



universität  
wien

# DISSERTATION

Titel der Dissertation

„The structural organization of the human telomerase“

Verfasser

Mag. Michael Wildauer

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt. Studienblatt:

A 094 490

Dissertationsgebiet lt. Studienblatt:

Molekulare Biologie

Betreuerin / Betreuer:

Mag. Dr. Christina Waldsich, Privatdoz.

# Table of Contents

1.	Introduction .....	8
1.1.	The importance of RNA .....	8
1.2.	Telomerase – a mysterious complex.....	9
1.3.	Telomerase RNA – a highly interesting RNA .....	10
1.3.1.	A pseudoknot entangles the template.....	12
1.3.2.	The CR4/CR5 domain – the activation region .....	16
1.3.3.	Maturation depends on the scaRNA domain.....	17
1.4.	TERT – a close relationship with telomerase RNA .....	19
1.4.1.	The TEN domain .....	20
1.4.2.	Telomerase RNA binding domain.....	21
1.4.3.	The reverse transcription domain.....	23
1.4.4.	The end of TERT – C-terminal extension .....	24
1.5.	Elongating the ends.....	26
1.6.	The shelterin complex is guarding the telomeres.....	28
1.7.	Biogenesis and maturation of hTR .....	30
1.8.	Telomerase and influence on disease .....	32
1.8.1.	Telomere-associated diseases – Dyskeratosis congenita .....	32
1.8.2.	Telomerase and cancer .....	33
1.9.	Investigating the structure of biological macromolecules.....	35
2.	Aims of the work.....	37
3.	Materials and Methods.....	38
3.1.	Bacterial strains.....	38
3.2.	PCR-based cloning strategies .....	38
3.3.	Agarose gels and restriction digests.....	39
3.4.	Competent cells and bacterial transformation .....	40
3.5.	Plasmid preparation and sequencing.....	41
3.6.	Culturing of HEK293 cells .....	41
3.7.	Transfection of HEK293 cells.....	42
3.8.	UV cross-linking and total RNA isolation.....	43
3.9.	Reverse transcription .....	44
3.10.	Denaturing PAGE.....	45
3.11.	SDS-PAGE.....	46

3.12.	Western blotting & chemiluminescence detection .....	48
3.13.	Gel-purification of DNA oligos.....	49
3.14.	5' end-labeling of DNA oligos.....	49
3.15.	Direct telomerase assay .....	50
3.16.	Data analysis.....	51
4.	Results.....	52
4.1.	Experimental setup – In vivo UV cross-linking of RNA.....	52
4.2.	The structural organization of human telomerase RNA .....	54
4.2.1.	Organization of the pseudoknot domain .....	54
4.2.2.	Assessing the conformation of the template region.....	58
4.2.3.	Only a few cross-links are found in the CR4/CR5 domain.....	60
4.2.4.	The CAB box is cross-linked within the scaRNA domain .....	62
4.3.	hTERT induces structural changes within hTR .....	65
4.3.1.	The triple helix is influenced by hTERT binding.....	66
4.3.2.	The template region is not altered upon hTERT binding .....	70
4.3.3.	The CR4/CR5 domain is strongly influenced by hTERT binding .....	72
4.3.4.	The scaRNA domain is left unaltered upon hTERT binding.....	77
4.4.	Identification of intramolecular RNA-RNA cross-links .....	80
4.4.1.	The structural organization of the pseudoknot domain .....	82
4.4.2.	The topology of the important CR4/CR5 domain .....	87
4.4.3.	The conformation of the scaRNA domain .....	88
5.	Discussion .....	90
5.1.	The structural organization of hTR in vivo .....	90
5.1.1.	The conformation of the pseudoknot domain <i>in vivo</i> .....	91
5.1.2.	The CR4/CR5 domain harbours only few cross-links .....	92
5.1.3.	The topology of the scaRNA domain.....	93
5.2.	hTERT-dependent conformational changes within hTR.....	94
5.2.1.	hTERT-induced conformational changes in the pseudoknot .....	95
5.2.2.	The influence of the CR4/CR5 domain on hTERT binding .....	96
5.3.	Identification of cross-linking partners .....	97
5.3.1.	The triple helix is formed <i>in vivo</i> .....	97
5.3.2.	The CR4/CR5 domain is in proximity of the triple helix .....	99
6.	References .....	101
I.	Addendum .....	i
I.	List of abbreviations.....	i

I.	List of figures.....	ii
II.	List of tables.....	iii
III.	List of oligonucleotides and plasmids.....	iv
IV.	Reagents and Kits.....	v
V.	Publications.....	vii
	RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction .....	vii
	Chemical probing of RNA in living cells .....	viii
VI.	Acknowledgements .....	ix
VII.	Curriculum vitae.....	x

## Abstract

Telomerase is a large RNP complex which adds repetitive elements to the telomeres, the ends of the chromosomes. By doing so, telomerase is responsible for the maintenance of the telomeres. The two key components responsible for this task are **a)** the telomerase RNA (TR) and **b)** the protein telomerase reverse transcriptase (TERT). The interplay of them allows telomerase to counteract the loss of information during cell division, which is due to the so-called end replication problem, while maintaining the proliferation potential of the cell. Furthermore, active telomerase permits a cell to divide itself beyond the Hayflick limit, thereby literally becoming immortal. However, unlimited as well as uncontrolled cell division is a hallmark of most cancer types known today and it is not surprising that telomerase is active in over 90% of known cancer types. Despite the bio-medical importance of telomerase, our understanding of the structure of telomerase and its mechanics is still in its infancy. We are interested in elucidating the structural conformation of human telomerase RNA (hTR) within the telomerase RNP complex. So far, most studies on telomerase have been carried out *in vitro* and did not take into account most of the accessory proteins present in the telomerase RNP complex. To compensate for this, we decided to use an *in vivo* approach. By using an *in vivo* UV cross-linking assay, we obtained structural information about key elements within human telomerase RNA which can be divided into three domains: **a)** the pseudoknot, **b)** the CR4/CR5 domain and **c)** the scaRNA domain. In total we identified 70 cross-linked nucleotides in the human telomerase RNA. With 46 cross-links the majority is located in the pseudoknot domain, in detail in the region which forms the conserved triple helix. The remainder is distributed between the CR4/CR5 domain and the scaRNA domain. Binding of hTERT strongly influences the cross-linking pattern of nucleotides in the pseudoknot region as well as in the CR4/CR5 domain. In detail, multiple nucleotides in the triple helix as well as in the corresponding linker J2b/3 show drastic changes in their cross-linking intensity. While most of the residues show an increase in cross-linking intensity in absence of hTERT, C104 is the only nucleotide for which a decrease in cross-linking intensity was observed. This makes C104 in J2b/3 a prime candidate for a direct cross-link to hTERT. For the CR4/CR5 domain, the most significant changes were observed in the P6.1 stem and the adjacent junction J6.1/5. The observed changes in the cross-linking pattern are similar to those observed in the crystal structure of mTERT bound to mTR in medaka. We therefore suggest that binding of hTERT might also lead to drastic rearrangements in the three-way junction in the CR4/CR5 domain using a similar mechanism to mTR, where P6.1 acts as a conformational switch upon mTERT binding. To identify the interaction partner of the previously identified cross-links, we created mutated hTR constructs. We were able to identify several spatial constraints for hTR. For example, we have good evidence that the triple helix is indeed formed *in vivo*. In addition, the template region as well as the CR4/CR5 domain is in close proximity to the pseudoknot domain. This is in good agreement with previous studies, which showed an

involvement of residue U307 in catalysis. Furthermore, it is likely that the cross-links in the 5' pocket and in scaRNA domain are directed to H/ACA snoRNA binding proteins, as we do not observe changes in the cross-linking pattern independent of hTERT. Finally, our data is in good agreement with *in vivo* DMS probing of hTR. In summary, our results provide new insights into the structural organization of human telomerase RNA and refine the current model of telomerase structure.

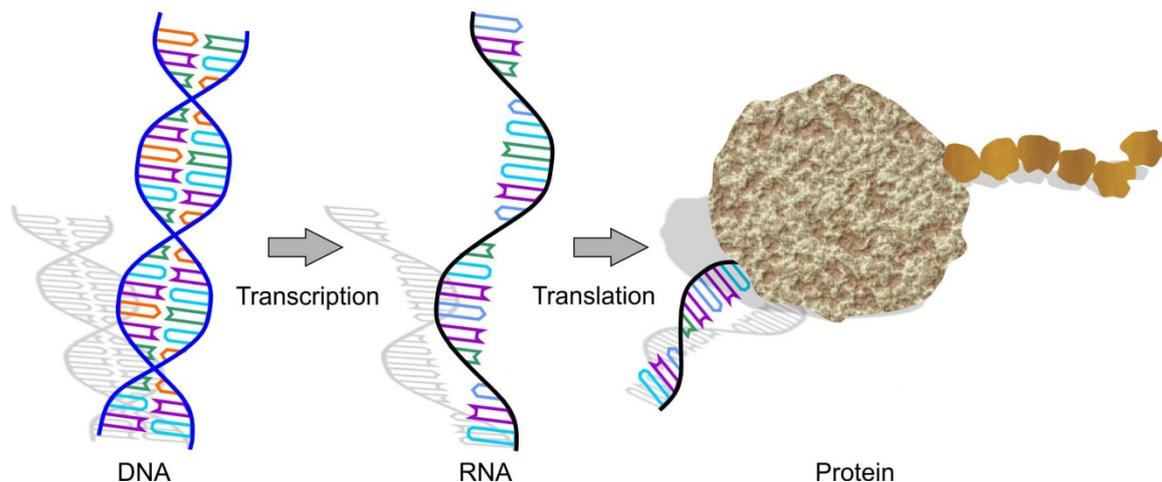
Telomerase ist ein großer Ribonucleinkomplex, dessen Aufgabe die Synthese von repetitiven Elementen am Ende der Chromosomen ist. Dadurch trägt die Telomerase zur strukturellen Integrität der Chromosomenenden, den Telomeren bei. Diese Aufgabe wird von den zwei Hauptbestandteilen der Telomerase erledigt: **a)** die RNA Komponente, genannt Telomerase RNA (TR) und **b)** dem dazugehörigen Protein, der Telomerase Reverse Transcriptase (TERT). Der Telomerasekomplex kann dadurch die ständige Verkürzung der Chromosomen nach einer Zellteilung, dem sogenannten Endreplikationsproblem, verhindern und die Solllänge der Telomere erhalten. Telomerase wird dadurch zu einem wichtigen Faktor, um die Proliferationskapazität einer Zelle zu erhalten. Einerseits ist das für sich schnell teilendes Gewebe wie blutbildende Zellen und diversen Schleimhäuten wichtig, auf der anderen Seite befähigt Telomerase auch die Entstehung von Krebszellen. Durch das Potential, sich unbegrenzt teilen zu können, und per Umgehung von natürlichen Regelmechanismen wie Apoptose und Seneszenz, kann eine gesunde Körperzelle sich zur Krebszelle entwickeln. Dieses zweischneidige Schwert macht Telomerase zu einem interessanten Ziel für Krebstherapien. Leider ist unser Wissen über die genaue Struktur und den Aufbau der Telomerase noch immer sehr eingeschränkt, da sehr viele Studien auf Daten aus *in vitro* Experimenten stammen und sich daher nur bedingt auf eine lebende Zelle übertragen lassen. Wir haben uns daher entschlossen, einen neuartigen Versuchsaufbau zu verwenden, der auf UV Vernetzung basiert, aber *in vivo* durchgeführt wird. Damit haben wir 70 Nukleotide identifiziert, die vernetzt sind. Die überwältigende Mehrheit von 46 Nukleotide befindet sich im Pseudoknoten, während sich der Rest auf die CR4/CR5 Domäne sowie die scaRNA Domäne verteilt. Wir haben gezeigt, dass die Bindung von hTERT das Vernetzungsmuster der Nukleotide im Pseudoknoten, genauer gesagt im Bereich der konservierten Dreifachhelix, signifikant beeinflusst. Die meisten Nukleotide zeigen eine Steigerung der Vernetzungsintensität, nur C104 zeigt eine Reduzierung. Diese Beobachtung lässt den Schluss zu, dass C104 einen direkten Kontakt mit dem Protein hTERT ausgebildet haben könnte und daher eine wichtige Rolle in der Bindung von hTERT an hTR spielt. Der zweite Bereich, in dem wir große Veränderungen im Vernetzungsmuster festgestellt haben, ist die konservierte CR4/CR5 Domäne. Hier spielt besonders die Helix P6.1 eine große Rolle als auch die anschließende Kreuzung J6.1/5. Das festgestellte Vernetzungsmuster legt nahe, dass die Bindung von hTERT über einen ähnlichen

Mechanismus erfolgen könnte, der in Medaka identifiziert wurde. Hier findet eine komplette Umgestaltung der Dreivegekreuzung statt und P6.1 fungiert als Schalter, je nachdem hTERT gebunden ist oder nicht. Anhand unserer Daten postulieren wir einen ähnlichen Bindemechanismus für die humane Telomerase RNA. Letztlich haben wir für die identifizierten Vernetzungen den Interaktionspartner bestimmt. Dafür haben wir mutierte hTR Konstrukte verwendet, die es uns erlaubt haben, einige räumliche Begrenzungen für das Molekül aufzustellen. Erstens zeigen unsere Daten, dass der Pseudoknoten sich in räumlicher Nähe zur Vorlage und der CR4/CR5 Domäne befindet. Diese Beobachtung ist im Einklang mit früheren Studien, die gezeigt haben, dass Nukleotid U307 für die Katalyse eine wichtige Rolle spielt. Zusätzlich postulieren wir, dass die Vernetzungen an den Nukleotiden in der scaRNA Domäne höchst wahrscheinlich zu einem der H/ACA snoRNA-Bindeproteine gehen, da wir keine Änderungen im Vernetzungsmuster feststellen konnten, unabhängig ob hTERT vorhanden war oder nicht. Zusammengefasst helfen unsere Daten, das bereits vorhandene Modell der humanen Telomerase RNA weiter zu vervollständigen um einen noch genaueren Einblick zu erhalten.

# 1. Introduction

## 1.1. The importance of RNA

Over the last three and a half billion years, life on earth has developed an astonishing number of species. The domain of eukaryotes alone is thought to contain roughly ten million species<sup>1</sup> and this does not include the two other domains of life (prokaryotes and archaea). Each of these organisms consists of DNA, protein and RNA in a simplified view. The hierarchy between these completely different classes of molecules was strongly shaped by Francis Crick's central dogma of molecular biology: DNA is the template for RNA which in turn is the template for proteins (figure 1.1).



**Figure 1.1. Central dogma of molecular biology.** In a cell the flow of information starts with double-stranded DNA, which is transcribed by the cellular machinery to single-stranded RNA and used by ribosomes during translation as a blue print to synthesize the majority of cellular proteins. This idea was first brought up by Francis Crick.

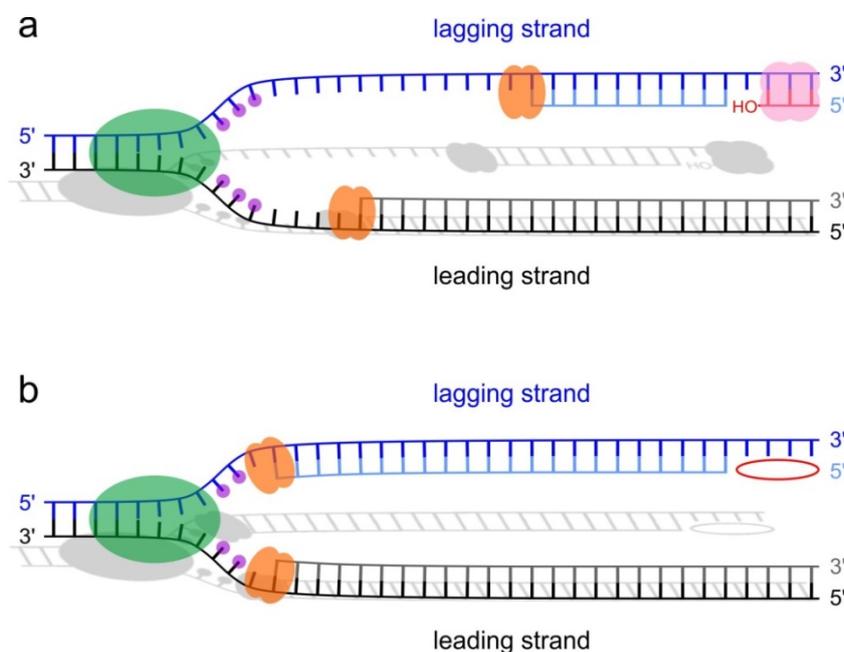
In this classification the role of RNA was understood as a simple messenger. This perception changed completely during the 1980s with the discoveries of Sidney Altman and Thomas Cech<sup>2,3</sup>. "For their discovery of catalytic properties of RNA" they were jointly awarded the Nobel prize in 1989<sup>4</sup>. Now the scientific awareness of RNA was altered and it became evident that RNA plays a much more profound role than just being a simple messenger. Today it is commonly accepted that RNA is involved in almost every aspect of metabolism in the living cell (e.g. mRNA, tRNA and rRNA in protein synthesis, miRNA and siRNA in gene regulation, snoRNA in RNA modification, and lncRNA in epigenetic regulation, among others).

Which properties allow RNA to carry out such a variety of tasks? Why is it impossible for DNA to fulfil similar tasks as well? First, DNA is almost exclusively found to be double-stranded, while RNA is most of the time single-stranded. When forming a double helix the presence of the 2'-OH group at the ribose forces the molecule to adopt another helix geometry than a similar DNA molecule. While DNA

is most commonly found to adopt the so-called B-form (wide major groove, narrow minor groove), RNA assumes the A-form (narrow and deep major groove, shallow minor groove). These differences strongly influence how a protein can bind and/or interact with one of these molecules. Second, uracil (U) instead of thymine (T) allows, besides to the four canonical Watson-Crick base-pairs, the possibility of forming a G-U wobble base-pair as well. Together with the ability of a RNA molecule to fold back upon itself, the number of possible pairing combinations drastically increases and allows the formation of complex tertiary architectures. In this regard the 2'-OH group of the ribose also plays an important role as H-donor and -acceptor. Taken together, chemical differences allow a RNA molecule to carry out many different tasks within the cell.

