catalytic activity. We report the first structure of a complete CR4/5 domain, by solution-state nuclear magnetic resonance (NMR) spectroscopy. The mdCR4/5 forms a Y-shaped structure in which the positions of all three
NMR spectroscopy and structure calculations

NMR spectra were recorded on Bruker DRX 500, 600 and Avance 800 MHz spectrometers equipped with HCN cryoprobe or QXI probe at 10°C for exchangeable and at 20°C for non-exchangeable proton spectra. NMR spectra were processed and analyzed using XWINNMR 3.5 (Bruker), TOPSPIN (Bruker) and Sparky 3.110. (University of California, San Francisco, CA, USA).

The imino and amino exchangeable protons were unambiguously assigned from 2D NOESY, 2D 13N-correlated CPMG NOESY, 2D 1H-15N HMQC and 2D H5(C5C4N)H spectra (34). RNA base pairs were confirmed with 2D HNN-COSY experiments (35,36). The 2'-hydroxyl protons were identified as described (32). The standard H1'-base proton sequential assignments were initially obtained from analysis of 2D NOESY and 2D TOCSY spectra with unlabeled RNA samples. The assignments for all non-exchangeable protons were achieved from analysis of 2D NOESY, 2D 1H-13C HSQC, 2D TOCSY, 2D HCCH-COSY and 3D HCCH-TOCSY (37–39). A suite of 2D-filtered/editon NOESY (F2f, F1fF2e, F1fF2f and F1eF2f) experiments on base-specific 13C,15N-labeled RNAs was used to resolve ambiguous assignments in overlapped regions and to obtain nuclear Overhauser effect (NOE) restraints, as previously described (40). The sugar pucker was determined as C2'-endo (δ = 145 ± 30°) for residues 187–189 and 205–208 that showed strong H1'-H2' cross-peaks in 2D DQF-COSY (41). 3JH2P and H3CP were measured using 13P spin echo difference CT-HSQCs to determine the ε dihedral angles for the loop residues (42). The base conformation of G207 was determined to be syn (χ = 60 ± 30°), based on the presence strong intranucleotide H1'-H8 and weak H2'-H8 NOE cross-peaks in a 2D NOESY. All other bases were restrained to anti (χ = −160 ± 30°). For structural refinement and dynamic characterization, one-bond C-H (C1'H1', C2'H2, C5'H5, C6H6 and C8H8) and N-H (N1H1 and N3H3) residual dipolar couplings (RDCs) were measured on uniformly 13C,15N-labeled samples in the absence and presence of ~15 mg/ml Pr1 phage (ASLA Biotech, Ltd.) at 20°C on the 800 MHz spectrometer. C-H RDCs were measured from the splittings of 1H-13C doublets along the 1H dimension using 2D 1H-13C S'CT-HSQC experiments (43), and N-H RDCs were measured and averaged from the splittings of 1H-15N doublets along both the 1H and 15N dimensions using standard 2D 1H-15N HSQC experiments. A total of 77 one-bond C-H and N-H RDCs were obtained for mdCR4/5 in the absence of Mg2+. For structural refinement, all C-H and N-H RDCs were input as normalized C-H RDCs with a fixed bond length of 1.0 Å as previously described (44).

Inter-proton distances were measured from normalized cross-peak volumes in 2D NOESY as well as 2D-filtered/edited NOESY experiments acquired with various mixing times (40). NOE distance restraints were classified as very strong (2.5 ± 1.0–0.7 Å), strong (3.5 ± 1 Å), medium (4.5 ± 1 Å), weak (5.5 ± 1 Å) and very weak (6.5 ± 1 Å). The dihedral angle restraints (α, β, γ, δ, ε, χ, ψ2 and ζ) were used in the structure calculations, where the δ, ε, and χ were determined experimentally as described earlier in the text. The α, β, γ, ψ2 and ζ dihedral angles for the nucleotides involved in base pairs in the helix were constrained to the A-form values (–62.1 ± 30°, 179.9 ± 30°, 47.4 ± 30°, 37.3 ± 30° and –73.6 ± 30°, respectively).
Final structure calculations included hydrogen bond distance restraints and weak base pair planarity restraints for the 16 Watson–Crick (WC) base pairs and 3 G·U base pair as previously described (32). Two hundred initial structures of mdCR4/5 were calculated using XPLOR-NIH 2.9.8 starting from an extended unfolded RNA using the NOE distance and dihedral angle restraints following XPLOR protocols. The 50 lowest energy structures were further refined by inclusion of RDCs for the final structures.