## 1.2. Telomerase – a mysterious complex

One of the many different tasks RNA can fulfil inside a cell is the maintenance of the chromosome ends. The importance of the so-called telomeres became evident in the 1930s, when McClintock and Muller observed independently that the ends of chromosomes were protected from end-to-end fusion events which usually take place at the site of a double-strand break<sup>5,6</sup>. Based on this finding McClintock proposed a special role and structure for the ends of the chromosomes and named them telomeres<sup>5</sup>. However, the exact organization and function remained enigmatic for decades. In 1961 the American gerontologist Leonard Hayflick proved the theory of unlimited cell division wrong. Hayflick discovered that, contrary to established theories, cells cultured *in vivo* can only divide a certain number of times and undergo programmed cell death (apoptosis) when exceeding this number<sup>7</sup>. The Hayflick limit acts as a cellular clock, thus enabling a cell to count how many divisions it already made. The fact that some germ line cells and many tumour-derived cell lines are not affected by the Hayflick limit remained a mystery.

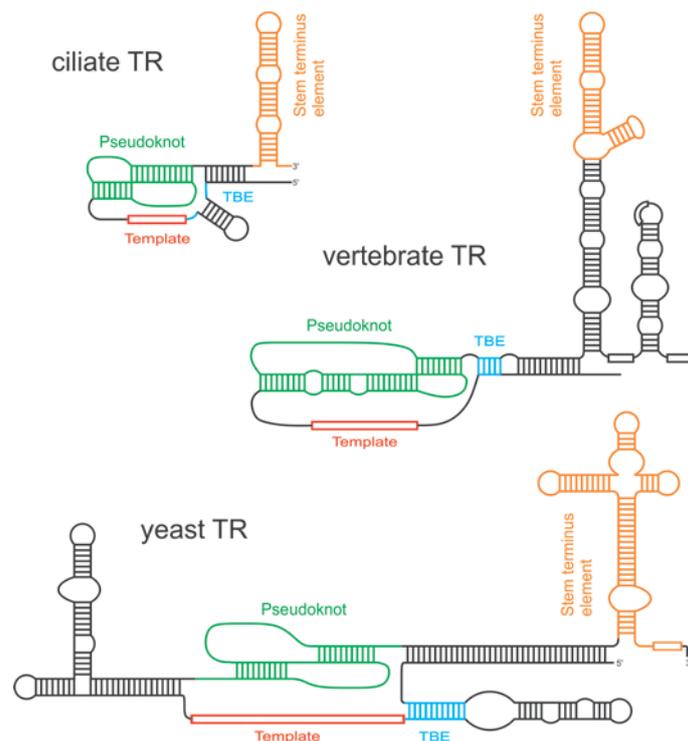


**Figure 1.2. Schematic representation of DNA synthesis.** **a)** The helicase (green ellipse) unwinds the DNA helix, while single-strand binding proteins (purple circles) prevent base pairing. While the leading strand (black) is continuously replicated by the DNA polymerase (depicted in orange), the lagging strand (blue) is dependent on Okazaki fragments with their free 3'-OH synthesized by primase (pale red). **b)** Okazaki fragments are removed and the DNA is synthesized in the gaps. On the 3' end (e.g. at the telomeres) of the lagging strand DNA cannot be replicated and information is lost (depicted by red ellipse).

The next important step on the journey towards the discovery of telomerase took place ten years later, in 1971. The Russian scientist Alexey Olovnikov formulated the so-called end replication problem<sup>8</sup>: during DNA replication the lagging strand cannot be replicated completely due to intrinsic mechanics of DNA replication, which only works in 5' to 3' direction and requires a free 3'-OH group to form the bond to the next nucleotide. As a consequence the last nucleotides from the 3' end of the lagging strand are not replicated and the chromosome length decreases with each round of replication (figure 1.2). Olovnikov proposed the following solution in 1971: a) the existence of repetitive sequences at the 3' end of chromosomes which act as a buffer during DNA replication and preserve the genetic information<sup>8</sup>; b) the existence of a special DNA polymerase in germ line and tumour cells which can completely synthesize the lagging strand during DNA replication<sup>8</sup>. In addition, he concluded that the Hayflick limit is the consequence of losing too much of the repetitive sequences and therefore forcing the cell into senescence or apoptosis<sup>8</sup>. Finally, in 1989 the long-proposed enzyme was discovered in *Tetrahymena thermophila* by Carol Greider and Elisabeth Blackburn and named telomerase<sup>9</sup>. It turned out that telomerase counteracts the shortening of the telomeres by using an intrinsic RNA template to elongate the 3' end of the telomeres with repetitive sequences<sup>9</sup>, just as proposed by Olovnikov almost 20 years earlier. Interestingly, telomerase is strongly down-regulated in most somatic cells and is only found to be active during embryogenesis and in stem cells, hematopoietic cells and about 90% of all known cancers<sup>10</sup>. Nowadays telomerase even found its way into popular science and media. Located at the intersection between longevity and cancer, telomerase is often seen as the mythical fountain of youth. This perception seems to be overrated and wishful thinking. However, recent findings clearly showed that telomerase is involved in many human diseases and its regulation affects many types of cancer (see section 1.8). In summary, it took nearly 80 years from the first observations by Muller and McClintock to the actual discovery of telomerase by Greider and Blackburn<sup>5,6</sup>. Countless studies have investigated the role of telomerase in different species and deepened our understanding of this fascinating complex. **Telomerase RNA – a highly interesting RNA**

The telomerase holoenzyme consists of two major players: **a)** the telomerase RNA (TR, TER, TERC) and **b)** the protein telomerase reverse transcriptase (TERT). These two compounds are sufficient *in vitro* to obtain a functional telomerase complex, however *in vivo* the situation is different, as many

accessory proteins are needed to form a functional complex<sup>11-13</sup>. Telomerase RNA is either transcribed by RNA polymerase III (*T. thermophila*) or RNA polymerase II in human followed by extensive RNA processing (see section 1.7)<sup>14-16</sup>. In general, telomerase RNAs often have a very high GC content and are therefore challenging to work with. Despite being found in every species with active telomerase, TR shows a considerable variety in length among the different species<sup>17</sup>. The shortest telomerase RNA with a length of about 150 nucleotides was found in ciliates, whereas yeast species have the longest TRs described so far with roughly 1300 nucleotides<sup>18</sup>. With all TRs carrying out the same function in the cell, structural rather than sequence conservation is a key characteristic of these RNAs<sup>17</sup>. Indeed, every single telomerase RNA discovered so far possesses the following structural elements (figure 1.3): **a**) the template region and the template boundary element which determine the repeat sequence added to the 3' end of the telomere; **b**) the pseudoknot, which is indispensable for telomerase activity; and **c**) the stem terminus element (STE) which binds to TERT and is important for telomerase activity. In addition, in vertebrates and yeast there is one additional region: the scaRNA domain which is involved in telomerase biogenesis, maturation and localization<sup>19</sup>.



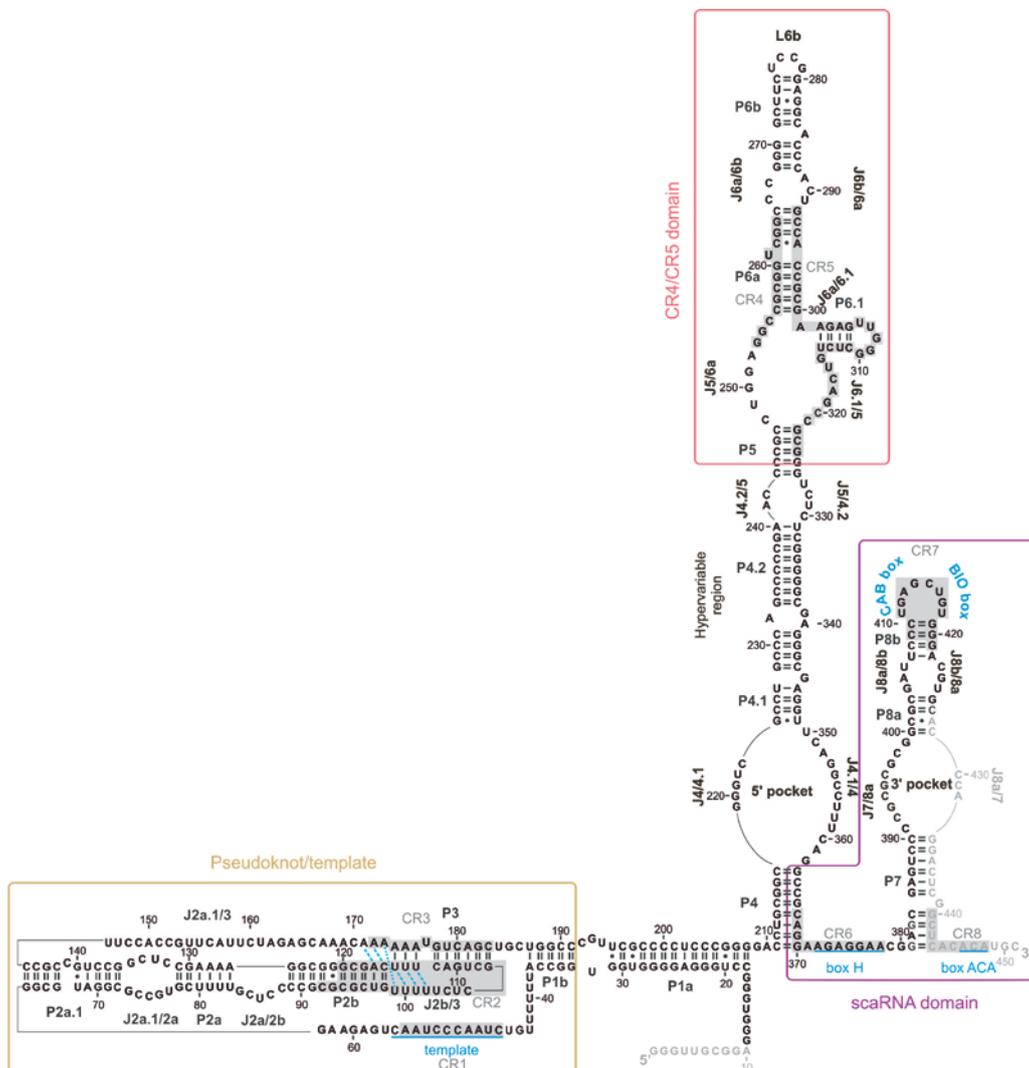
**Figure 1.3. Comparison of different telomerase RNAs.** Schematic representation of telomerase RNAs from ciliates, vertebrates and yeasts. The common elements are highlighted in different colors: the template (red box), the pseudoknot (green), the template boundary element (TBE; blue) and the stem terminus element (STE; in orange).

The presence of highly complex elements such as the pseudoknot implies that the TR must adopt a well-defined structure. This process is called RNA folding and is a common process for most RNAs

found within a cell. In the next sections the hallmarks of telomerase RNA will be explained in more detail.

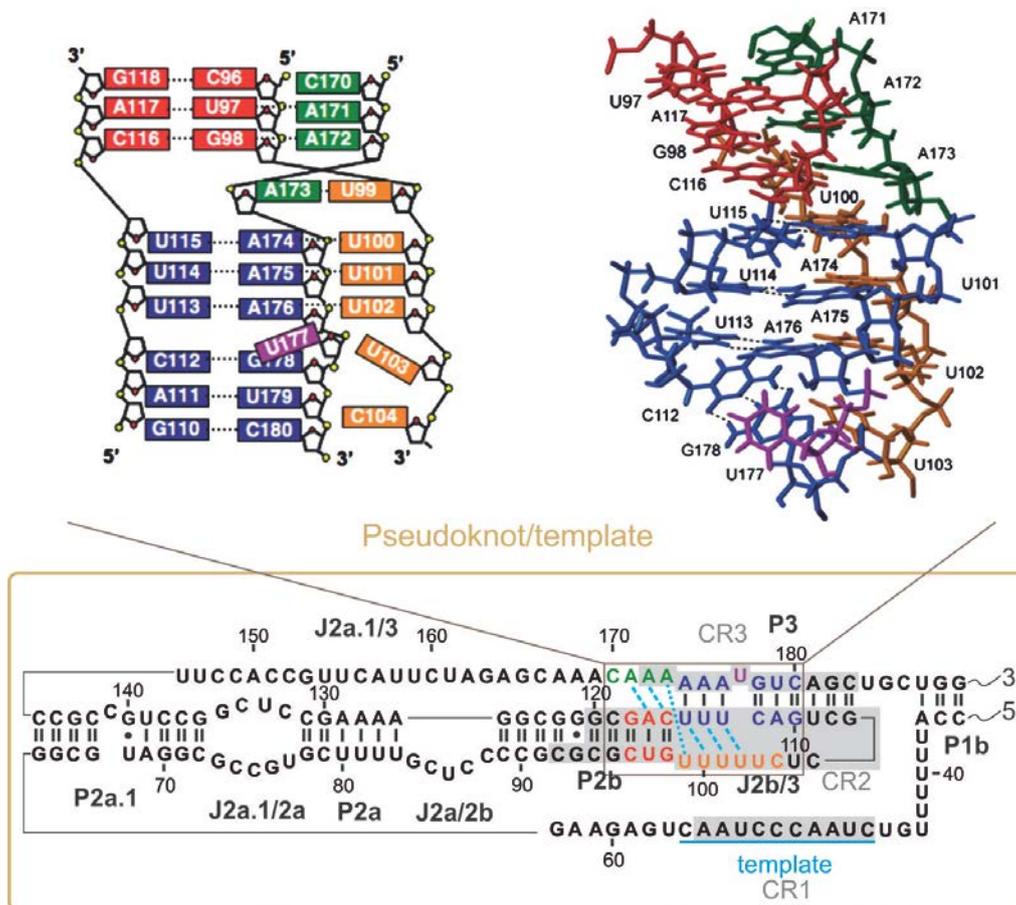
### 1.3.1. A pseudoknot entangles the template

Phylogenetic studies by Chen et al. identified eight conserved regions (CR1-8) within vertebrate telomerase RNAs<sup>17</sup> (figure 1.4) with CR1-3 being located in the template and pseudoknot domain. CR1 represents the intrinsic RNA template used by the telomerase complex to elongate the telomeres<sup>17</sup>. It is flanked by the template boundary element (TBE), which consists of stem P1b, and stem P2a.1<sup>20</sup>. In ciliates, an additional element called template recognition element (TRE) was identified 3' of the actual template, which is also involved in template recognition and positioning<sup>20</sup> but absent in vertebrates.



**Figure 1.4. Secondary structure of the human telomerase RNA.** Human telomerase RNA can be divided into three domains. The pseudoknot/template domain is boxed in red, whereas the CR4/CR5 domain is boxed in red and the scaRNA domain in purple, respectively. Interesting elements are depicted in blue. The base triples in the pseudoknot are indicated by blue dashed lines and the non-canonical U•A base pair is marked with a blue dotted line.

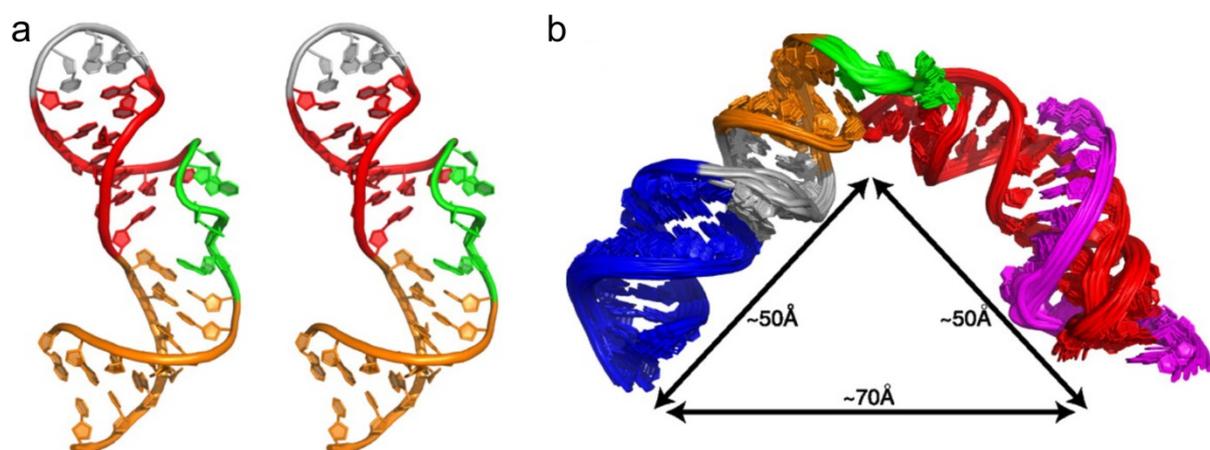
The TBE plays an important role in defining which nucleotides are reverse transcribed, prevents read-through and template translocation<sup>18,20,21</sup>. Telomere elongation is thought to be carried out according to the accordion RNA model proposed by Berman et al<sup>20</sup>. In short, TBE (and TRE in ciliates) contract and stretch themselves, while nucleotide addition and template translocation take place (see section 1.5). The template consensus sequence (CR1) for vertebrates is 5'-CUAACCCUAA-3' resulting in the hexameric telomere repeat 5'-GGTTAG-3'<sup>17</sup>. The sequence changes with evolutionary distance. In contrast, the template length and the resulting repeats are not strictly conserved. For example, invertebrates, ciliates and plants often have pentameric repeats, while fungi often have repeat lengths of 15 to 25 nucleotides<sup>22</sup>. In addition to the template region the core domain of human telomerase RNA (nucleotides 1 to 209) contains a pseudoknot as well. It is linked to the template region via a single-stranded linker and can be divided into the core pseudoknot (P2b/P3) and the extended pseudoknot (including P2a and P2a.1).