To account for different degrees of alignment due to inter-helical motions, the values for the magnitude (Da) and asymmetry (R) of the alignment tensor for each stem (P5, P6 and P6.1) of mdCR4/5 were first evaluated by analyzing their RDCs using the program RAMAH (45). During this initial order tensor analysis (46), idealized ‘A’-form helices corresponding to P5, P6 and P6.1 were generated using Insight II software (Accelrys) and used as input coordinates (47). These initial values were then used as the starting values and were further optimized using the structure calculated without RDCs. The optimized values for P5, P6 and P6.1 are Da = −17.1 Hz, 29.2 Hz and 10.4 Hz, respectively, and R = 0.63, 0.62 and 0.38, respectively. During the final structure calculations with RDCs, the force constants for the RDCs were gradually increased from 0.001 to 0.85, 0.50 and 1.40 kcal mol$^{-1}$ Hz$^{-2}$ for P5, P6, and P6.1, respectively, to account for the differences in the Da values. Experimental restraints and structural statistics for the 20 lowest energy structures are shown in Table 1. All structures were viewed and analyzed using MOLMOL (48) and PyMOL (DeLano Scientific LLC).

The inter-helical dynamics were characterized by the resultant generalized degree of order (GDO),

$$\vartheta = \sqrt{2(S_{xx}^2 + S_{yy}^2 + S_{zz}^2)/3}$$

where $S_{xx}$, $S_{yy}$, and $S_{zz}$ are the order tensor values in the principle axis system obtained from order tensor analysis of each stem using RAMAH program with the RDC refined structure. The relative motions between each helix were then measured by the relative ratios of $\vartheta$ values determined for P5, P6, and P6.1, where $\vartheta_{int,P5} = \vartheta_{P5}/\vartheta_{P6}$ and $\vartheta_{int,P6.1} = \vartheta_{P6.1}/\vartheta_{P6}$, and $\vartheta_{int}$ is the inter-helical generalized degree of order (also called GDO$\text{int}$). Assuming a cone motional model, the amplitude of inter-helical motion can be described by the cone radius angle, $\Psi_{cone}$, which can be calculated from $\vartheta_{int}$ values based on the equation $\vartheta_{int} = 1/2 \cos \Psi_{cone}$ (1 + $\cos \Psi_{cone}$) (49).

For structure determination of mdCR4/5 in the presence of Mg$^{2+}$, the line broadening observed in the presence of Mg$^{2+}$ precluded obtaining accurate NOE-based distance constraints. Therefore, an NMR-based model structure was determined by generating a starting structure using the same NOE and dihedral angle restraints that were used for mdCR4/5 without Mg$^{2+}$, and refining this structure with 60 RDCs acquired on the mdCR4/5 sample with Mg$^{2+}$, prepared as described earlier in the text. Owing to differences in chemical shifts, NMR line broadening and spectral overlap in the two samples, there are some differences in the sets of RDCs collected (see Supplementary Table S1). The RDCs report on structural differences between the samples with and without Mg$^{2+}$.

### Modeling of mdCR4/5–mdTERT (medaka TERT) complex

The medaka TRBD (mdTRBD) domain was generated using the SWISS MODEL workplace server with the TRBD sequence of mdTERT (50) and the crystal structure of TRBD from Takifugu rubripes (PDB ID: 4LMO) (51) as a structural template. The modeled mdTRBD domain was then superimposed on the TRBD domain in the structure of the Tribolium castaneum TERT (PDB ID: 3KYL) (52), and the best fit was used to replace the Tribolium castaneum TRBD with the modeled mdTRBD. The model structure of mdCR4/5 in the presence of Mg$^{2+}$ was used and positioned on the mdTRBD in a way that satisfied the constraints (within 10 Å) from the cross-linking study (53). The resulting position of the P6.1 loop at the interface between the CTE and TRBD results in some steric clash, suggesting that some conformational rearrangement would need to occur.

### Direct telomerase activity assays

For telomerase activity studies, the DNA (a gift from Prof Julian J.L. Chen) for a full-length (312 nt) wild-type medaka TR was cloned into pUC19 (pUC19mdTR). The mdCR4/5 mutants were made by site-directed mutagenesis (Stratagene). Both wild-type and mutant RNAs were transcribed in vitro using linearized DNA templates obtained from polymerase chain reaction (PCR) amplification of the pUC19mdTR. RNA products were purified by 5% polyacrylamide gel electrophoresis and electroelution as described earlier in the text and concentrated.

N-terminally FLAG-tagged mdTERT in pCITE4a was expressed using the TNT quick-coupled transcription/translation RRL system (Promega) according to the manufacturer’s instruction. Briefly, mdTERT protein
was translated in 10 μl of RRL for 1 h at 30°C. *In vitro* synthesized mdTR was added to a final concentration of 1 μM, and telomerase RNP was reconstituted by incubating the mixture for 30 min at 30°C. *In vitro* telomerase activity was measured by using the direct telomere extension assay. For telomere extension reactions, 3 μl of the reconstituted telomerase RNP was added to a final concentration of 1 μM, and telomerase RNP was reconstituted by incubating the mixture for 30 min at 30°C.