**Figure 1.5. Secondary structure map of the hTR core domain.** The 5' domain of human telomerase RNA with nucleotides 35 to 189 is shown. Gray areas refer to the conserved regions CR1 to CR3. The area boxed in brown is depicted on the top with the left panel showing the bases in the pseudoknot. In the right panel the NMR solution structure of the pseudoknot is shown. The top panel was adapted from Kyoon et al., 2008<sup>23</sup>.

In human telomerase RNA (hTR) the core pseudoknot is formed by the helices P2b, P3 and the single-stranded areas J2b/3 and J2a.1/3 (figure 1.5). The conserved regions CR2 and CR3 correspond to elements forming the pseudoknot<sup>17</sup> (figure 1.5). The importance of the pseudoknot was revealed by the identification of disease-related mutations (Dyskeratosis congenita, aplastic anaemia), which reduced or even abolished telomerase activity in those patients<sup>24</sup>. A NMR solution study of the hTR pseudoknot by Theimer et al. provided structural insights into the loss of function: an extensive network of tertiary interactions was identified at the pseudoknot, resulting in the formation of a triple helix<sup>25</sup>. Of utmost importance are several Hoogsteen base triples (figure 1.5) between P3 and J2b/3 (U113-A176•U102, U114-A175•U101 and U115-A174•U100, and), and P2b and J2a.1/3 (A117-U97•A171 and C116-G98•A172) as well as a non-canonical base pair formed between J2b/3 and J2a.1/3 (U99•A173)<sup>25</sup>. All of these nucleotides show an extremely high degree of conservation among vertebrates (80% for A171, 90% for A173 and U115, 100% for the remainder) underlining their crucial role for telomerase activity<sup>17</sup>.

The P2b helix of the core pseudoknot expands towards the template to form the extended pseudoknot. The elements starting from the template 3' end towards P2b are stems P2a.1 and P2a separated by an internal loop (J2a.1/2a and J2a/2a.1). The extended pseudoknot is connected to the core pseudoknot via the bulge J2a/2b. Helix P2a.1 and the adjacent internal loop J2a.1/2a are specific for mammalian and extend helix P2a<sup>17,26</sup>. Correct formation of P2a.1 is needed for telomerase activity in human, however mutational analysis revealed that the helical character of this element is more important than the nucleotide sequence of P2a.1<sup>27</sup>.



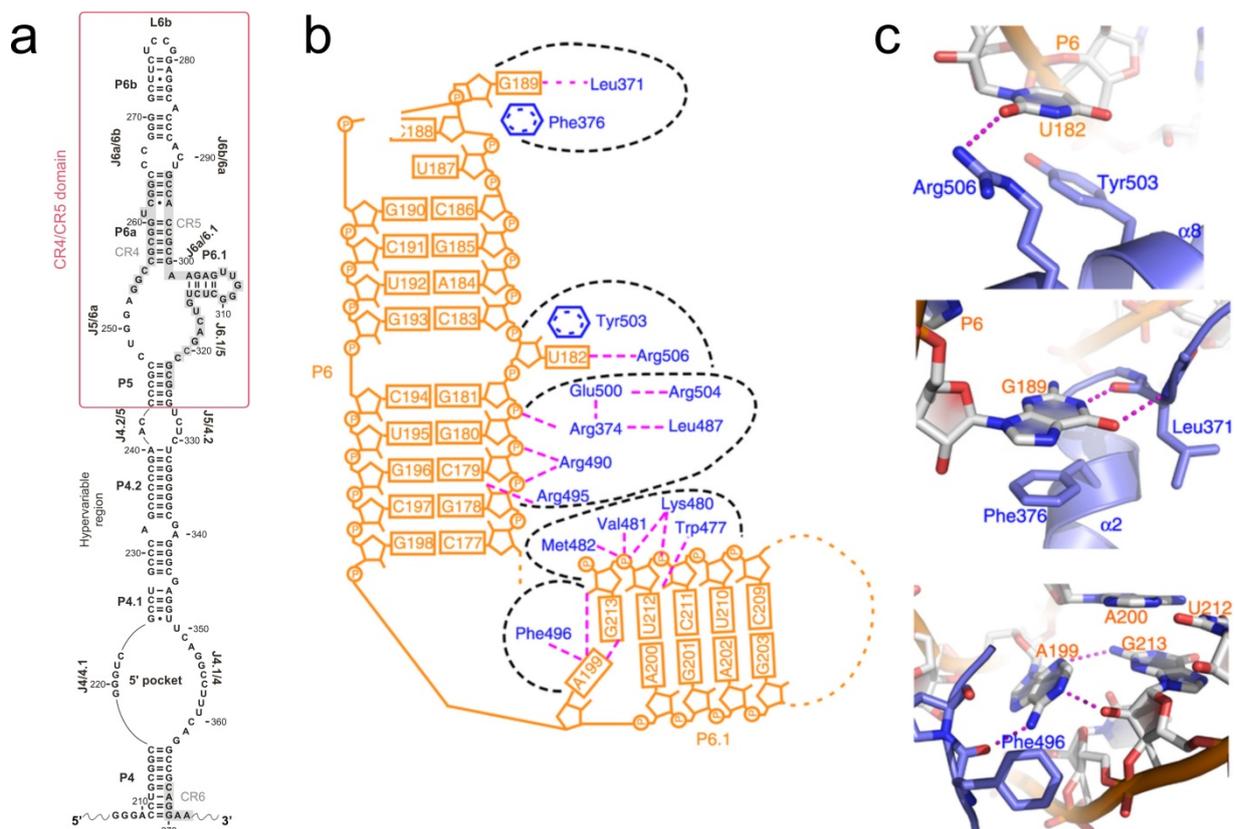
**Figure 1.6. Geometry of J2a/2b in hTR.** **a)** A stereo view of the NMR solution structure for junction J2a/2b. Nucleotides in stem P2a are coloured orange, stem P2b is shown in red and the five nucleotides of J2a/2b are depicted in green. Nucleotides added for loop closure are shown in gray. **b)** Computational model of the hTR pseudoknot. Stem P2a.1 is coloured blue, the internal loop J2a.1/2a and J2a/2a.1 is shown in gray and stem P3, which forms the pseudoknot together with P2b, is depicted in magenta. All other elements are coloured as in **a)**. Figure was adapted from Zhang et al., 2010<sup>28</sup>.

Junction J2a/2b is found in all vertebrates TRs and is strongly conserved in mammals<sup>17</sup>. The length of this asymmetric internal loop is 5 nucleotides<sup>28</sup>, effectively dividing helix P2 into two parts, P2a and P2b, respectively. Among mammalian TRs the first 5' nucleotide is a highly conserved G, whereas the remainder of the loop consists mostly of pyrimidines (5'-GCUCC-3' in human)<sup>28</sup>. This has major implications for the geometry of this region<sup>28</sup>. The NMR solution structure revealed that the presence of this bulge introduces a sharp bend between helices P2a and P2b with limited interhelical flexibility and therefore strongly determines the core pseudoknot structure<sup>28</sup> (figure 1.6). The resulting shape resembles an inverted V with a distance of approximately 70 Å between the ends of the extended pseudoknot (G64-C145 in P2a.1 and G107-C183 in P3). These distance constraints have important implications for telomerase function according to a computational model by Zhang and colleagues<sup>28</sup>: according to their simulation the fully base-paired RNA-DNA duplex between template and telomeric DNA contains ten base pairs which equal 26 Å in length. Together with the single-stranded nucleotides flanking the template both on the 3' side as well as on the 5' side, the RNA-DNA duplex can be placed between the two ends of the pseudoknot<sup>23</sup>. However this is only made possible by the sharp bend introduced via J2a/2b because a flat pseudoknot would be too long to be spanned by the RNA-DNA duplex and its flanking sequences. Mutating the bulge residues, moving the bulge to the other strand or shortening it drastically decreased telomerase activity (less than 10% of WT activity), underlining the importance of this structure<sup>28</sup>. Interestingly, the other internal loop composed of J2a.1/2a and J2a/2a.1 does not seem to introduce bending or flexibility between its flanking helices P2a.1 and P2a and is therefore not thought to have such drastic implications on the pseudoknot topology like J2a/2b<sup>28</sup>.

Despite this large amount of available structural information, the role of the pseudoknot for template positioning and hTERT binding is still unanswered. So far studies have failed to pinpoint specific nucleotides important for binding to hTERT. One explanation could be that several different interaction sites exist and result only in weak binding of the protein, thus implying a role for the pseudoknot in template translocation during catalysis<sup>25</sup>. On the other hand, the hTR-hTERT interaction may not depend on contacts to nucleobases but to the backbone of hTR. In contrast, with A176 and U177 two nucleotides which directly contribute to catalysis and to pseudoknot folding have been identified<sup>23,29</sup>. Being part of helix P3, both of them are positioned in a way that allows them to interact with the template and hTERT<sup>28</sup>. Substitution of the 2'-OH group of A176 to 2'-H drastically decreased telomerase activity<sup>29</sup>, while deletion of U177 changed the pseudoknot structure by destabilizing the Hoogsteen base triples<sup>23</sup>.

### 1.3.2. The CR4/CR5 domain – the activation region

The next region of interest, the stem terminus element or CR4/CR5 domain in vertebrates, is connected to the pseudoknot domain via a hypervariable region in which nucleotide conservation between species and even mammals is extremely low<sup>17</sup>. The secondary structure differs between species with a single terminal stem loop in ciliates or a stem loop emerging from a three-way junction as found in human and yeast<sup>17,30</sup> (figure 1.3). The role of the CR4/CR5 domain in telomerase assembly and function was established over the last years. In vertebrates the P6.1 helix is absolutely essential and when depleted telomerase activity *in vivo* is virtually abolished<sup>31</sup>. Furthermore, *in vitro* UV cross-linking studies showed that the CR4/CR5 domain might be in close proximity to the template, with long-range tertiary interactions formed between the loop L6.1 and the 5' as well as the 3' end of the template (A54-U307 and U306 to G44, U45 and C46)<sup>32</sup>. In addition, helix P6.1 also plays a role in binding to TERT, as mutants lacking the P6.1 helix show no detectable telomerase activity *in vitro* and are also not able to bind TERT<sup>31</sup>.



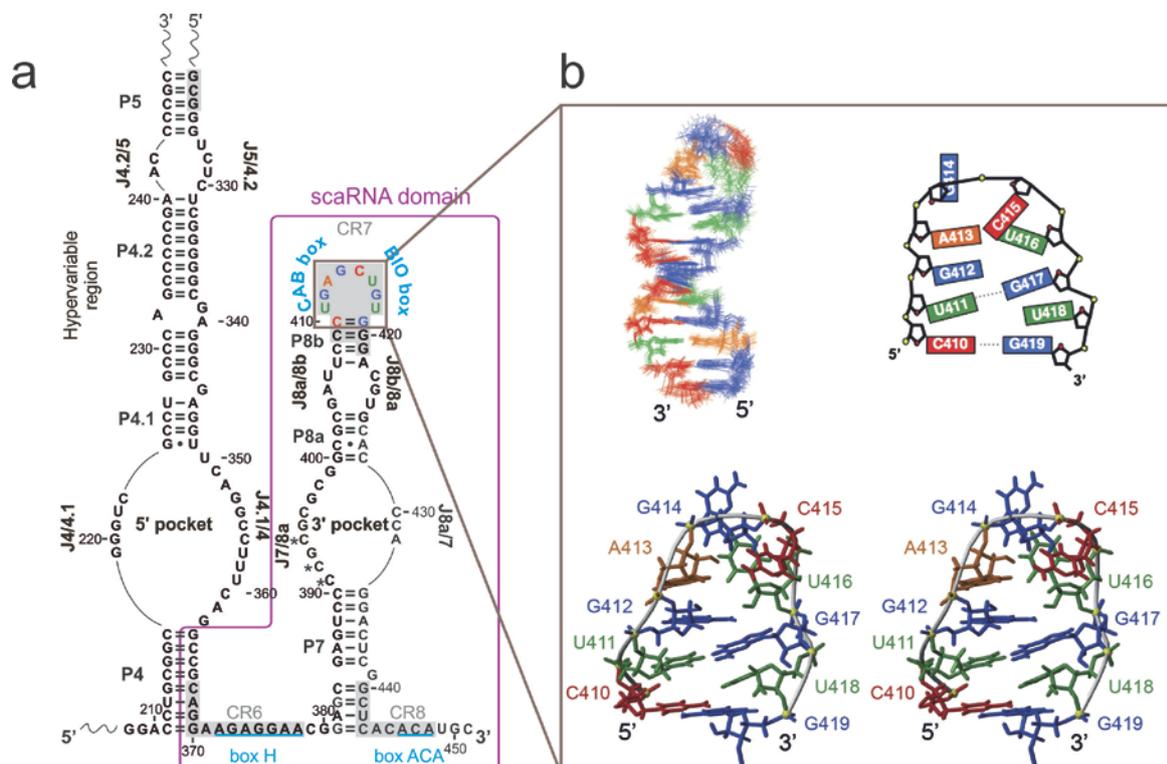
**Figure 1.7. The TRBD of TERT interacts with the CR4/CR5 domain in telomerase.** a) Secondary structure map of the human CR4/CR5 domain and the hypervariable region. Labelling is the same as in figure 1.4. b) Schematic representation of P6/P6.1 from medaka (orange) and its interaction partners in medaka TRBD (blue). H-bonds are depicted as purple dashed lines; black dashed lines represent contact patches. c) Detailed neighbourhood of U182, G189 and A199 when bound to mTERT. Figure adapted from Huang et al., 2014<sup>33</sup>.



domain, but this domain is absolutely needed *in vivo* for telomerase biogenesis<sup>13,16,35,36</sup>. So what is the exact role of the scaRNA domain?

Originally, scaRNAs (small Cajal body-specific RNAs) have been identified as a subclass of snoRNAs (small nucleolar RNAs), which translocate to the Cajal body instead of the nucleolus<sup>37,38</sup>. Both scaRNAs and H/ACA snoRNAs are involved in pseudouridylation of other RNAs, a task which is carried out by RNP complexes<sup>37-39</sup>. In addition to snoRNAs, which are characterised by the conserved H box (5'-ANAGGA-3') and an invariant ACA trinucleotide, scaRNAs possess an additional element, called CAB box (Cajal body box) consisting of 5'-UGAG-3'<sup>36-38</sup>. Adjacent to the CAB box, the BIO box is located, which is important for RNA accumulation, while the CAB box is required for correct localization to the Cajal bodies (see section 1.7)<sup>19,40</sup>. All of these elements are embedded in the secondary structure in single-stranded regions (figure 1.9) with box H in the hinge region, the CAB and BIO boxes in the terminal loop of the second hairpin (L8b in hTR) and the ACA trinucleotide in the tail<sup>40</sup>.

The NMR solution structure for CR7 harbouring CAB and BIO box and snoRNA U64 revealed some similarities despite different nucleotide composition in the terminal loops: both loops have a U•G wobble base pair formed by the first nucleotide (U) and the second to last nucleotide (G) in the loop; the terminal nucleotide of the loop is a uridine in both cases which is always unpaired<sup>41</sup> (figure 1.9). Taking this into account, Theimer et al. concluded that the first nucleotides proximal to the P8b stem are involved in hTR processing and accumulation<sup>41</sup>.



**Figure 1.9. Solution structure of the terminal hairpin of the scaRNA domain.** **a)** The stem terminus element and its functional elements are depicted. Box H is colored in light green, CAB and BIO box are shown in purple and orange, respectively. The ACA trinucleotide is shown in red. The CR7 region is boxed and enlarged in **b)**. Upper left panel: Superposition of the 20 lowest energy structures for the CR7-containing stem loop. Upper right panel: Schematic representation of nucleotides within the CR7 region. Lower panel: Stereo view of the CR7 region based on the NMR solution structure (PDB 2Q2H). Figure adapted from Theimer et al., 2010<sup>42</sup>.

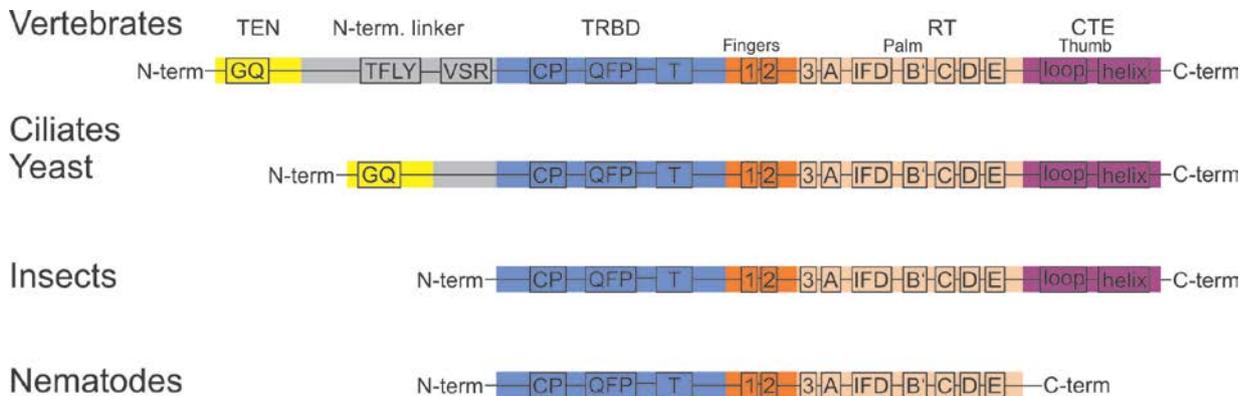
Taken together, the human telomerase RNA can be divided into three major domains with the first two being important for catalysis (pseudoknot/template domain and the CR4/CR5 domain), whereas the third one (scaRNA domain) is important for biogenesis and localization of hTR<sup>23,25,35,43,44</sup>. However, for a functional telomerase RNP complex, the protein TERT must be present as well.

#### **1.4. TERT – a close relationship with telomerase RNA**

With the telomerase RNA providing the necessary template for synthesizing the telomeric repeats, the catalytic core of the telomerase RNP complex is located on the protein part named telomerase reverse transcriptase (TERT). The class of reverse transcriptases was discovered independently in 1970 by Howard Temin and David Baltimore<sup>45,46</sup>. Until then, every known DNA polymerase was dependent on a DNA template (e.g. the leading strand during DNA replication, figure 1.2) to synthesize the complementary DNA strand. Instead, the reverse transcriptase relies on a RNA template to synthesize the complementary DNA strand. In the case of the telomerase RNP complex the RNA template is located within the telomerase RNA (see section 1.3.1). Human TERT (hTERT) consists of 1132 amino acids, has a molecular mass of approximately 127 kDa and can be divided into four functional domains (figure 1.10): **a)** the TEN domain (telomerase essential N-terminal domain), **b)** the TRBD domain (telomerase RNA binding domain), **c)** the RT domain (consisting of fingers and palm) and **d)** the CTE (C-terminal extension), also called thumb domain<sup>47</sup>. Interestingly, the RT domain shows remarkable similarities to RT motifs of viral origin, like that of the HI-virus, which could point to a common ancestral mechanism<sup>48</sup>. In contrast to the highly conserved RT domain, other elements of TERT are species-dependent. For instance, the TEN domain as well as the thumb motif are only found in ciliates and vertebrates, but are completely absent in insects (TEN domain) and nematodes (TEN and thumb motif)<sup>47,49,50</sup>. The reduced telomerase activity observed in some insect and nematode species might be explained by the lack of the TEN domain<sup>51-53</sup>.

However, TERT proteins have one unique feature not found in any other RT protein: the ability to translocate after successful DNA synthesis and realign for an additional cycle of DNA synthesis without primer dissociation (see section 1.5)<sup>54</sup>. Therefore the catalytic activity of the telomerase RNP complex is considered to be twofold: the ability to elongate a given DNA substrate by one or more nucleotides is termed nucleotide addition processivity (NAP), while initiation of a new round of

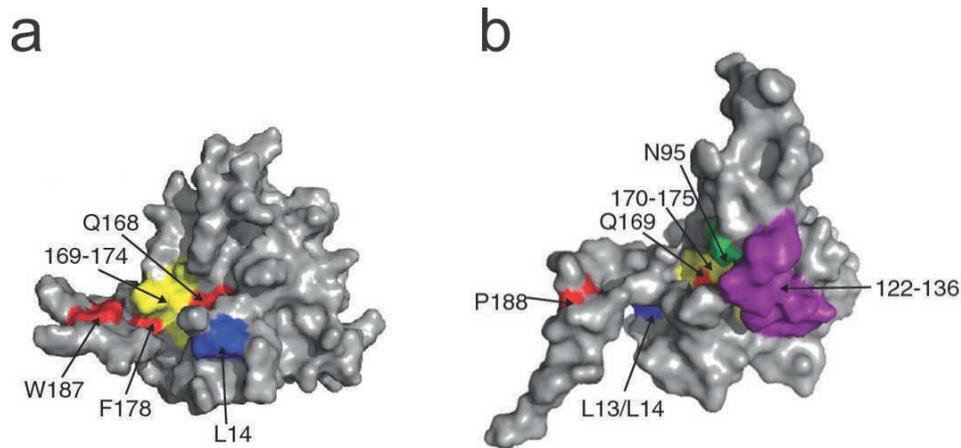
elongation after successful DNA synthesis is called repeat addition processivity (RAP). Within the telomerase RNP complex one domain might be crucial for either NAP or RAP, but can be dispensable for the other one<sup>28,55–58</sup>.



**Figure 1.10. Organization of TERT proteins from different species.** Cartoon representation of TERT proteins from vertebrates, ciliates, yeast, insects and nematodes. The TEN domain is depicted in yellow and the TRBD domain in blue. The RT domain is split in fingers (orange) and palm (crème) motifs. The CTE or thumb domain is shown in purple. The distinct motifs of each domain are labelled GQ, TFLY, VSR, CP, QFP, T, 1, 2, 3, A, IFD, B', C, D, E, loop and helix.

### 1.4.1. The TEN domain

Located at the N-terminus, the telomerase essential N-terminal domain is the first of the four functional domains of hTERT. It is connected via a flexible linker of species-dependent length to the TRBD domain and is the least conserved domain within known TERT proteins<sup>47,59,60</sup>. However, three residues have been found to be conserved among all available TERT sequences: in human these are G145, Q169 and G172<sup>61</sup>. In the *T. thermophila* TEN domain both glycines adopt specific torsion angles ( $\phi/\psi$ ) only possible for glycines, thereby inducing a sharp turn between  $\alpha 5$  and  $\alpha 6$  and  $\beta 6$  and  $\alpha 7$ <sup>61</sup>. Based on the *T. thermophila* crystal structure, homology modelling of the human TEN domain shows a comparable protein fold, in which G145, Q169 and G172 are superimposable on their counterparts in *T. thermophila*<sup>62</sup>. In contrast to the topology-related function of G145 and G172, Q169 was shown to be essential for catalysis<sup>62</sup>. Both in yeast (Q146 in Est2p) and *T. thermophila* (Q168 in tTERT) mutation of this residue leads to a drastically reduced interaction between protein and ssDNA, impaired function *in vitro* as well as growth deficiency and telomere loss<sup>61,63,64</sup>. In human TERT the mutation Q169 shows basically the same phenotype as in yeast and ciliate, however binding to ssDNA is only decreased when using a TERT fragment consisting only of amino acids 1-300<sup>65</sup>. This strongly suggests that additional DNA binding sites are present in the human protein, but still points to a conserved role of Q196 within the telomerase RNA complex<sup>65</sup>. The motif in which G145, Q169 and G172 are located, is called the GQ motif<sup>61,66</sup> and enables the TEN domain to bind DNA by forming a groove on the protein surface which is often referred to as “anchor site”<sup>62,67,68</sup> (figure 1.11).



**Figure 1.11. Structural organization of the TEN domain.** **a)** The *T. thermophila* TEN domain based on a crystal structure<sup>61</sup>. **b)** Homology model of the human TEN domain based on **a)**. Amino acids coloured in red in **a)** have been shown to cross-link to ssDNA primers, human homologues are also coloured in red<sup>61,63</sup>. Blue coloured amino acids have been shown to be important for RAP<sup>69</sup>. Residues depicted in yellow in **b)** are involved in ssDNA binding<sup>62</sup> and their homologues are also shown in yellow in panel **a)**. The purple patch denotes residues in hTERT involved in protein-protein interactions and TERT recruitment to the telomeres<sup>70</sup>. Residue N95 coloured in green was shown to be involved in DNA binding in human<sup>71</sup>. Figure adapted from Jurczyk et al., 2011<sup>62</sup>.

The anchor site of the TEN domain is responsible for RAP by binding ssDNA such as the 3' overhang of the telomeres<sup>72</sup>, however the exact mechanism is not yet known. Detailed studies on yeast (Est2p), ciliate (tTERT) and human (hTERT) TEN domain have revealed that besides the three absolutely conserved residues (G145, Q169 and G172 in human) there are species-specific residues important for telomerase function. While L14 is the key contributor to RAP in *T. thermophila*, human TERT needs a lysine either at position 13 or 14<sup>69</sup>. The current model of how the TEN domain contributes to repeat addition processivity (RAP) positions the TEN domain in close proximity to the catalytic center, at which it realigns the DNA-RNA duplex after DNA synthesis<sup>62</sup>.

In addition, the first 15 amino acids of the TEN domain have been reported to act as a nucleolar localization signal, directing TERT to the nucleolus and subsequently to the Cajal bodies<sup>73</sup>. Together with the CAB box present in hTR, both components are capable of reaching their destination for RNP assembly.

#### 1.4.2. Telomerase RNA binding domain

In contrast to the relatively poor sequence conservation of the TEN domain, the TRBD (telomerase RNA binding domain) sequence has a higher degree of conservation among all known TERT proteins<sup>74</sup>. Despite species-dependent differences in the motifs the core motifs CP, QFP and T are present in all TERT proteins<sup>74</sup>. Nomenclature with regards to the vertebrate-specific TFLY and VSR

motifs is not unified, as some studies count TFLY and VSR to the TRBD domain<sup>75</sup>, while others position them in the linker region connecting TEN and TRBD domain (figure 1.10)<sup>76</sup>. In this thesis TFLY and VSR motifs are considered to be part of the linker region, but are explained in this section due to their physical proximity to the TRBD domain.

As already implied by the name the TRBD's main characteristics are the extensive contacts to their TR counterparts. High quality structural information are available on the TRBDs of the beetle *T. castaneum*<sup>55</sup> (tcTRBD), the ciliate *T. thermophila*<sup>56</sup> (tTRBD) and the vertebrate *T. rubripes*<sup>77</sup> (trTRBD). All of three structures show a fold consisting mostly of helices and homology in their overall topology<sup>55,56,77</sup> (figure 1.12). The motifs CP and T form an extended pocket on the protein surface which was shown to be involved in binding of the template boundary element (TBE) in all three structures<sup>55,56,77</sup>. In addition, in vertebrate TRBD (trTRBD) the motifs TFYL and VSR which consists of helix  $\alpha 1$  and  $\alpha 4$ , respectively, extend the binding pocket built by CP and T<sup>77</sup>. Nonetheless, direct involvement of these motifs in binding to the TBE was not observed.



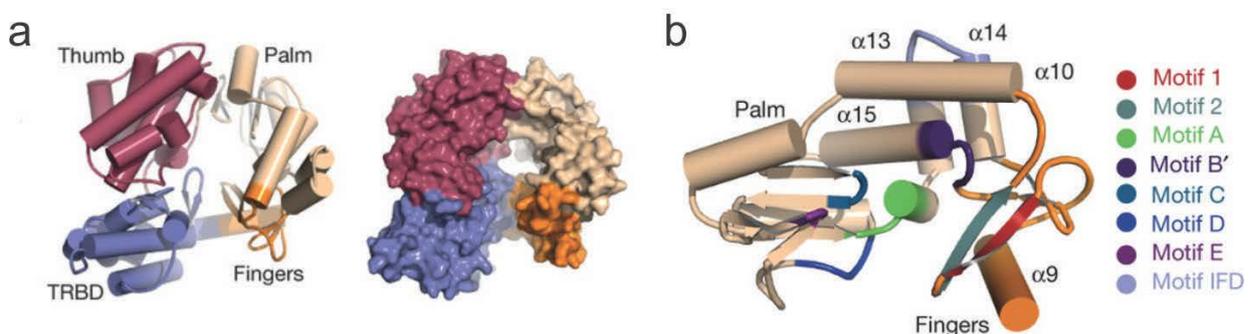
**Figure 1.12. Alignment of TRBDs from different species.** Aligned sequences from *O. latipes* (Ol\_Tert), *D. rerio* (Dr\_Tert), *H. sapiens* (Hs\_Tert), *M. musculus* (Mm\_Tert), *X. laevis* (Xl\_Tert), *T. castaneum* (Tc\_Tert), *T. thermophila* (Tt\_Tert) and *S. pombe* (Sp\_Tert) are shown. Motifs within the TRBD are boxed in orange, the topology is depicted with blue cylinders for helical parts or blue arrows for  $\beta$ -sheets, respectively. Figure adopted from Huang et al., 2014<sup>33</sup>.

Recently the high-resolution structure of the medaka mTRBD-CR4/CR5 complex was reported, revealing the first detailed insights into the TRBD RNA-binding interface<sup>33</sup>. While only three nucleotides (U182, G189 and A199 or U261, G268 and A301 in hTR) play a key role in telomerase RNA to contact the protein (see section 1.3.2), several amino acids are necessary for correct recognition and binding of medaka CR4/CR5 domain (figure 1.7): arginine 506 (glutamate 539 in hTERT) and leucine 371 (leucine 382 in hTERT) form hydrogen bonds with U182 and G189; stacking

interactions with two aromatic amino acids (tyrosine 503 and phenylalanine 376) help to stabilize those interactions, while A199 forms a hydrogen bond with phenylalanine 496 (valine 529 in hTERT)<sup>33</sup>. The interactions between U182 and tyrosine 503 have also been proposed on the basis of cross-linking experiments<sup>75</sup>. In addition, extensive contacts to the TR backbone of P6 are made by residues arginine 374, 490 and 495 as well as tryptophan 477, lysine 480, valine 481 and methionine 482 to the backbone of P6.1<sup>33</sup>. Interestingly, only phenylalanine 496 is part of a known motif (QFP), whereas all other residues are not part of any motif (Leu 371 and Phe 376 are part of helix  $\alpha 2$  and Tyr 503 as well as Arg 506 are located in helix  $\alpha 8$ )<sup>33</sup>. This suggests a mode of action for CR4/CR5 recognition that is based on the highly conserved structure of the corresponding telomerase RNA rather than on high sequence conservation in the TRBD of the corresponding TERT protein<sup>33</sup>.

### 1.4.3. The reverse transcription domain

The next functional domain in line is the reverse transcriptase (RT) domain. It consists of the seven conserved RT motifs (1, 2, A, B', C, D and E) as well as motifs only found in TERT proteins (motif 3 and IFD)<sup>48</sup>. Unlike the TRBD with its mainly  $\alpha$  helical fold, the RT domain consists of both  $\alpha$  helices and  $\beta$  sheets which are organized into two subdomains: fingers with motifs 1, 2 and the palm domain consisting of motifs 3, A, IFD, B', C, D and E<sup>55</sup>. Despite the lack of a TEN domain in *T. castaneum* TERT the overall organization is thought to represent the correct TERT conformation (figure 1.13)<sup>55,78</sup>. This is supported by several lines of evidence: **a)** the overall ring-shaped topology of the domains resembles closely those of evolutionary related viral RTs, such as HIV RT, viral RNA polymerases and B-family DNA polymerases<sup>55,79–81</sup>; **b)** extensive contacts between the N-terminal TRBD and the C-terminal thumb domain, which are mostly of hydrophobic nature, were observed<sup>55</sup>; **c)** the spatial arrangement of the domains form a pore which is about 21 Å in depth and about 26 Å wide<sup>55</sup>. These dimensions are almost identical to a DNA-RNA duplex expected in the vicinity of the active site.



**Figure 1.13. Structural organization of *Tribolium castaneum* TERT.** **a)** Cartoon and surface representation of *T. castaneum* TERT. **b)** Close up of the RT domain of *T. castaneum*. The motifs are coloured and their localization is shown on the cartoon representation of the RT domain. Domain colours are the same as in figure 10. Figure adapted from Gillis et al., 2008<sup>55</sup>.

A closer look on the organization of the RT domain reveals that the IFD motif (insertion in fingers) is located at the interface between palm and fingers (figure 1.13). It is primarily made up of helices  $\alpha 13$  and  $\alpha 14$  and orientated towards the outside of the inner hole, while making extensive contacts to helices  $\alpha 10$  and  $\alpha 15$ , thereby influencing significantly the local organization of those helices<sup>55</sup>. As both helices  $\alpha 10$  and  $\alpha 15$  are oriented towards the inner side of the central hole and K210 of  $\alpha 10$  is within coordination distance to a hypothetical RNA-DNA duplex, the IFD motif plays an important role for the organization of the RT domain<sup>78</sup>. As a consequence, point mutations diminishing the interactions between  $\alpha 14$  and  $\alpha 10$  are expected to severely disturb the organization, this was shown to be the case for yeast TERT (Est2p), in which substitutions of four amino acids to alanine (L589A/Y590A/I591A/D592A) led to decreased nucleotide addition processivity (NAP)<sup>55,82</sup>.

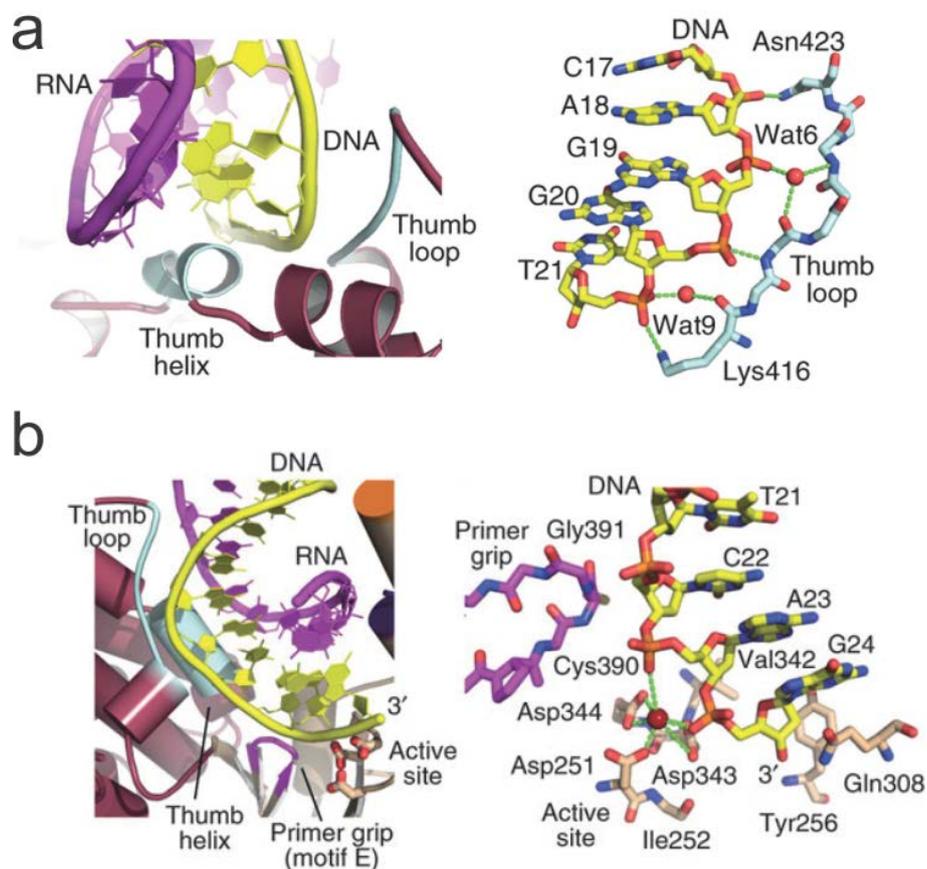
The catalytic centre of the protein consists of three highly conserved aspartates (D251, D343 and D344 in *T. castaneum* and D712, D868 and D869 in human, respectively) forming a catalytic triade responsible for nucleotide addition processivity<sup>48,55</sup>. D251 is part of the motif A, while D343 and D344 are both located in motif C<sup>48</sup>. The former is embedded in strand  $\beta 7$ , while the latter two are situated in the loop connecting strands  $\beta 10$  and  $\beta 11$ <sup>48,55</sup>. Like other reverse transcriptases, TERT most likely employs a two-metal ion mechanism for catalysis, relying on two magnesium ions for activating the nucleophile and stabilizing the leaving group<sup>48</sup>. Replacing the active-site aspartates to alanines was shown to impair telomerase activity<sup>48,83,84</sup>. Correct alignment of the incoming dNTP to the active site is achieved by conserved residues in motifs 1, 2, B' and D as well as motifs A and C<sup>55,78</sup>. Most of these motifs have already been characterized to play a role in template and/or nucleotide binding as well<sup>85,86</sup>. Taken together, the RT domain of TERT is mainly responsible for providing nucleotide addition processivity via an evolutionary conserved core which is extended by the IFD motif.