RESULTS

Secondary structure of mdCR4/5

The mdCR4/5 construct for NMR study includes the full sequence (nts 170–220) (Figure 1, boxed region). The secondary structure of mdCR4/5 was determined from the NMR data by identifying the base pairs in P5, P6, P6.1 and the junction. The WC and G·U base pairs were identified from the NOEs between imino and imino (Figure 2A), imino and aromatic and imino and amino protons in 2D 1H-1H NOESY and 2D 15N-correlated NOESY spectra, and confirmed by 2D HNN-COSY spectra (Figure 2B). Observed base pairs in stems P5, P6 and P6.1 agree with those predicted by phylogenetic analysis (30). The P5 stem shows sequential imino to imino cross-peaks from the G170-C220 base pair to the C174-G216 base pair (Figure 2B, cyan). P6 forms a helix containing one G·U base pair and a single bulge nucleotide U182 and is capped by a 3-nt terminal loop. The G180·U195 base pair is identified by the typical NOE pattern between the imino resonances in the base pair and to the neighboring base pairs in 2D NOESY spectra (Figure 2A, green). The sequential connectivity between the imino cross-peaks of the P6 stem in the 2D NOESY spectrum are disrupted by the bulge U182, and the remaining imino proton connectivities for base pairs above the bulge were assigned using the NOEs observed in other chemical shift regions. In P6.1, four sequential cross-peaks between the imino protons from G201 to G207 were identified (Figure 1A, orange). Interestingly, base pairs A200-U212 and U204-U208 that were predicted by phylogenetic analysis are not formed in P6.1 of mdCR4/5. Instead, we observed an unexpected NOE cross-peak between G203H1 and G207H1, suggesting that G207 base pairs to U204 (see later in the text). The junction flanked by P5, P6 and P6.1 has a unique base pairing pattern, which is also different from the prediction from phylogenetic analysis. Except for U214, all the imino protons in the junction are involved in hydrogen bonding for base pairing. In the imino region of the 2D NOESY spectrum, a G·U → C-G → G·U sequential cross-peak pattern was observed (Figure 2A, magenta). These base pairs were assigned as G175·U213, C176-G213 and G198·U212 are in the junction. The 2D HNN COSY spectra in (B) show a direct correlation between imino protons and their hydrogen-bonded nitrogen atoms in the canonical base pairs (upper box: G-C, lower box: A-U). Cross-peak labels are colored as in (A).

Solution structure of mdCR4/5

The structure of mdCR4/5 in 10 mM sodium phosphate (pH 6.3), 100 mM KCl was calculated using 1085
NOE-based distance restraints with 623 inter-residue NOEs. In addition to the NOE restraints, the relative helical orientations were determined by incorporating 77 RDCs in the structure calculations (Table 1 and Supplementary Table S1). The 20 lowest energy structures (Figure 3A) have a root-mean-square deviation of 0.93 Å for all heavy atoms.

The solution structure of mdCR4/5 reveals that the P5, P6 and P6.1 helices are splayed in a Y-shape around a well-structured junction. P6, the connecting junction nucleotides (J5,6), and P5 stack sequentially on each other (Figure 3). Helical angle analysis showed that the helical axis of P5 is inclined ~60° and ~85° relative to P6 and P6.1, respectively, and the P6.1 helix is inclined from the P6 helical axis by ~35° [analyzed by Curves 5.3 (54)]. P5 forms a six WC base-paired A-RNA helix with an average major groove width of ~11 Å. P6 has eight base pairs including one G•U wobble pair and is capped by a 3-nucleotide loop, and P6.1 has three WC base pairs capped by a UUCGU pentaloop. The nucleotides in the three-way junction (G175–C177, G198–A200 and U212–U215) form a unique tertiary structure with unpredicted base pairings (Figure 4). Nucleotides G175 and C176 at the 5/6 junction form canonical wobble and WC base pairs with U215 and G213 in the 5/6.1 junction, respectively. There is a slight twist (~20°) and roll toward the major groove between the C174-G216 base pair of P5 and the bottom junction G175•U215 base pair. C177 and U214 are not involved in base pairing, but stabilize the junction by providing additional base-stacking interactions—the C177 base is stacked on the C176 base, and the U214 base stacks between the bases of G213 and U215 at the top of P5. The phosphate backbone unwinds after C177, and the P6 stem starts at the G178-C197 base pair. Nucleotide U212, which has been predicted to form a terminal A-U base pair in the P6.1 stem of vertebrate TR (18), instead forms a wobble base pair with G198. The bases of G198•U212 are almost perpendicular to those of C176-G213 and G178-C197, resulting in a sharp turn in the phosphate backbone between C197 and G198 (Figure 4A). The conserved A199 and A200 are stacked in between G198•U212 and G201-C211 at the bottom of P6.1.

The mdCR4/5 P6 has a single bulge in the stem and a 3-nucleotide loop (Figure 3B, green, Supplementary Figure S1A). In the P6 helix, U182 is bulged out of the helix and lies perpendicular to the major groove. This universally conserved pyrimidine bulge induces an S-shape kink in the helix backbone. The P6 loop is well structured, with U187 pointing into the minor groove and C188 and G189 into the major groove of the P6 helix.