#### **1.4.4. The end of TERT – C-terminal extension**

The last of the functional domains of the TERT protein is the C-terminal extension (CTE) or also referred to as thumb. It can be further divided into two substructures, called thumb loop and thumb helix, respectively<sup>55</sup>. Despite sequence conservation between species, some nematodes completely lack the CTE<sup>50</sup>, which is interesting as both the TRBD and the CTE are part of the ring structure observed in the *T. castaneum* TERT<sup>55,78</sup>. As the presence of the CTE was shown to promote human telomerase activity, but is dispensable in yeast<sup>87</sup>, it is tempting to speculate that the role of the CTE was evolutionary solved in different ways (i.e. other proteins take over its role)<sup>47</sup>.

In the *T. castaneum* crystal structure of active TERT the CTE forms extensive interactions with the TRBD domain and is positioned in a remarkably similar way than the thumb domain of the HIV RT<sup>78</sup>. Its two substructures, thumb loop and thumb helix are predominantly formed by a helical fold<sup>55</sup>.

However, bio-informaticians could not identify a homologous fold, implying an entirely new conformation for the CTE<sup>55</sup>. The main role of the CTE seems to be the correct positioning of the DNA relative to the active site of the protein<sup>78</sup>. This is mostly done via extensive contacts of residues in the thumb and loop with the DNA (figure 1.14). Of particular importance is the thumb helix  $\alpha$ 15, which directly binds to the minor groove of the DNA-RNA duplex<sup>78</sup>, which is similar to the role of helix H found in viral RTs<sup>88,89</sup>. The thumb loop is oriented almost parallel to the axis of the DNA-RNA duplex, thereby allowing extensive contacts of both its sidechains (Lys 416 and Asn 423) and surrounding water-molecules with the phosphodiester backbone as well as the ribose moieties of the DNA<sup>78</sup>. The entirety of these interactions position the 3' DNA nucleotides within coordination distance to the 'primer grip' in motif E as well as in proximity to the active site<sup>78</sup>.



**Figure 1.14. Positioning of the DNA-RNA duplex at the active site of the TERT protein from *T. castaneum*.**  
**a)** Left panel: Overview of the thumb domain with its helix and loop substructures relative to the RNA-DNA duplex (purple and yellow, respectively). Right panel: Atomic resolution of the interactions between the thumb loop and the DNA backbone. Note that Lys 416, Asn 423 as well as two water molecules coordinate the exact position of the DNA backbone. **b)** Left panel: Overview of the active site of *T. castaneum* TERT with the thumb loop, thumb helix, the primer grip (motif E) and the RNA-DNA duplex (purple and yellow, respectively). Right panel: Motif E (primer grip) interacts with the DNA backbone adjacent to the active site of the TERT protein. The magnesium ion in the active site is shown as a red sphere. Figure has been adapted from Mitchel et al., 2010<sup>78</sup>.

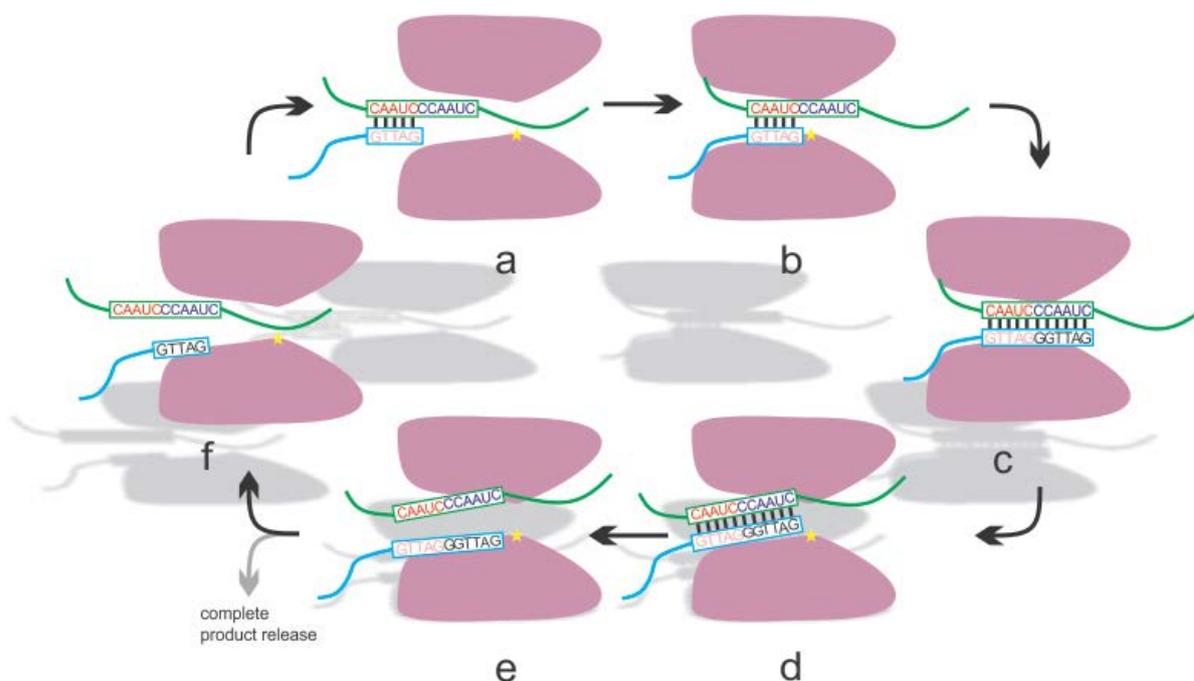
In summary, TERT is a unique reverse transcriptase with a conserved core, which points towards an evolutionary link with other RTs. It consists of four domains (TEN, TRBD, RT and CTE), which are not found in every species and therefore might play species-specific roles<sup>54,55,78</sup>. The solution of the crystal structure of TERT from *T. castaneum* both in absence and presence of the substrate DNA/RNA helix led to remarkable progress in understanding the overall structure and topology of TERT<sup>55,78</sup>.

### 1.5. Elongating the ends

Telomerase is a unique member within the family of reverse transcriptases. In contrast to all other RNA-dependent DNA polymerases telomerase releases single-stranded DNA products rather than double-stranded RNA-DNA duplexes<sup>90</sup>. This is necessary to allow telomerase to realign its RNA template and to proceed to the next round of DNA synthesis, a process which is commonly referred to as repeat addition processivity (RAP) and the distinctive hallmark of telomerase. While RAP is a unique feature of telomerases, it is not surprising that telomerases are also capable of nucleotide addition processivity (NAP) given the similarity of their RT domain with those of other prototypical RTs (e.g. HIV RT, viral RNA polymerases and B-family DNA polymerases<sup>55,79-81</sup>). The interplay of RAP and NAP needs careful orchestration, a fact which makes telomerase's mode of action much more complex and difficult to understand.

Elongation of telomeric DNA is a stepwise process and can be separated into the following steps (based on the assumption that the telomerase RNP complex has been properly assembled, localized and bound to the telomeres, see sections 1.6 and 1.7): **a)** initial binding of the 3' end of the DNA to the RNA template; **b)** correct positioning of the RNA-DNA duplex in the active site; **c)** extension of the DNA primer; **d)** release of the duplex from the active site; **e)** separation of the RNA from the DNA strand; **f)** repositioning of the template and **a)** with binding of the new 3' end of telomeric DNA to the RNA template a new round of repeat addition can occur (figure 1.15).

The processivity of telomerase raises some interesting questions. First, telomerase must exhibit an extremely stringent control on repeat length and needs to prevent read-through, as the incorporation of nucleotides outside the canonical template region would prevent successful re-alignment of the newly synthesized telomeric DNA with the canonical RNA template<sup>90</sup>. Second, after elongating the DNA primer the newly made RNA-DNA duplex must be separated again, a reaction which is thermodynamically unfavourable but takes place nonetheless<sup>90</sup>. Finally, repositioning of the template at the beginning of a new round of DNA synthesis somehow favours telomerase to bind the short RNA-DNA duplex rather than the old, fully synthesized repeat<sup>90</sup>.



**Figure 1.15. The telomerase catalytic cycle.** Telomerase employs a variety of steps in order to achieve both nucleotide addition processivity and repeat addition processivity. **a)** first binding of template RNA (green) to the 3' end of the telomeric DNA (blue) occurs, **b)** followed by correct positioning of the DNA/RNA duplex at the active site (yellow star). **c)** The DNA primer is elongated for six nucleotides and then the elongation comes to a stop. **d)** Afterwards the product is released from the active site, and **e)** strand separation takes place in order to prepare the RNP complex for a new round of DNA synthesis or the telomeric DNA dissociates and is released (gray arrow). **f)** The template is repositioned and with the initial base pairing of the new 3' end of telomeric DNA with the RNA template the cycle starts again. Note that the alignment region of the template is shown in red, whereas the actual template is shown in dark blue. Newly synthesized DNA nucleotides are shown in black. The TERT enzyme is shown in violet.

So far, most of the studies on telomerase focused on understanding the regulation of the repeat length and on defining the template. A study by Brown et al. in 2014 showed convincingly that indeed the template itself plays an important role in repeat addition processivity<sup>91</sup>. In fact, a single nucleotide (U49) in the human template is responsible for sensing when a full repeat has been added to the telomeric DNA<sup>91</sup>. The first synthesized A base pairs with U49 in hTR, allowing telomerase to precisely pause after adding three additional nucleotides and reaching the end of one canonical repeat<sup>91</sup>. The sequence-specific pause site permits the release of the RNA-DNA duplex from the active site (figure 1.15, step d) and seem to work synergistically with the P1b template boundary element (TBE)<sup>91</sup>.

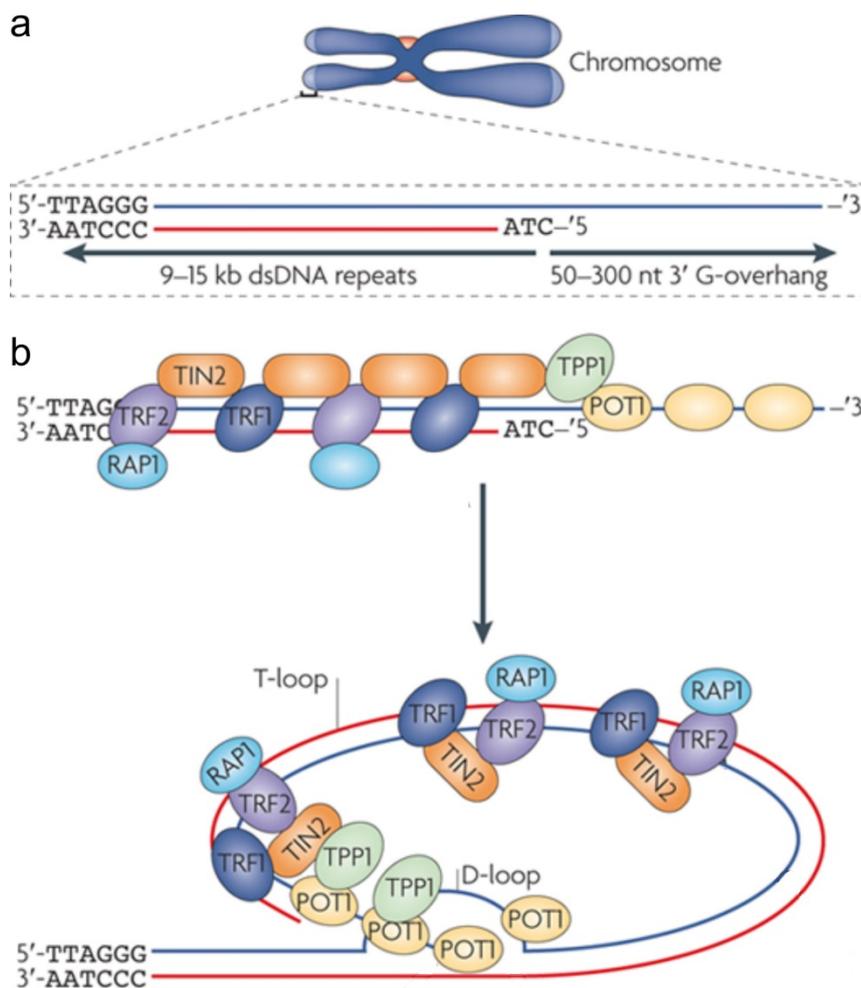
In addition, numerous other factors which influence telomerase activity have been identified. For example, the concentration of primer and dNTPs, sequence of the primer, temperature and

regulatory proteins<sup>92-96</sup>. Only a few of all the contributors to telomerase processivity are well understood, while others remain to be characterised or to be discovered.

### **1.6. The shelterin complex is guarding the telomeres**

The ends of a chromosome are called telomeres and consist usually of several kilobases of the corresponding species-dependent telomeric repeat sequence. For instance in human the repeat sequence is 5' GGTTAG 3'<sup>97</sup>. A typical human telomere can vary between 10 and 15 kb in length<sup>98-100</sup>, whereas a mouse telomere is profoundly longer with 25 to 50 kb<sup>101</sup>. However, the reason for this difference is not yet known. On the other hand, the minimum length of a telomere in human was shown to be around 1 kb, as shorter telomeres induce senescence<sup>102</sup>. In contrast to double-strand breaks, telomeres are protected from mechanisms that are normally triggered upon detection of double-strand breaks, like DNA-damage response, homologous recombination or non-homologous end joining (NHEJ)<sup>103</sup>. Capper and Baird showed that a stretch of 13 telomeric repeats corresponding to 78 base pairs is the minimal length to prevent telomere-telomere fusion<sup>104,105</sup>.

The very end of a mammalian telomere consists of a G-rich 3' overhang instead of being blunt-ended<sup>106,107</sup>, a feature conserved in eukaryotes. In mammals the 3' overhang varies between 50 and 300 nucleotides (figure 1.16), being among the longest in the eukaryotic kingdom<sup>97</sup>. This is different from the telomeres of lower eukaryotes which often have shorter 3' overhangs, e.g. ciliated protozoans, like *Tetrahymena thermophila*, only have 12 to 16 nucleotides at the 3' strand of their chromosomes<sup>108,109</sup>. The exact mechanism for generating the vertebrate long 3' tail is not known, but telomerase has been ruled out as a potential candidate<sup>110,111</sup>. While the last base of the G-strand can consist of any of the four canonical dNTPs, the 5' end of the C-strand consists exclusively of 3'-ATC-5'<sup>112</sup>, suggesting that the 3' overhang is generated by a specific cleavage event from a nuclease<sup>106</sup> (figure 1.16). The ends of telomeres form an architecturally interesting structure, which is thought to protect and hide the telomere terminus from the DNA damage repair machinery<sup>113</sup>. The so-called T-loop is formed as the 3' overhang of the G strand intrudes the double-stranded region and forming base pairs with the complementary C-strand, while the original G-strand in that region is forced to form a D-loop (displacement loop)<sup>113</sup>. The size of the T-loops varies considerably between different organisms and does not seem to be related to the function of telomere protection<sup>97</sup>. Furthermore, it is not clear how the formation of T-loops is triggered and works nor is it known if the T-loops represent structures, which are present throughout the entire cell cycle or only present in certain stages of the cell cycle<sup>97</sup>. T-loop formations are energetically unfavourable and it is therefore not surprising that various proteins called the shelterin complex help to stabilize them.



**Figure 1.16. Schematic representation of human telomere organization.** a) Human telomeres consist of several kilobases of canonical double-stranded 5' GGTTAG 3' repeats ending with a 3' overhang. The strands are referred to as G-strand (containing the canonical repeat sequence) and its complementary C-strand, respectively. b) Telomeres are covered by the proteins of the shelterin complex (TPP1, POT1, TIN2, TRF1, TRF2 and RAP1). To protect the chromosome end the 3' overhang folds back and invades the dsDNA, displacing the G strand. This results in formation of a T-loop and a D-loop (displacement loop). Figure adapted from O'Sullivan et al., 2010<sup>114</sup>.

Other common features found at the ends of telomeres are G-quadruplex structures of G quartets which form planar, square configurations of four Gs via Hoogsteen base pairing<sup>115</sup>. Although there is evidence that G quadruplexes form spontaneously *in vitro* and are as well present *in vivo*, all known shelterin proteins bind unfolded, single-stranded DNA, but have the ability to catalyse both folding and unwinding of G quadruplexes<sup>116–118</sup>.