In the P6.1 pentaloop, the first four nucleotides form a structure similar to the well-characterized UUCG tetraloop, and the last nucleotide, U208 is flipped out (Figure 4C). Unlike the secondary structure prediction, U204 does not form a base pair with U208; no imino protons of U204 and U208, and no inter-residue NOEs between U204 and U208 are observed. Instead the conserved G207 forms a sheared pair with U204 as seen in the UUCG tetraloop (55). Because the U204 2'-hydroxyl proton resonance and the G207 imino proton resonance both had broad linewidths, the expected NOE between the 2'-hydroxyl proton of U204 and G207H1 that is usually observed in the U·G base pair in a UUCG tetraloop were not seen. However, there are inter-residue NOEs between G203 imino and G207 amino and imino protons, and between U204 sugar protons and G207 imino protons, indicative of the sheared U204•G207 base pair seen in the structure. Also as seen in standard UUCG tetraloops, the bases of U204, C206 and G207 stack on U203, U204 and C209, respectively, which increases the stability of P6.1, and the universally conserved U205 is flexibly positioned in the minor groove.

In the mdCR4/5 structure, the P6.1 loop is positioned near the minor groove of P6 near base pair C183-G193, and the flipped out loop nucleotide U208 points toward this groove. This U208 base is dynamic, as evidenced by the intensities of the C6H6 and C5H5 cross-peaks in the 1H-13C HSQC spectra (Figure 5A and Supplementary Figure S2A). There is a sharp turn in the backbone between C197 at the bottom of P6 and G198 (where it forms the G198•U212 base pair), which results in close approach of the phosphodiester backbone from P6 C194 through C197 to P6.1. Only a few long distance (>4.5 Å)
NOEs are observed between P6.1 and P6, specifically between C209 base and sugar at the top of P6.1 and the G193 sugar (2'H and 2'OH) in the middle of P6, indicative of close approach of these two helices at least some of the time. However, no NOEs indicative of close-packing of the P6.1 loop or stem against P6 are observed.

Inter-helical dynamics

To investigate how fixed the position of the three stems are relative to each other, we carried out order tensor analysis of the RDC data measured for each stem, which provides information on inter-helical dynamics faster than millisecond time scale. The obtained stem-specific order parameters indicate that both stems P5 and P6.1 have a large amplitude of inter-helical motions relative to P6, as predicted by $\theta_{\text{int,P5}} = \theta_{P5}/\theta_{P6} = 0.55 \pm 0.03$ and $\theta_{\text{int,P6.1}} = \theta_{P6.1}/\theta_{P6} = 0.30 \pm 0.02$, where $\theta_{\text{int}}$ is the inter-helical GDO as described in ‘Materials and Methods’ section (46). Using the cone-motional model to estimate the inter-helical dynamics, the determined amplitudes of the motion for P5 and P6.1 are $\sim 49{\text{o}}$ and $\sim 60{\text{o}}$, respectively, relative to P6. These results are consistent with the paucity of inter-helical NOEs observed between P6 and P6.1. We conclude that the determined solution structure is the averaged structure of a dynamic ensemble.

Divalent metal ion binding and Mg$^{2+}$ induced global change of mdCR4/5 structure

Given the dynamic nature of mdCR4/5 in the absence of divalent cations, we investigated whether divalent cations had any effect on its folding. To determine the effect of Mg$^{2+}$ on the mdCR4/5 structure, MgCl$_2$ was added in increments up to 30 mM to a $\sim$0.5 mM mdCR4/5 sample and chemical shift changes were monitored by $^1$H-$^1$C HSQC experiments (Figure 5 and Supplementary Figure S2A). Most of the residues show small chemical shift changes and/or line broadening upon addition of Mg$^{2+}$ ($\Delta\delta_{\text{max}} <$0.12, Supplementary Figure S2C), indicating fast or intermediate exchange between Mg$^{2+}$-free and Mg$^{2+}$-bound conformations. The NOESY cross-peak patterns of the imino proton resonances that identify the base pairs are the same (Supplementary Figure S3) and the NOESY cross-peak patterns for the non-exchangeable spectra are also similar (data not shown), indicating that addition of Mg$^{2+}$ does not cause significant local structural changes or changes in base pairing.

The largest chemical shift perturbations ($\sim$0.1 ppm) are observed for resonances from G180 and G189 in the P6 stem and loop, respectively, C177, G213 and U214 in the junction and G201 in P6.1. Except for G189, all of these residues are either directly involved in or near a G·U wobble pair, which is known to be a divalent metal-binding site (56). Mg$^{2+}$ binding could stabilize the flexible loop nucleotide G189. Negligible changes in chemical shifts and signal intensities were observed for $^1$H-$^1$C resonances from the non–base-paired loop or bulge residues: U187 and C188 in the P6 loop, U205, C206 and U208 in the P6.1 loop and U182. Based on their large relative intensity in $^1$H-$^1$C HSQC spectra, these residues are more dynamic than base-paired residues. In contrast, relatively large chemical shift changes are observed for non base-paired residues in the junction (C177, U214, A199 and A200).
To further characterize the divalent metal ion binding sites in mdCR4/5, we used paramagnetic line broadening experiments with Mn$^{2+}$. Mn$^{2+}$ is a sensitive and direct probe for identifying metal binding residues in RNA using NMR (33). The paramagnetic Mn$^{2+}$ induces a distance dependent ($r^{-6}$) line broadening in resonances of residues proximal to bound Mn$^{2+}$ that results in reduced NMR signals. Addition of 40 mM Mn$^{2+}$ results in line broadening of resonances, observed in $^1$H-$^{13}$C HSQC spectra, of residues near the G·U base pairs in the P6 stem (G180 and G181) and in the junction (C197, G198, U212 and G213), consistent with the results with Mg$^{2+}$ (Supplementary Figure S2B). Line broadening was observed around tandem purine bases throughout mdCR4/5: G169 through A172 in P5, A184, G189 and G190 in P6, and A199 through G203 in P6.1. The latter region is also where there is a sharp turn in the backbone between P6 and J6.1, and divalent metal ion binding could help reduce the electrostatic repulsion between the phosphate backbones at this turn.

We next investigated the effect of Mg$^{2+}$ on the overall folding of mdCR4/5. Titration of Mg$^{2+}$ up to 30 mM did not saturate the chemical shift perturbations, suggesting weak binding affinity of Mg$^{2+}$ to RNA. Significant NMR line broadening was also observed at higher Mg$^{2+}$ concentrations, which may be largely attributed to non-specific Mg$^{2+}$ binding to RNA. Consequently, further analysis at higher saturating Mg$^{2+}$ concentration was not possible due to difficulties in resonance assignments and NOE-based distance measurements in NMR spectra. Therefore, we calculated a model structure of mdCR4/5 in the presence of Mg$^{2+}$ starting with the NMR restraints previously used for mdCR4/5 without Mg$^{2+}$ and a set of RDC data obtained for the sample in the presence of Mg$^{2+}$. As described earlier in the text, as the RNA retains the same base pairs and NOESY cross-peak patterns in the presence of Mg$^{2+}$, we conclude that the local RNA structures such as the helices, P6 and P6.1 loops, the U182 bulge and the junction are not significantly changed. Based on this assumption, the structure was calculated using the initial structure calculated without RDCs for the sample without Mg$^{2+}$ as a starting structure and then refining with a set of 60 RDCs measured on the sample with Mg$^{2+}$, prepared as described in the ‘Materials and Methods’ section (see Supplementary Table S2 for structural statistics).
The calculated model structure of mdCR4/5 with Mg$^{2+}$ showed a change in the relative orientations of P5, P6 and P6.1 (Figure 5E). In the presence of Mg$^{2+}$, P5 and P6 are nearly coaxial, with only $\sim 2.5^\circ$ helical angle deviation between each other. The angle between the helical axes of P6 and P6.1 increases from $\sim 35^\circ$ to $\sim 65^\circ$ (Figure 5B and C). The internal GDO values for P5 and P6.1 are $\theta_{\text{int},P5} = \theta_{\text{P5}}/\theta_{\text{P6}} = 0.57 \pm 0.02$ and $\theta_{\text{int},P6.1} = \theta_{\text{P6.1}}/\theta_{\text{P6}} = 0.49 \pm 0.02$, respectively. In the cone-helical motion model as described earlier, the amplitude of the inter-helical dynamics are $\sim 47^\circ$ for P5 and $\sim 50^\circ$ for P6.1 relative to P6, respectively. This indicates that both P5 and P6.1 remain mobile relative to P6 even in the presence of Mg$^{2+}$ ions, albeit with a lower amplitude of motions.

**Effect of mdCR4/5 junction nucleotide substitutions on telomerase activity**

To identify if any structural elements of the mdCR4/5 are correlated with telomerase function, we incorporated nucleotide substitutions into the full-length mdTR and performed direct telomerase activity assays in vitro using mdTERT translated in RRL (Figure 6). We first investigated the effect of base pairing in the junction using a compensatory mutagenesis strategy (Figure 6B, lane 10–12). In the NMR solution structure of mdCR4/5, three base pairs, G175-U215, C176-G213 and G198-U212, are formed in or next to the junction (Figure 6A, boxes). Simultaneous disruption of all three base pairs by the mutations G175U/C176U/G198U or U212A/G213A/U215A (Figure 6B, lane 10–11) surprisingly showed no effect or slightly increased telomerase activity. Compensatory mutations that potentially restored base pairing also had little effect on activity in vitro (Figure 6B, lane 12). Hence, formation of the base pairs in the junction of mdCR4/5 does not appear to be required for catalytic activity.