The shelterin complex consists of six proteins (TPP1, POT1, TIN2, TRF1, TRF2 and RAP1) and numerous shelterin complexes cover the telomeres<sup>97</sup>. Telomeric repeat binding factor 1 and 2 (TRF1 and TRF2) bind to the double-stranded repeats, whereas protection of telomeres (POT1) localizes to the single-stranded repeats in the 3' overhang and the D-loop<sup>97</sup>. TIN2 binds to TRF1 and TRF2 as well

as to TPP1, which itself is bound to POT1, whereas RAP1 interacts with TRF2<sup>97</sup> (figure 1.16). The shelterin complex was shown to be stable even when no telomeric DNA is present<sup>119</sup>. In general, covering of telomeric DNA with proteins, such as the shelterin complex, was also observed in other organism and some shelterin proteins, like POT1 and TRF, have conserved orthologues in yeast as well as in ciliates<sup>97</sup>.

It is also noteworthy that some species have found alternative ways despite telomerase to protect their telomeres. For instance, the silkworm *B. mori* still has the canonical insect repeats (TTAGG), but in addition its telomeres contain also insertions of non-LTR retrotransposons, which is of particular interest, as the telomerase enzyme of *B. mori* shows little or almost no processivity<sup>120-122</sup>. It is believed that in *B. mori* the retrotransposons are exclusively responsible for the maintenance of the telomeres<sup>122,123</sup>. Therefore, *B. mori* might represent an organism in transition between two telomeric maintenance systems.

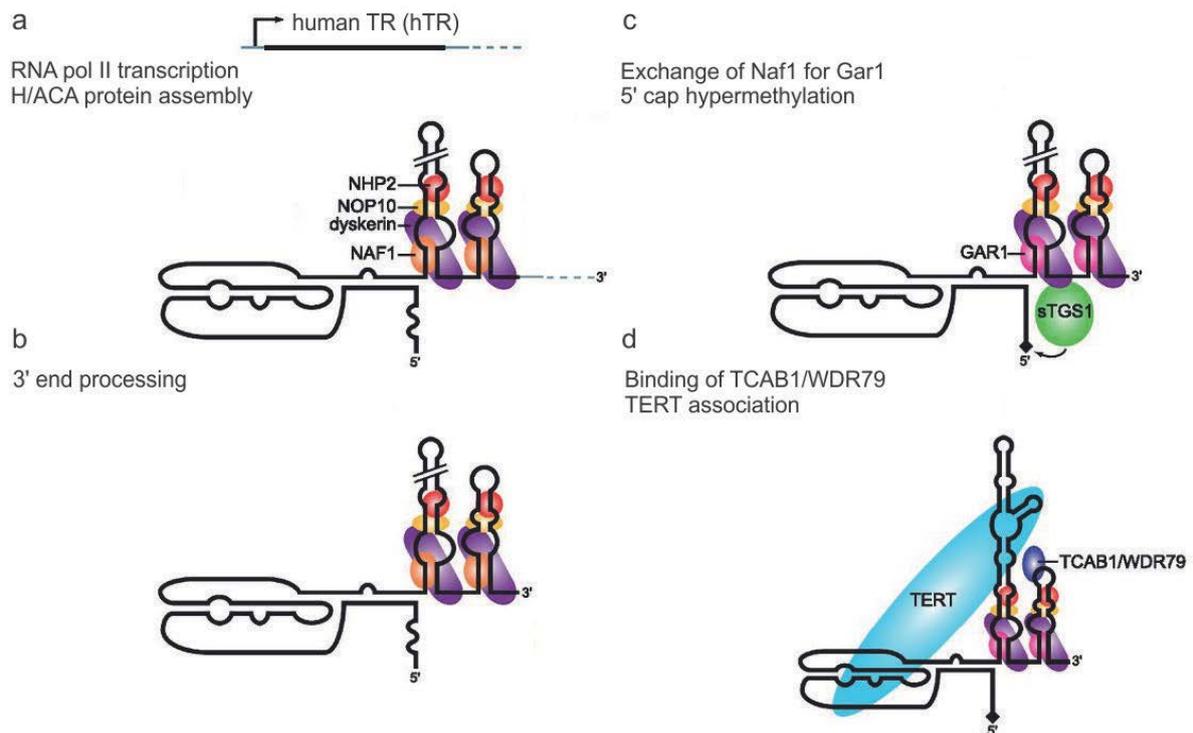
An entire different situation is found in one of the main model organisms in biology: the fruit fly *D. melanogaster*. Here, genes coding for TERT or the integral RNA component TR were not found<sup>124</sup>. Instead, three different classes of non-LTR retrotransposons have been identified (HeT-A, TAHRE and TART), which are exclusively found at the chromosome ends and form a terminal retrotransposon array<sup>124</sup>. These retrotransposons have actually taken over the role of telomerase RNP and prevent telomere erosion. Both examples, *B. mori* and *D. melanogaster*, illustrate nicely how evolution has come up with different, independent solutions for the same problem.

## 1.7. Biogenesis and maturation of hTR

Obtaining a functional telomerase RNP complex *in vivo* is a complex process (figure 1.17). The RNA component TR and the protein TERT undergo completely different steps of maturation and processing, before being assembled into a functional RNP complex. *In vitro* the situation is quite simple with only TR and TERT being necessary to obtain telomerase activity<sup>11,12</sup>, while *in vivo* additional protein co-factors are needed for correct processing of the telomerase RNA, the assembly of the RNP complex and trafficking of the mature complex to the telomeres<sup>19</sup>.

As already mentioned earlier, human telomerase RNA is transcribed as a precursor by RNA polymerase II<sup>16</sup>. The 3' end of hTR is processed in a similar yet not fully understood way to canonical H/ACA snoRNAs to yield the mature 451 nucleotide product<sup>14,15</sup>. The structural similarity of hTR to other H/ACA snoRNAs points to a shared mechanism for RNP assembly between hTR and H/ACA snoRNAs. Usually, H/ACA snoRNAs act as a guide to target other RNAs for pseudouridylation, such as ribosomal RNAs or snRNAs<sup>125,126</sup>. Indeed, human telomerase RNA relies on the same pre-assembled protein scaffold as canonical H/ACA snoRNAs consisting of four proteins, the pseudouridylase

Dyskerin, Nhp2, Nop10 and the H/ACA RNP assembly factor Naf1<sup>19</sup>. The pre-assembled protein complex is loaded co-transcriptionally onto hTR, whereby one set of proteins binds to each hairpin<sup>127–129</sup>. In contrast to H/ACA snoRNAs which require a total of two protein sets, one bound by each hairpin, only one set of proteins bound to the 3' hairpin is necessary for hTR<sup>40</sup>. The BIO box located in the terminal loop (L8b) of the hairpin was shown to promote H/ACA RNP assembly and therefore shifts the balance towards successful RNP assembly and away from degradation of the precursor by exonucleolytic cleavage<sup>40</sup>.



**Figure 1.17. Biogenesis and maturation of the human telomerase RNP complex.** Schematic overview of the different steps in telomerase maturation. **a**) hTR is transcribed by RNA polymerase II and loaded with the H/ACA protein complex consisting of Dyskerin (purple), Nhp2 (red), Nop10 (yellow) and Naf1 (orange). **b**) The 3' end of hTR is processed to obtain the mature product of 451 nucleotides. **c**) The entire complex is routed to the Cajal bodies, Naf1 (orange) is exchanged for Gar1 (pink) and the TMG cap is added by sTgs1 (green). **d**) Interaction with Tcab1 (dark blue) prolongs Cajal-body localization and during S-phase hTERT (light blue) binds to complete the mature telomerase RNP complex. Figure adapted from Egan and Collins, 2012<sup>19</sup>.

After correct assembly of hTR with Dyskerin, Nhp2, Nop10 and Naf1 the entire RNP complex is transported via Phax and Nop140 to the Cajal bodies, a site of RNP remodelling<sup>130,131</sup>. In this respect Naf1 is replaced with the mature component Gar1<sup>128</sup>. Finally, direct interaction between the CAB box located in the terminal loop (L8b) and the protein Tcab1 ensures continuous localization of the RNP complex at the Cajal bodies<sup>132,133</sup>.

In order to obtain a functional telomerase the protein component TERT must associate with the complex. While hTR is mostly processed in the Cajal bodies, immunofluorescence studies identified hTERT to be located in different compartments than hTR<sup>134</sup>, such as nucleoplasmic foci. The data suggests that human telomerase RNA and hTERT are separated for most of the cell cycle and only co-localize throughout S-phase, during which co-localization of hTR and hTERT was observed at the periphery of the Cajal bodies as well as at telomeres<sup>134</sup>. Strikingly, not all telomeres of the cell were covered with telomerase, this was only the case for a small subset, suggesting that only a fraction of the total amount of cellular telomeres is elongated during one cell division<sup>134</sup>. In addition, hTERT is likely to bind to the mature hTR RNP complex and not during one of the previous steps.

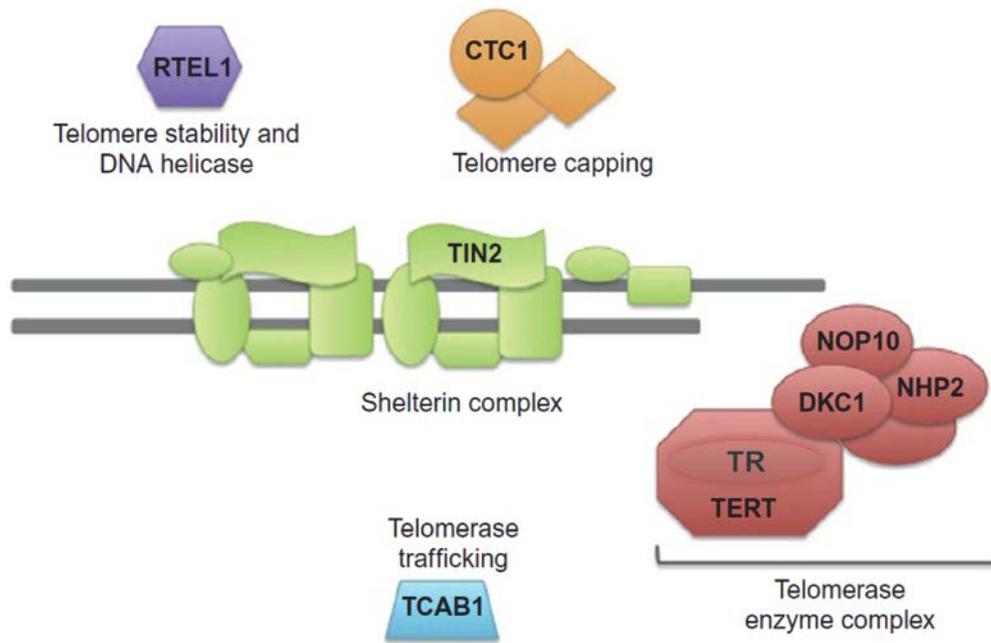
Like in all natural processes, a failure in one or more steps of telomerase RNP processing and assembly might have drastic consequences. With telomerase being highly active in proliferative tissues, decreased amounts or defective telomerase have been suggested to be the cause of a number of diseases<sup>135</sup>.

## **1.8. Telomerase and influence on disease**

Since the discovery of telomerase, mutations either in the hTR gene, the hTERT gene or genes coding for shelterin components or H/ACA RNA-binding proteins have been identified as the primary reason for different genetic diseases. It is now commonly accepted to refer to the entirety of telomere-associated diseases as telomere biology disorders (TBD)<sup>10</sup>.

### **1.8.1. Telomere-associated diseases – Dyskeratosis congenita**

The first clinical description of syndromes of Dyskeratosis congenita (DC) are dating back to the 1900s<sup>136</sup>. The patients suffer from a multi-system disorder with three main diagnostic patterns, including nail dystrophy, abnormal skin pigmentation around the chest and neck and oral leukoplakia<sup>10</sup>. Several associated disorders like bone marrow failure, stem cell defects or a higher susceptibility for cancer are also frequently observed<sup>137</sup>. The unifying characteristics of all DC patients are their significantly shorter telomeres compared to healthy individuals of the same age<sup>138</sup>. The identification of mutations in the gene *DKC1* coding for Dyskerin as a major source for this inherited disease provides the connection between the observed clinical syndromes and the fact that telomere length is drastically reduced<sup>10</sup>. This hypothesis is strengthened by findings that the cellular levels of pseudouridylation of rRNAs, the main task of Dyskerin, are not altered and, indeed, telomerase dysfunction is the primary reason for Dyskeratosis congenita<sup>139</sup>. In contrast, a knock-out of Dyskerin proved embryogenic lethal in mice<sup>140</sup> as well as in yeast<sup>141</sup>, but this is due to failure of pseudouridylation of rRNAs<sup>141</sup>.



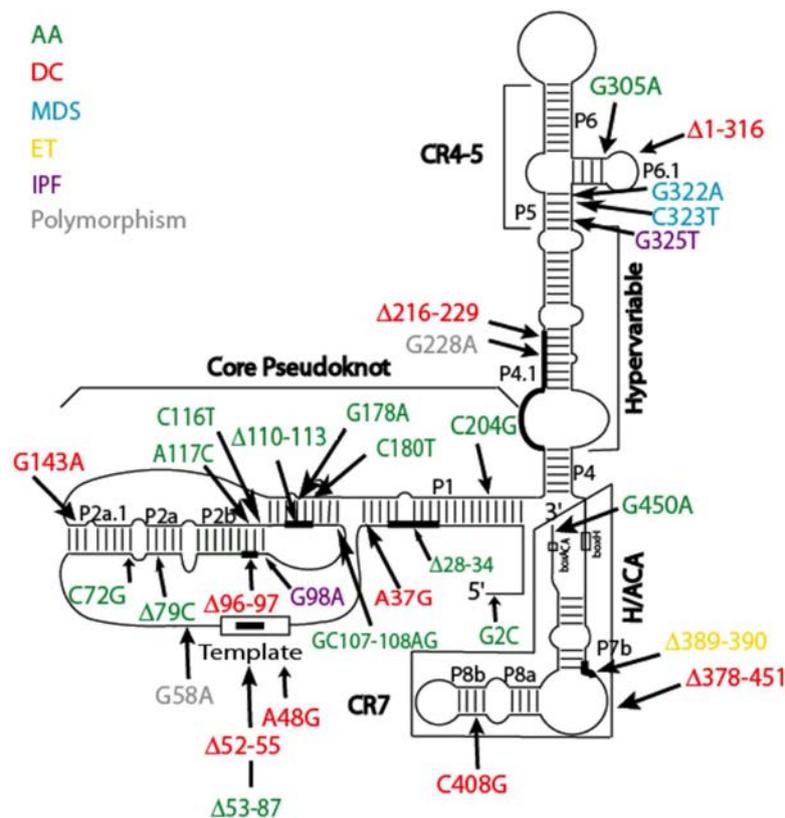
**Figure 1.18. Mutations associated with Dyskeratosis congenita.** Mutations in the proteins Rtel1, Ctc1, Tin2 and Tcab1 are associated with DC and are involved in either telomere stability and capping, the shelterin complex or telomerase trafficking. Proteins that are part of the telomerase RNP complex were also found to contain mutations in DC patients: TERT, Dyskerin, NOP10 and NHP2. The telomerase RNA was also found to contain mutations which are associated with DC. This figure was adapted from Savage et al., 2014<sup>10</sup>.

Nowadays, additional mutations in several genes coding for various telomerase-associated proteins or the telomerase RNA have been identified to cause DC (figure 1.18). Thus, mutations were only identified either in the shelterin complex (Tin2), proteins responsible for telomerase stability (Rtel1) and telomere capping (Ctc1), telomerase localization (Tcab1) or in components of the telomerase RNP complex (Nop10, Nhp2, dyskerin, TERT and hTR)<sup>10</sup>. Interestingly, the important role of telomerase RNA and its structure is reflected by the fact that several mutations in the pseudoknot domain, in the CR4/CR5 domain as well as in the scaRNA domain are associated with DC and other pre-mature aging syndromes<sup>24</sup> (figure 1.19). Most of them are thought to alter the secondary and/or tertiary structure (e.g. the triple helix in the pseudoknot domain), thereby disrupting the functional telomerase RNP complex<sup>142</sup>.

### 1.8.2. Telomerase and cancer

Telomerase expression is tightly regulated throughout the different cells in the human body. In early embryogenesis telomerase is highly expressed and active, but is down-regulated afterwards in most somatic tissues<sup>143</sup>. In contrast, highly proliferative tissues containing stem-cell-alike populations, such as the hematopoietic system in the bone marrow, skin and intestine still express telomerase<sup>144-146</sup>, indicating that telomerase expression and the proliferation capacity of certain tissues are closely

connected. This link is therefore demonstrated by the fact that almost 90% of human cancer types have been identified as telomerase-positive<sup>147</sup>. What could be the reason for normal cells to transform into malignant ones? Normal cells lacking telomerase expression sooner or later reach their Hayflick limit and enter senescence (also called mortality state 1 or M1)<sup>148</sup>. If the checkpoints for cell cycle arrest are absent, the cell continues to divide and enters crisis (referred to as M2 state) due to telomere fusion-bridge-breakage cycles followed almost universally by apoptosis<sup>148</sup>. Both senescence and crisis (M1 and M2) are very potent anti-cancer mechanics protecting the human body<sup>148</sup>. However, as a consequence of a rare mutation event allowing a cell to elongate its telomeres (e.g. by expression of telomerase) and thus starting to proliferate again, a potentially cancer-causing cell developed. Recent studies have identified several mutations in the promoter region of hTERT in different cancer types<sup>149</sup>. Strikingly, in melanoma (skin cancer) mutations in the hTERT promoter were found in 89% of the cases<sup>150,151</sup>. Specifically, cytidine to thymidine transitions in a consensus sequence that is required for binding of a transcription factor results in a two to four-fold increase in hTERT expression<sup>150,151</sup>.



**Figure 1.19. Disease-related mutations sites within hTR.** Summary map showing naturally occurring mutations within human telomerase RNA. Each change in the sequence is colour-coded based on the disease it is associated with: AA (aplastic anemia), DC (Dyskeratosis congenita), MDS (myelodysplastic syndromes), ET (essential thrombocythemia) and IPF (idiopathic pulmonary fibrosis). Figure adapted from Carroll et al., 2009<sup>142</sup>.

In comparison to the enormous amount of cells in a human being (approximately  $3.7 * 10^{13}$  cells<sup>152</sup>) and the large number of cells dividing, differentiating and dying every day (e.g. in the gastrointestinal tract or the bone marrow), the protection against cancer works surprisingly well. Replicative aging (senescence) due to telomere shortening, apoptosis for cells reaching crisis and the absence of active telomerase represent powerful tools to prevent the formation of cancer cells most of the time. With telomerase being active in most cancer types, new therapeutic studies focus on inhibiting telomerase activity in cancer cells<sup>153-156</sup>.

## 1.9. Investigating the structure of biological macromolecules

In order to shed light on such a complex structure, like the human telomerase RNP complex, a variety of different techniques have been developed and applied. Many biological macromolecules can be crystallized, resulting in a repeating arrangement of the molecule. Upon exposure to a beam of X-rays this grid-like structure produces a diffraction pattern which contains information about the crystal's internal structure and hence the molecules the crystal was made of. The first biological macromolecule, for which X-ray crystallography was used successfully to solve the structure, was myoglobin in 1958<sup>157</sup>. The importance of this new method is also reflected by the fact that the Nobel Prize in Chemistry in 1962 was jointly awarded to John Kendrew and Max Perutz for "*for their studies of the structures of globular proteins*"<sup>158</sup>. Nowadays X-ray crystallography has developed into a key method for solving complex biological macromolecules. The method has no size limitation thus as long as a macromolecule is crystallisable, X-ray crystallography can be applied to study its structure.

On the other hand, crystals are static and therefore the molecule can be only investigated in one specific, frozen conformation. For instance, biological macromolecules often have regions which are flexible and might not adopt a single conformation (e.g. the flexible tails in the murine prion protein<sup>159</sup>) and therefore are hard to investigate in a static crystal structure. The dynamics within a protein, for instance, can be studied much better when using solution nuclear magnetic resonance spectroscopy. The fact that an atomic nucleus, consisting of an odd number of protons and/or neutrons, interacts with a magnetic field by absorbing and re-emitting electromagnetic radiation was already discovered by Edward Mills Purcell and Felix Bloch in the 1940s<sup>160,161</sup>. However, it took 40 years, before NMR was successfully used to solve a protein structure in 1984 by the group of Kurt Wuethrich<sup>162</sup>. Both, X-ray crystallography and NRM have developed into key methods for assessing biological macromolecules and are nowadays used on a routine basis.

Despite being powerful tools, both X-ray crystallography and solution NMR come with limitations. While crystallography is utterly dependent on the molecule of choice forming a useable crystal, NMR has considerable size limitations (<30 kDa) allowing only the analysis of small molecules or isolated

domains from a larger molecule. Furthermore, both methods rely on *in vitro* samples and might therefore not accurately represent the situation *in vivo*. This can be solved by using chemical compounds which modify the RNA or protein structure and can easily penetrate cell walls in order to be used *in vivo*. These approaches are commonly known as chemical probing and can often be applied *in vitro* as well as *in vivo*. One of many advantages of chemical probing is the fact that there is no limitation in size, therefore allowing the probing of complex macromolecules. In our lab many different approaches for investigating RNP complexes are used (e.g. UV cross-linking, Pb<sup>2+</sup> cleavage, chemical probing as well as SHAPE protocols) allowing us to thoroughly analyse a RNA molecule of choice. The combination of different approaches allows collecting a plenitude of information on base pairing, long-range tertiary interactions and ligand-induced structural changes<sup>163-165</sup>. Currently, improved protocols for DMS probing, SHAPE as well as UV cross-linking allow for parallel analysis of many RNAs at once through subsequent deep sequencing<sup>166-168</sup>. In addition, UV cross-linking coupled with mass spectroscopy can be used to identify RNA-protein interactions, but the low yield often poses a considerable problem for subsequent mass spectroscopy.

By incorporating data from chemical probing and phylogenetic information on the one hand and X-ray crystallography as well as NMR data on the other hand, the structure of a given macromolecule can often be much more refined and improved. An example for such a combined approach is the CASP experiment for protein folding<sup>169</sup>. Although originally intended to improve structure prediction algorithms by comparing their results to crystal structures, the computational predictions can also improve areas in the crystal with only poor electron density<sup>170</sup>. By using a combined approach of X-ray crystallography, NMR, chemical probing and computational predictions, the structural organization of many molecules have been analysed nowadays.

## 2. Aims of the work

Organic life is dependent on chemical reactions carried out by a multitude of different proteins and RNAs. In order to carry out these functions, each enzyme must adopt a precisely defined structure, leading to the conclusion that understanding living cells necessitates understanding the structure of its building blocks. The telomerase RNP complex is known for 25 years and the more structural details emerge, the more advanced is our understanding of this fascinating machinery.

In the Waldsich lab we are interested in elucidating RNA structure and RNA folding. Telomerase consists of approximately 55% RNA, which is essential for telomerase function. To obtain a functional complex hTR and hTERT must assemble into an active telomerase complex and by doing so an extensive network of either RNA-RNA and/or RNA-protein interactions must form. However, only limited information about this interaction network is available and most of these studies have been carried out *in vitro*. It is therefore necessary to develop a method to investigate RNA-RNA (and RNA-protein) interactions in the telomerase complex *in vivo* with single-nucleotide resolution to shed more light onto its structural organization.

By employing *in vivo* UV cross-linking, I first aimed to characterise the sites within hTR that are spatially close to other hTR elements or hTERT, thus resulting in a cross-link. Second, it was of interest to determine whether the cross-linking efficiency depends on the presence of hTERT. To identify potential RNA-RNA contacts or RNA residues close to hTERT, a genetic approach was combined with UV cross-linking.

By successfully solving these challenges I want to find answers to the following questions: **a)** What regions of hTR are involved in RNA-RNA contacts?, **b)** How do these sites change in the absence of the protein hTERT? and **c)** Where is a potential binding site within hTR for hTERT?

### 3. Materials and Methods

#### 3.1. Bacterial strains

Cloning and transformation were performed with *E. coli* strain XL1-blue (Stratagene). The genotype is endA1 gyrA96(nal<sup>R</sup>) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB<sup>+</sup> lacI<sup>q</sup> Δ(lacZ)M15] hsdR17(r<sub>K</sub><sup>-</sup> m<sub>K</sub><sup>+</sup>).

#### 3.2. PCR-based cloning strategies

Mutations in hTR were introduced to the plasmid pBS U1 hTR by Fastcloning<sup>171</sup>. Summarized shortly this method relies on the *in vivo* methylation of bacterial plasmids obtained from *E. coli*. In contrast to bacteria PCR products are not methylated as they are produced *in vitro*. Using the methylation-sensitive restriction enzyme DpnI the template of bacterial origin can be degraded and only the PCR products containing the desired mutations are transformed into the appropriate *E. coli* strain.

Affinity tags for hTERT were introduced with conventional overlap PCR. In short two PCR reactions were performed using primers F1 and R1 and F2 and R2. F1 is binding upstream of the insertion, while R1 contains the tag to be inserted and is covering the insertion site. The second primer F2 is partially overlapping and complementary to R1 and contains the affinity tag. Primer R2 is binding downstream of the insertion site. The final product is generated by a third PCR reaction using the F1/R1 and F2/R2 PCR reactions as a template. Primer F1 and R2 are added after 5 cycles and guarantee the synthesis of the full-length product. Table 3-1 shows the details of the PCR protocol, table 3-2 shows the reaction setup for a single PCR reaction.

Step	Conventional and Overlap PCR		Fastcloning	
	[°C]	[sec]	[°C]	[sec]
Denaturation [°C]	94	300	98	180
# of cycles	25		25	
Denaturation [°C]	94	30	95	10
Annealing [°C]	Dependent on primer T <sub>M</sub>		55	30
Elongation [°C]	72	Enzyme dependent	72	180
Final elongation [°C]	72	300	72	300
Storage [°C]	4	∞	4	∞

Table 3-1. Summary of PCR protocols used for molecular cloning.

Component	Conventional PCR	Overlap PCR	Fastcloning
10x buffer	5 µl	10 µl	5 µl
10mM dNTPs	1 µl	2 µl	1.25 µl
Template DNA [10ng/µl]	5 µl	5 µl	1 µl
10µM forward primer	1 µl	1 µl	0.5 µl
10µM reverse primer	1 µl	1 µl	0.5 µl
10mM MgCl <sub>2</sub>	5 µl	-	-
10mM MgSO <sub>4</sub>	-	15 µl	-
14.1M DMSO	2.5 µl	5 µl	-
GoTaq polymerase [5u/µl]	0.25 µl	-	-
Pfu polymerase [3u/µl]	-	0.8 µl	-
Pfu Ultra II polymerase	-	-	1 µl
ddH <sub>2</sub> O	29.25 µl	60.2 µl	40.75 µl
Total volume	50 µl	100 µl	50 µl

**Table 3-2. Pipetting scheme for conventional, overlap and Fastcloning PCR.**

All plasmids generated for this thesis are listed below (table 3-3), while the primers used are listed in the addendum (table iii-1).

Plasmid	Mutation
pCW135	U39C/U40C/U41C/U42C/U43C
pCW137	U86C/C87U/C88U/C89U
pCW138	U100C/U101C/U102C/U103C/C104U/U105C
pCW139	C112U/U113C/U114C/U115C
pCW117	U146A/U147A
pCW163	C152G/G153U/U154A/U155A/C156G/A157U/U158A/U159A/C160G
pCW187	A176G/U177C/C180U/U184C
pCW189	U249C
pCW190	C290U/U291C
pCW191	U307C
pCW192	C313U/U314C/U316C/C317U
pCW145	U350C
pCW146	U357C/U358C/U359C/C360U
pCW147	G404A/A405G/U406C/U407C/C408U
pCW148	U416C/G417A/U418C

**Table 3-3. Mutated hTR constructs used for mapping RNA-RNA interaction sites.**

### 3.3. Agarose gels and restriction digests

PCR products were analysed on a 0.8% native agarose gel for efficiency and specificity. The respective bands were cut out under UV-light and processed with the Wizard© SV Gel and PCR Clean-Up System (Promega). Restriction digests were performed with enzymes purchased from New

England Biolabs. Total volume of each restriction digest was either 10 µl for control digests or 200 µl for preparative digests. 1 unit of enzyme was used to digest 1 µg of DNA. Buffer and BSA were added according to the manufacturer's protocol. For enzymes for which no heat inactivation was possible the Wizard© SV Gel and PCR Clean-Up System was used for removal of the enzyme.

### 3.4. Competent cells and bacterial transformation

Chemically competent XL1-Blue *E. coli* were used for transformation. Cells were prepared according to the following protocol:

1. Inoculate 1 ml of overnight culture in 100 ml LB medium at 37 °C and 180 rpm.
2. Grow the culture to OD<sub>600</sub> 0.5 – 0.6.
3. Once the OD is reached, cool the flask on ice for 20 minutes and harvest cells by centrifugation at 4 °C and 4000 rpm for 10 minutes.
4. Resuspend the cell pellet in 5 ml of ice-cold KCM buffer and incubate on ice for 15 minutes.
5. Freeze 100 µl aliquots in liquid nitrogen and store at – 80 °C.

Transformation is carried out as follows:

1. Mix 20 µl of 5x KCM, 1 µl of DNA solution [10ng/µl] and 79 µl ddH<sub>2</sub>O with 100 µl of thawed KCM cells.
2. Incubate on ice for 20 minutes.
3. Perform heat shock at 37 °C for 5 minutes.
4. Add 200 µl of pre-warmed LB medium and incubate at 37 °C for another 60 minutes (with moderate shaking).
5. Plate transformation mix on LB plates containing appropriate antibiotics for selection.

#### KCM buffer

LB medium	75 ml
PEG 3350 [w/v]	10 g
DMSO	5 ml
1M MgSO <sub>4</sub> [f.c. 100mM]	10 ml
1M MgCl <sub>2</sub> [f.c. 100mM]	10 ml

Add ddH<sub>2</sub>O up to 100 ml, sterile filter the solution.

#### 5x KCM

2.5M KCl [f.c. 500mM]	6 ml
1M CaCl <sub>2</sub> [f.c. 150mM]	4.5 ml
1M MgCl <sub>2</sub> [f.c. 250mM]	7.5 ml

Add ddH<sub>2</sub>O to 30 ml, store at 4 °C

#### LB medium

Tryptone [f.c. 1%]	10 g
Yeast extract [f.c. 0.5%]	5 g
NaCl [f.c. 1%]	10 g

Add ddH<sub>2</sub>O to 1 l and autoclave, store at RT.

#### LB plates

Tryptone [f.c. 1%]	10 g
Yeast extract [f.c. 0.5%]	5 g
NaCl [f.c. 1%]	10 g
Ampicillin [f.c. 100µg/ml]	1 ml

Add ddH<sub>2</sub>O to 1 l and autoclave. Add ampicillin when solution has cooled down to 50 °C shortly before pouring the plates. Store plates at 4 °C.

### 3.5. Plasmid preparation and sequencing

MINI and MIDI preps have been performed with the PureYield™ Plasmid Miniprep System (Promega) and the PureYield™ Plasmid Midiprep System (Promega), respectively. Both kits were used according to the manufacturer's protocol. All newly cloned plasmids have been sent to sequencing at Microsynth, Austria.

### 3.6. Culturing of HEK293 cells

All experiments using human cells were performed with the HEK293 cell line derived from human embryonic kidney. Dulbecco's modified Eagle medium (DMEM, Sigma Aldrich) supplemented with 10% FCS (Sigma Aldrich) was used for culturing the cells. Splitting took place each third day, the dish was rinsed once with 1x Dulbecco's Phosphate Buffered Saline (Sigma Aldrich) and trypsinized (Sigma Aldrich). Cells were incubated at 37 °C and 5% CO<sub>2</sub>.

Cells were passaged around 40 times before a new aliquot was thawed and used to establish a new culture. Cryo-stocks for freezing were prepared according to the following protocol:

1. Remove media from 10 cm plate and wash the plate with 1x DPBS.
2. Trypsinize cells by adding 1 ml of 1x trypsin and incubate for 2 minutes at 37 °C and 5% CO<sub>2</sub>.
3. Add 9 ml of fresh media and transfer cell solution into a 15 ml falcon tube.
4. Centrifuge at 4 °C, 1200 x g for 5 minutes.
5. Remove supernatant and wash pellet in ice-cold 1x DPBS.
6. Repeat step 4.
7. Resuspend the cell pellet in DMEM/10% DMSO and prepare 1 ml aliquots.
8. Transfer cells to – 80 °C for at least 24 hours before storing them permanently in liquid nitrogen.

For thawing a cryo-stock and establishing a new culture, the following protocol was used:

1. Carefully thaw cryo-stock on ice.
2. Mix with 9 ml DMEM supplemented with 10% FCS.
3. Centrifuge at 4 °C, 1200 x g for 5 minutes.
4. Remove supernatant and resuspend pellet in fresh DMEM with 10% FCS.
5. Transfer cell suspension to a 10 cm plate and incubate at 37 °C and 5% CO<sub>2</sub>.

### **3.7. Transfection of HEK293 cells**

All plasmids used for this thesis were transiently transfected into HEK293 cells using FuGENE®HD (Promega). Transfection reactions were carried out in 6-well plates seeded the day prior to the experiment with a cell density of  $6 * 10^5 \frac{cells}{well}$ . The total amount of plasmid transfected for 1 well were 4 µg with a ratio of 1:5 for RNA:protein coding plasmids. A typical transfection reaction consisted of 13.3 µl FuGENE FuGENE®HD, 4 µg of plasmids and ddH<sub>2</sub>O to the final volume of 150 µl. Proper complex formation between plasmids and the transfection reagent was ensured by a 30 minutes incubation step at RT. Afterwards medium was exchanged with 3 ml of fresh one and the transfection mix was added drop-wise to each well. Each 6-well plate was incubated for 24 hours before being washed with 1x PBS, trypsinized and expanded to 150mm dishes (3 wells of a 6-well plate were used for 1 150mm dish). After each dish has reached 95% confluency (typically 72 hours after transfection) UV cross-linking was performed

### 3.8. UV cross-linking and total RNA isolation

After reaching 95% confluency dishes were rinsed twice with ice-cold PBS (Sigma Aldrich), placed on an ice-cold metal block and subjected to 200mJ/cm<sup>2</sup> UV-radiation in a Stratalinker© 2400. Control samples were kept on ice during UV-treatment. Afterwards PBS was carefully removed and 700 µl CHAPS lysis buffer were added to each dish. Cells were then collected with a cell scrapper and transferred to a 1.5 ml tube to proceed with total RNA isolation. Therefore the cells were incubated on a rotating wheel at 4 °C for 20 minutes to complete cell lysis. To ensure complete digestion of proteins by Proteinase K all samples were put at 37 °C for 1 minute and cleared by centrifugation at 14 000 x g, 4 °C for 30 minutes. The supernatant was then transferred to a fresh 1.5 ml tube, filled up to 500 µl total volume (if necessary) with CHAPS lysis buffer. Subsequently phenol extraction was performed as described below:

1. Add 500 µl of water-saturated phenol (Applichem) to the sample. Vortex thoroughly.
2. Centrifuge at 14 000 x g, 4 °C for 5 minutes.
3. Transfer aqueous phase to a new 1.5 ml tube, add 500 µl chloroform:isoamylalcohol [24:1 [v/v]].
4. Repeat step 2.
5. Precipitate the aqueous phase with 120 µl 3M NaOAc pH 5.0 and 600 µl isopropyl alcohol o/n.
6. Centrifuge sample at 14 000 x g, 4 °C for 30 minutes.
7. Take off supernatant, air-dry pellet for 10 minutes and resuspend the RNA in 15 µl ddH<sub>2</sub>O.
8. Measure concentration in a 1:30 dilution via an UV-VIS spectrometer (Nanodrop2000C).