The conserved P6.1 terminal A-U base pair in vertebrate TR (A200 and U212 in mdCR4/5) (18,20) predicted by phylogenetic comparison does not form in the solution structure of mdCR4/5. Instead, A200 is unpaired and U212 forms a U*G base pair with G198. Thus, we next examined whether an A200-U212 base pair might be important for mdCR4/5 folding and telomerase catalytic activity even though they are not paired in the solution structure. The compensatory substitutions A200U/U212A have no effect on telomerase activity (Figure 6, lane 7). The A200C substitution that would disrupt the previously predicted A200-U212 base pair decreased activity somewhat to 79% (Figure 6B, lane 5). The U212G that would disrupt A200-U212 would also disrupt the G198-U212 base pair seen in the solution structure and decreases telomerase activity to 60% (Figure 6B, lane 9). However, the U212C substitution that would disrupt A200-U212 and at the same time replace the G198-U212 base pair with a G-C base pair decreases activity to 20% (Figure 6B, lane 8). Finally, G198U, which would change the G198-U212 base pair to U-U but leave the A200-U212 base pair intact decreases activity only slightly to 87% (Figure 6B, lane3), consistent with the lack of effect of this mutation in the context of disruption of all three junction nucleotides discussed earlier in the text. Thus, it seems likely that the G198-U212 base pair seen in the solution structure is not present in or important for the complex formation with mdTERT TRBD, whereas the A200-U212 base pair predicted by phylogeny may be. This would be the terminal base pair of the P6.1 helix and if present would most likely play a role in stabilizing the helix.

A199 is a universally conserved nucleotide linking P6 and P6.1. Nevertheless, the substitution A199C had no significant effect on telomerase activity (Figure 6B, lane 5). In previous studies on hTR, the same result was shown for the mutation on the corresponding (A301) residue, where A301U (see Supplementary Figure S1C, FL CR4/5 for the sequence identity) substitution had only a minor impact on hTR-hTERT (human TERT) interaction and catalytic activity (29). In contrast to the negligible and small effects of individual A199C and A200C substitutions, respectively, simultaneous substitution of A199C/A200C significantly reduced catalytic activity (Figure 6B lane 6). However, this decrease can be explained by the misfolding of mdCR4/5 on A199C/A200C double mutation. Although
secondary structure prediction by Mfold (57) (Supplementary Figure S4) does not calculate the internal loop base pairing, it does predict the formation of the catalytically essential P6.1. Mfold generates only one conformation for A199C with the P6.1 stem-loop intact (Supplementary Figure S4B), but in the case of the A200C substitution two mdCR4/5 conformers were generated, where the less thermodynamically stable structure does not form the catalytically essential P6.1 hairpin (Supplementary Figure S4C, right). The presence of an inactive conformer in A200C could explain the decrease in telomerase activity compared with the A199C mutant. For the A199C/A200C double mutant, mdCR4/5 is predicted to fold into a stem-loop-stem hairpin structure, where the substituted C199 and C200 are base-paired with G190 and G189, respectively. In this case, the P6.1 stem-loop is completely lost, which could explain the almost complete loss of telomerase activity (Figure 6B, lane 6 and Supplementary Figure S4D). Consistent with the predicted change in overall fold, the A199C/A200C substituted TR migrates faster than the other wild-type and mutant mdCR4/5 RNAs in a nondenaturing gel (Figure 6C).

In summary, none of the mutations in junction residues including disruption of the three base pairs seen in the solution structure and changes in the conserved nucleotides decrease telomerase activity <75%, with the exception of U212G (60%) and U212C (20%). These mutations change both the observed G198-U212 and the predicted U212-A200 base pairs, so it is difficult to rationalize their effect. Previous studies of mdCR4/5 show that the minimal domain for >75% TRBD binding comprises nucleotides 176–214 (53). This minimal domain retains P6, P6.1 and the junction nucleotides but deletes P5. Similarly, for hCR4/5, residues 253–318, which comprise P6ab and P6.1 with 3 and 4 single-stranded nucleotides at the 5' and 3' ends, respectively, are sufficient to confer 50–70% activity relative to full-length hCR4/5 when combined in trans with the core domain and reconstituted with hTERT (32). Therefore, the role of the internal loop residues and P5 residues is likely to help position P6, P6.1 and flanking residues for interaction with TERT in the context of the full-length TR. The sequence of the internal loop residues including the conserved nucleotides may also be important for proper RNA folding to maintain the P6.1 and P6 stem-loops.

Effect of mdCR4/5 loop and bulge substitutions on telomerase activity

We next investigated the importance of the conserved P6 bulge (U182), found in all vertebrate TRs, and the loop G189 residues for telomerase activity. In the previous UV cross-linking studies of mdCR4/5 with the TRBD of mdTERT, the bulge U182 and P6 loop residue U187 were identified as being intimately involved in TRBD binding (53). Deletion of U182 or substitution to U182C eliminated or reduced by ~3-fold, respectively, the binding of mdCR4/5 to mdTERT TRBD (53). However, our activity assays using full-length TR and TERT showed that U182C substitution slightly increased telomerase activity (Figure 6B; lane 13). The single bulge nucleotide in the P6 stem is not sequence conserved, but is present in all vertebrate TRs except for one murine species (50). Because the bulge U182 base is looped out into the major groove of the P6 helix and has no tertiary interactions with other parts of mdCR4/5, the U182C substitution is unlikely to change the P6 stem structure. Thus, although previous cross-linking and binding assays of CR4/5 with the isolated TRBD from TERT indicate that U182 may form specific contacts with TRBD, changing it to C does not appear to reduce the interaction with full-length TERT sufficiently to affect activity.

G189, which is 100% conserved in teleost fish and conserved as U or G for all identified vertebrate TR (42 species), is near the P6 loop residue U187 that cross-links to mdTRBD. Our assays show that G189A substitution has no effect on activity (Figure 6B lane 14). Substitution of all three loop-residues (U187/C188/G189) with a GAAA tetraloop has been shown to nearly abolish binding of mdCR4/5 to the isolated TRBD (53). As with U182, this suggests that G189 may be involved in TRBD binding, but is not directly involved in telomerase activity. The differences in binding and telomerase activity can be explained by the fact that telomerase activity was assayed in the context of full-length TR and TERT, whereas the binding assays were performed with mdCR4/5 and TRBD only. Modeling based on the cross-linking studies predicts that the P6.1 loop interacts with the TERT CTE domain that is absent in the TRBD binding assay, which would provide additional binding affinity (53).

Finally, we tested the importance of the P6.1 loop residues for telomerase activity. P6.1 is an essential and conserved stem-loop in the CR4/5 domain of vertebrate TR (28,29,32,53,58). Base pairing of P6.1 is critical for TRBD binding, which is required for telomerase activity, whereas the loop is important for catalysis but not for TERT binding (28,29). A previous study showed that simultaneous mutation of the two absolutely conserved P6.1 loop nucleotides, U205A/G207C in mdCR4/5, did not affect TERT binding affinity, but decreased activity to 15% in vitro (53). The same mutation in murine TR eliminated telomerase activity (28). In the mdCR4/5 P6.1 loop structure, 204-UUCG-207 forms a tetraloop, and U208 is flipped out toward the minor groove of P6 (Figure 4C). Our results show that deletion of U208 has no effect on activity (Figure 6B, lane 16). Substitution of U205A/G207C would be expected to result in structural changes to the loop, whereas deletion of the bulge U208 should have little effect. These results show that the non-conserved U208 is not required for telomerase activity.

Interestingly, the double mutation of the non-conserved nucleotides C206G/U208G that changes the sequence of the P6.1 loop to that of hTR P6.1 (UUGGG) decreased telomerase activity to ~35% of wild-type mdTR. Because hTR P6.1 loop folds into a structure different from mdTR (Supplementary Figure S1B) (58), we propose that telomerase activation by P6.1 is loop sequence specific.

DISCUSSION

In this work, we present the first structure of a complete CR4/5 domain, determined by NMR. Medaka TR has the
simplest and smallest CR4/5 domain and contains all the conserved elements, i.e. a three-way junction flanked by three helices P5, P6 and P6.1. The solution structure of mdCR4/5 revealed that P6 and P6.1 fold into a V-shape conformation, and P6 is stacked on P5 through the three-way junction. Although the TERT-free CR4/5 forms a three-way junction structure that is stabilized by well-ordered base pairs and stacking interactions at the junction region, none of these structural elements appear to be essential for telomerase activity. These results, together with the observed intrinsic dynamics of the three stems relative to each other, suggest that the three-way junction of mdCR4/5 could undergo a conformational rearrangement that transitions the TERT-free three-way junction and relative positions of P6 and P6.1 into the telomerase-active conformation.

Based on comparison of TR sequences from fungi, yeasts and vertebrates, a consensus three-way junction secondary structure was proposed, with a single conserved A nucleotide (A199 in mdCR4/5) connecting P6.1 with P6 and conserved UG connecting P6.1 to P5 (U212, G213 in mdCR4/5; in hCR4/5 there are two additional non-conserved nucleotides), and a conserved A-U base pair at the bottom of P6.1 (A200-U212 in mdCR4/5) (20). The predicted three-way junction secondary structure is different from what we observe in solution. In particular, the A200-U212 base pair does not form, rather U212 pairs with G198. Formation of this base pair results in a sharp turn in the backbone between P6 and P6.1. G•U base pairs seem to be important for stabilizing RNA backbone turns (56), e.g. in many junctions and at the sharp turns in tRNA^Phe (59). The conserved purines AA GAG (199–203) may also contribute to stabilization of the close contact between phosphates at the tight turn between P6 and J6/6.1, as these consecutive purines were identified as divalent cation binding sites. Given that disruption of the three-way junction base pairs observed in the mdCR4/5 solution structure had little effect on telomerase activity in vitro and the phylogenetic predictions for junction base pairing, it is likely that the free mdCR4/5 structure represents a stable inactive state. It can be envisioned that shifting G198-U212 to A200-U212 could release the tight turn between P6 and P6.1 helices and the compact base stacking from G198 to A200, which would induce the repositioning of P6.1 and P6 and/or open the junction region that might be important for interaction with TERT. We note that other rearrangements of the junction pairing, including tertiary interactions, might also occur in the complex of mdCR4/5 with TERT.

Previous telomerase activity assays with mutations on the conserved residues in the internal loop of hCR4/5 have suggested that some of the conserved three-way junction residues are involved in direct tertiary contact with TERT, and others may be required for organizing the internal loop conformation for TERT binding (29). Because the most important structural element in CR4/5 for TERT binding as well as telomerase activity is P6.1 formation (25), it is probable that the three other conserved nucleotides (A199, G213 and U214 in mdCR4/5) observed in the internal loop of CR4/5 might function to help ensure the formation of an active three-way junction structure and proper formation of P6.1.

Outside the junction, there are three regions of CR4/5 that are important in terms of structure, TRBD binding and activity: a conserved pyrimidine bulge in P6, the P6 loop and the P6.1 loop. The structure of mdCR4/5 P6.1 loop is well defined but is different from that of hCR4/5 (Supplementary Figure S1B). The hCR4/5 P6.1 loop contains pseudo-uridines (Ψ) at two potential positions (306 and 307), and structures for it have been determined both with and without these Ψs (32,58). In the absence of Ψ, the first three bases (U306, U307 and G308) in P6.1 loop shows some structural similarity to mdCR4/5 P6.1 loop, where U306 and G308 are positioned in the major groove with G308 stacking on U306, and the conserved U307 points out on the minor groove side. However, unlike mdCR4/5 P6.1 loop, the first U306 forms a wobble pair with G310 and the conserved G309 is flipped out to the minor groove. The structure of Ψ-modified hCR4/5 P6.1 loop shows different orientations of loop residues than the unmodified P6.1 loop, where Ψ306-G310 base pair is intact but G308 and G309 are all in the minor groove side and Ψ307 is stacked above Ψ308. Based on these structural differences, we propose that telomerase activation by P6.1 is loop sequence specific.

In the middle of P6, the conserved pyrimidine bulge nucleotide U182 sticks out from the major groove and results in the S-shape twist of the phosphate backbone. A similar kink is also observed at the conserved bulge (C262) in hCR4/5 P6 structure (Supplementary Figure S1A) (27). This unique structural feature, along with the deleterious effect on TRBD binding of deletion of mdCR4/5 U182 (53), and the minimal effect of U182→C substitution on activity, suggests that the conserved single pyrimidine bulge in vertebrate CR4/5 may affect proper positioning of the P6 loop for effective TERT binding or it may be important for TERT binding.

To gain insights into the potential interaction between the CR4/5 and TERT, we made a model structure of the mdCR4/5–mdTRBD complex based on the UV cross-linking data shown in a previous study, where three uridines, U182 in P6, U187 in P6 loop and U205 in P6.1 loop, were found to cross-link to mdTRBD residues Y503, F355 and W477, respectively (see ‘Materials and Methods’ section) (Supplementary Figure S5, left) (53). We tested modeling the complex using both Mg2+-free and -bound mdCR4/5 structures by manually placing each of these uridine bases to be within ~10 Å from the corresponding cross-linked amino acid. Interestingly, the Mg2+-bound CR4/5 structure, which has an inter-helical angle between P6 and P6.1 of 65° compared with 35° for the Mg2+-free CR4/5, fits better than the Mg2+−free one, in a way that all three cross-linking pairs can be better placed within the ~10 Å distance constraints. Given the mutational data and the intrinsic positional flexibility of the helices, it is plausible that the inter-helical angle between P6 and P6.1 can increase further in the mdCR4/5–TERT complex, which would bring mdCR4/5 closer to TRBD and allow some tertiary interactions between the conserved nucleotides in the junction and TRBD. In the...
with respect to each other, even in the presence of Mg2+.

...results in the P6, P6.1 and P5 stems being relative dynamic junction. The absence of inter-helical tertiary interactions same structure in the isolated forms and in the three-way packing of P6 and P6.1 were identified. The isolated indi-

...tions, the positions of the helices are often fixed by

...hammerhead ribozyme that performs precise phospho-

...riboswitches that recognize specific small metabolites or

...optimization of this interaction for activity.

STE loop (P6.1 loop in vertebrates) have co-evolved for

...U205 and G207 (U307 and G309 in hTR), has a large

...changes non-conserved nucleotides but not the conserved

...sequence from that of medaka, and changing the sequence

...conserved P6.1 loop nucleotides, U205A and

...Consistent with an interaction at the CTE, substitution

...TRBD and CTE together to close the TERT ring (53). It is possible that the three-way junction in

...rearranging to allow interaction between P6.1 and TERT at the TRBD-CTE interface.

COORDINATE DEPOSITION

Coordinates for the 20 lowest energy structures of mdCR4/5 have been deposited in the Protein Data Bank under accession code 2MHL, and chemical shifts have been deposited in the BioMagResBank under accession code 19634.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank E.J. Miracco for helpful comments.

FUNDING

National Institutes of Health [GM48123 to J.F.]; Korea Institute of Science and Technology [2V02900, 2E23810 (in part) to N.-K.K.]. Funding for open access charge: National Institute of Health [GM48123].

Conflict of interest statement. None declared.

REFERENCES