#### CHAPS lysis buffer

1M Tris-HCl pH 7.5 [f.c. 10mM]	100 µl	
5M NaCl [f.c. 400mM]	800 µl	
0.5M EGTA [f.c. 1mM]	20 µl	
1M MgCl <sub>2</sub> [f.c. 1mM]	10 µl	
5% CHAPS [f.c. 0.5% [v/v]]	1000 µl	
14.3M β-Mercaptoethanol [f.c. 7mM]	500 µl	add immediately before use
Proteinase K [f.c. 0.1mA/µl]	1142 µl	add immediately before use

Add ddH<sub>2</sub>O to a final volume of 10 ml and store at RT.

### 3.9. Reverse transcription

The reactions were performed as described below; all reactions are prepared on ice:

1. Pipette 1  $\mu$ l 4.5x hybridization buffer in a fresh 1.5 ml tube.
2. Add 2.5  $\mu$ l total RNA [10 $\mu$ g/ $\mu$ l].
3. Complement with 1  $\mu$ l of <sup>32</sup>P-labelled gene-specific primer.
4. Denature sample for 1 minute at 95 °C and immediately put on ice for 2 minutes.
5. Add 15.5  $\mu$ l of extension mix.
6. Incubate for 60 minutes at 55 °C.
7. To degrade template RNA add 3  $\mu$ l of 1M NaOH and incubate for 55 °C for 45 minutes.
8. Neutralize pH with 3  $\mu$ l 1M HCl; add 2.5  $\mu$ l 0.5M EDTA pH 8 and 2  $\mu$ l glycogen [10mg/ml].
9. Precipitate sample with 75  $\mu$ l of 0.3M NaOAc/EtOH and freeze at -20 °C for 60 minutes.
10. Centrifuge sample at 14 000 x g, 4 °C for 30 minutes.
11. Remove supernatant, air-dry the pellet for 10 minutes and resuspend in 8  $\mu$ l 7M Urea loading buffer.
12. Resolve half of the sample on a 8% denaturing PAGE (70 W, 140 minutes)

#### 4.5x Hybridization buffer

1M K-Hepes pH 7 [f.c. 225mM]                      45  $\mu$ l

1M KCl [f.c. 450mM]                                      90  $\mu$ l

Fill up with ddH<sub>2</sub>O to 200  $\mu$ l and store at – 20 °C.

#### Extension mix

5x Transcriptor RT buffer [f.c. 1x]                      4  $\mu$ l

10mM dNTPs [f.c. 1mM]                                      2  $\mu$ l

0.1M DTT [f.c. 5mM]    1  $\mu$ l

RNAse Inhibitor [40u/ $\mu$ l] [f.c. 10u]                      0.25  $\mu$ l

Transcriptor RT [20u/ $\mu$ l] [f.c. 5u]                      0.25  $\mu$ l

ddH<sub>2</sub>O    8  $\mu$ l

#### 7M urea loading buffer

Urea [f.c. 7M]	21 g
Sucrose [f.c. 25%[w/v]]	12.5 g
Bromophenol blue [f.c. 0.025%[w/v]]	10 µl
10x TBE [f.c. 1x]	5 ml

Fill up to 50 ml with ddH<sub>2</sub>O and store at RT.

#### 10x TBE buffer

Tris-HCl [f.c. 890mM]	108 g
Boric acid [f.c. 890mM]	55 g
EDTA [f.c. 20mM]	5.8 g

Fill up to 1 l with ddH<sub>2</sub>O, filter and autoclave. Store at RT.

### 3.10. Denaturing PAGE

For resolving the pool of cDNAs a denaturing polyacrylamide gel electrophoresis (PAGE) is performed. The gel is cast between glass plates (52cm x 33cm) separated by 0.4mm spacers. The usual volume for a typical gel is 70 ml denaturing 8% acrylamide solution with 40 µl of 6.7M tetramethylethylenediamine (TEMED) and 400 µl of 10% ammonium persulfate solution. Polymerization of the gel lasts for at least 60 minutes, followed by assembling the gel in the apparatus and placing a metal plate for heat dispersion. The gel is then pre-run with 1x TBE buffer at 65 W for 45 minutes. Before loading 4 µl of samples each well is cleared from urea by rinsing the well with 1x TBE. The actual run is then performed at 65 W for about 2 hours. In case of RT samples prepared with primer hTR\_404 a long run was performed (4 hours) as well. After disassembling the apparatus the gel sandwich is opened with a plastic wedge, transferred to Whatmann 3M paper, covered with saran wrap and dried on a vacuum slab gel drier at 80 °C for at least 60 minutes. The dried gel is then exposed to phosphor imager screen (GE Healthcare) for 24-48 hours and scanned by a Typhoon Trio (GE Healthcare).

Interestingly, nucleotides G214 to G219, A352 to C356 and C391 to C398 appear squashed on the gels and therefore the resolution of this stretch is too low to reliably analyse for cross-links at the aforementioned nucleotides. This finding was highly reproducible for all experiments performed. This

behaviour is most likely related to the method of primer extension than to *in vivo* UV cross-linking. The fact, that we observed very similar patterns when performing DMS probing on hTR, strengthens this explanation. Most likely, newly formed secondary structures during reverse transcription and/or intrinsic polymerase features might cause these artefacts.

#### 8% acryl amide solution

Urea [f.c. 7M]	210 g
40% acrylamide 19:1 [f.c. 8% [v/v]]	100 ml
10x TBE [f.c. 1x]	50 ml

Fill up to 500 ml with ddH<sub>2</sub>O and filter, store at 4 °C.

#### 10% ammonium persulfate

Ammonium persulfate [f.c. 10% [w/v]]	5 g
--------------------------------------	-----

Fill up to 50 ml with ddH<sub>2</sub>O, aliquot and store at – 20 °C.

### 3.11. SDS-PAGE

Cells were grown and transfected as described in section 4.7. The protocol for harvesting the cells was adapted from that outlined in section 4.8 as SIGMAFAST™ Protease Inhibitor Cocktail Tablets (Sigma) were added to the CHAPS lysis buffer. 5 µl of cleared lysate fraction were loaded on an 8% SDS-PAGE run with 1x GTS at 150 V until the bromophenol blue dye reaches the bottom of the gel. 5 µl of pre-stained protein marker (NEB) was loaded for control. Casting and running of the gel was done with a MIGHTY SMALL II apparatus (Hoefer). The size of the gels was 8cm x 7cm with 1.5mm spacers. The gel was then either stained with Coomassie Brilliant Blue to visualize protein bands or directly used for Western blotting. Table 3-4 shows the composition of the stacking and separating gel, respectively:

Component	8% stacking gel	3% separating gel
ddH <sub>2</sub> O	4.62 ml	4.1 ml
1.5M Tris HCl pH 8.8 [f.c. 375mM]	2.5 ml	-
1.5M Tris HCl pH 6.8 [f.c. 375mM]	-	417 µl
10% SDS	0.1 ml	50 µl

<b>30% acrylamide (29:1)</b>	2.67 ml	487.5 $\mu$ l
<b>10% APS</b>	0.1 ml	50 $\mu$ l
<b>TEMED</b>	10 $\mu$ l	5 $\mu$ l

**Table 3-4. Composition for the stacking and separation gel for SDS PAGE.**

#### 10x GTS

Tris Base [f.c. 250mM]	60.6 g
Glycine [f.c. 1.9M]	288 g
10% SDS [f.c. 1% [v/v]]	200 ml

Add ddH<sub>2</sub>O to 2 l, filter and store at RT.

#### 10x SDS-PAGE loading buffer

Tris-HCl pH 6.8 [f.c. 50mM]	0.33 ml
SDS [f.c. 2% [w/v]]	0.8 g
Glycerol [f.c. 10% [w/v]]	4 ml
2-mercaptoethanol [f.c. 1% [v/v]]	0.4 $\mu$ l
0.5M EDTA pH 8.0 [f.c. 12.5mM]	0.25 ml
Bromphenol blue [f.c. 0.02%[w/v]]	8 mg

Add ddH<sub>2</sub>O to 10 ml and store at RT.

#### Coomassie Brilliant Blue

Coomassie Brilliant Blue [f.c. 0.1% [w/v]]	1 g
Methanol [f.c. 50% [v/v]]	500 ml
Glacial acetic acid [f.c. 10% [v/v]]	100 ml
ddH <sub>2</sub> O	400 ml

Stir for 3 – 4 hours, filter and store at RT.

### 3.12. Western blotting & chemiluminescence detection

The SDS-PAGE gel was assembled into a blotting sandwich. All components were soaked in transfer buffer prior to use. The order from anode to cathode was the following: sponge, three layers of Whatmann paper 3MM, nitrocellulose membrane, polyacrylamide gel, three layers of Whatmann paper 3MM, sponge. The transfer was performed overnight in a wet-blotting chamber filled with transfer buffer at 4 °C and 25 V. The pore size of the membrane (Amersham Hybond ECL) used was 0.45 µm. Successful transfer was then confirmed by Ponceau staining. The membrane was transferred to a 50 ml tube and treated as described below in Table 3-5:

Procedure:	Workflow:
Blocking step:	120 minutes at RT with gentle shaking;
Solution used:	TBS with 0.1% Tween-20 and 5% milk
Washing step:	2x for 5 minutes each in 1x TBS
1 <sup>st</sup> antibody:	Incubate membrane o/n at 4 °C, 1:2000 dilution with gentle shaking
Antibody used:	α-FLAG M2 in TBS with 0.1% Tween-20 and 5% Milk
Washing step:	3x for 5 minutes each in 1x TBS
2 <sup>nd</sup> antibody:	Incubate membrane for 60 minutes, 1:10000 dilution with gentle shaking
Antibody used:	α-IgG HRP in TBS with 0.1% Tween-20 and 5% Milk
Washing step:	8x in 1x TBS+0.1% Tween-20 for 5 minutes

**Table 3-5. Workflow for Western blotting.**

After blotting the membrane is shortly dried between several layers of paper towels and incubated in 1 ml of the development solution (according to manufacturer's manual) from the Pierce ECL Plus kit (Thermo Scientific) for 5 minutes. Excess of development solution was removed by tipping the membrane with one corner on a paper towel. The dried membrane was then wrapped in Saran wrap and transferred into a cassette. Exposure to Amersham Hyperfilm™ ECL (GE Healthcare) lasted between 10 seconds and 1 minute. Subsequently the film was developed using an AGFA Curix 60.

#### 10x transfer buffer

Tris base [f.c. 250mM]	90.9 g
Glycine [f.c. 2.5M]	187.6 g

Dissolve in 1l ddH<sub>2</sub>O and store at RT.

#### Ponceau staining solution

Ponceau S [f.c. 2% [w/v]]	2 g
Trichloroacetic acid [f.c. 30% [v/v]]	30 ml
Sulfosalicylic acid [f.c. 30% [v/v]]	30 ml

Fill up to 100 ml with ddH<sub>2</sub>O and store at RT.

### 3.13. Gel-purification of DNA oligos

Primers were ordered from Sigma Aldrich in desalted grade. For purification 40 µl of 100µM primer are mixed with 40 µl of 7M urea loading buffer and separated on a 20% denaturing PAGE. The size of the gel is 20 cm x 22 cm with 1.5 mm spacers. Running parameters were 25 W constant for 60 to 90 minutes (all other parameters are identical to section 4.11). Gel run was completed when the bromophenol blue dye reaches the bottom of the gel. The gel was then carefully wrapped in Saran foil and transferred onto a fluorescent thin-layer chromatography plate and the DNA oligos were visualized by UV-light at 254 nm. The corresponding band was cut out from the gel and put in a 1.5 ml tube containing 0.5 ml elution buffer and put on shaking thermo mixer (37 °C, 4 hours). The gel remnants were collected by centrifugation for 2 minutes at 13000 x g. The supernatant was transferred to a new 2 ml tube and after adding 2 µl glycogen [10mg/ml] the samples was precipitated at -20 °C overnight with 1.25 ml EtOH. After centrifugation (13000 x g, 30 minutes, 4 °C) the pellet was resuspended in 15 µl ddH<sub>2</sub>O followed by concentration measurement of a 1:30 dilution via a UV-VIS spectrometer (Nanodrop2000c).

#### Elution buffer

1M Tris HCl pH 7.5 [f.c. 10mM]	0.5 ml
3M Sodium acetate pH 5.0 [f.c. 250mM]	4.16 ml
0.5M EDTA pH 8.0 [f.c. 2mm]	0.2 ml

Fill up with ddH<sub>2</sub>O to 50 ml and store at RT.

### 3.14. 5' end-labeling of DNA oligos

The DNA oligo was diluted to 10µM with ddH<sub>2</sub>O. 10 pmol of oligo were used for each labeling reaction with <sup>32</sup>P-γ-ATP. After assembling the labeling reaction (table 3-6), the sample was incubated for 45 minutes at 37 °C. After addition of 1.5 µl 0.5M EDTA pH 8 the tube was incubated for 1 minute at 95 °C and immediately put for 2 minutes on ice to increase the yield. To precipitate the sample 1 µl of glycogen [10mg/ml] was added followed by 35 µl 0.3M sodium acetate pH 5.0/EtOH. Precipitation lasted for 60 minutes at -20 °C followed by centrifugation at 4 °C, 13 000 x g for 30 minutes. The supernatant was carefully removed and the pellet was air-dried for 10 minutes. Finally the dried pellet was resuspend in 20 µl ddH<sub>2</sub>O and stored at -20 °C until further use. Table 3-6 shows the reaction set up:

Component	[ $\mu$ l]	[f.c.]
10 $\mu$ M DNA oligo	1	10 pmol
10x T4 PNK (10u/ $\mu$ l)	0.5	10 units
10x T4 PNK buffer	1	1x
<sup>32</sup> P- $\gamma$ -ATP (6000mCi/mmol and 10 $\mu$ Ci/ $\mu$ l)	3	30 $\mu$ Ci
ddH <sub>2</sub> O	4.5	-
<b>Total</b>	<b>10</b>	-

**Table 3-6. Setup of 5' end-labeling reaction.**

### 3.15. Direct telomerase assay

To characterize the activity and processivity of telomerase, direct telomerase assays (DTAs) were carried out. The protocol is based on that from Cristofari et al<sup>172</sup>. In brief, the transiently transfected HEK293 cells were lysed as described in sections 5.7 and 5.11. Concentration of these lysates was measured via Bradford assay. For each reaction (see table 3-7) 20  $\mu$ g of protein lysate was used. The sample was incubated at 30 °C for 60 minutes followed by addition of 5  $\mu$ l heat-inactivated RNase A (10mg/ml) and subsequent incubation at 37 °C for 10 minutes. To digest proteins present in the sample 15  $\mu$ l of Proteinase K solution (900mAU/ml) were added and incubated for 30 minutes at 37 °C. The sample was precipitated by adding 1  $\mu$ l of a <sup>32</sup>P-labeled 100mer oligo (serves as loading control), 2  $\mu$ l 0.5M EDTA pH 8, 1  $\mu$ l glycogen (10mg/ml) and 125  $\mu$ l 0.3M NaOAc/EtOH pH 5.0 were added and the sample was put at -20 °C for one hour. The sample was then centrifuged at 4 °C, 13 000 x g for 30 minutes and the dried pellet was resuspended in 10  $\mu$ l 7M urea loading buffer and analysed on an 8% denaturing PAGE (see section 5.11).

Component	[ $\mu$ l]	[f.c.]
Protein lysate [5 $\mu$ g/ $\mu$ l]	4	20 $\mu$ g
10x DTA buffer	2	1x
dNTPs (25mM dATP, 25mM dTTP, 0.1mM dGTP)	0.4	-
<sup>32</sup> P-dGTP (6000mCi/mmol and 10 $\mu$ Ci/ $\mu$ l)	2	20 $\mu$ Ci
10 $\mu$ M oligo (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub>	2	1 $\mu$ M
0.1M 2-mercaptoethanol	1	5 $\mu$ M
ddH <sub>2</sub> O	8.6	-

**Table 3-7. Reaction setup for direct telomerase assay**

### 3.16. Data analysis

Quantification of UV-induced cross-links was performed with ImageQuant 7.0 from GE Healthcare. The raw data for each gel was exported into Microsoft Excel and the local maximum for each nucleotide was extracted. In order to correct for loading differences between the – lanes and the + lane, a mean normalization factor from three different bands in the – lane was calculated and applied to the + lane. For the quantification of the UV cross-linking pattern in presence and absence of hTERT, the ratio of the normalized + lanes from hTERT + and hTERT – lanes was calculated. Changes above 1.5 and below 0.66 have been considered significant.

Nucleotide addition processivity (NAP) and repeat addition processivity (RAP) were calculated as described by Zaug et al<sup>69</sup>. In brief, the intensity for each repeat was divided by the amount of incorporated dGTP residues (e.g. 1 G for the first repeat, 4 for the second repeat and 7 for the third repeat). Total NAP was obtained by summing up the values for every single repeat, which was set to 100% for the wildtype. Processivity was determined by the formula  $R_{1/2} = -\frac{\log 2}{k}$  with k being the slope of the formula  $n = \frac{(\text{intensity of repeat } n)}{(\text{number of G's in repeat } n)}$ . Again, the wildtype value was arbitrary set as 100%.

## 4. Results

The human telomerase RNP complex consists of two major players: human telomerase RNA and the protein human telomerase reverse transcriptase. Both are needed *in vivo* to obtain a functional complex, together with additional accessory proteins<sup>11-13</sup>. In order to assess the spatial organization of the telomerase holoenzyme, we have developed and successfully applied an *in vivo* UV cross-linking assay on human telomerase RNA.

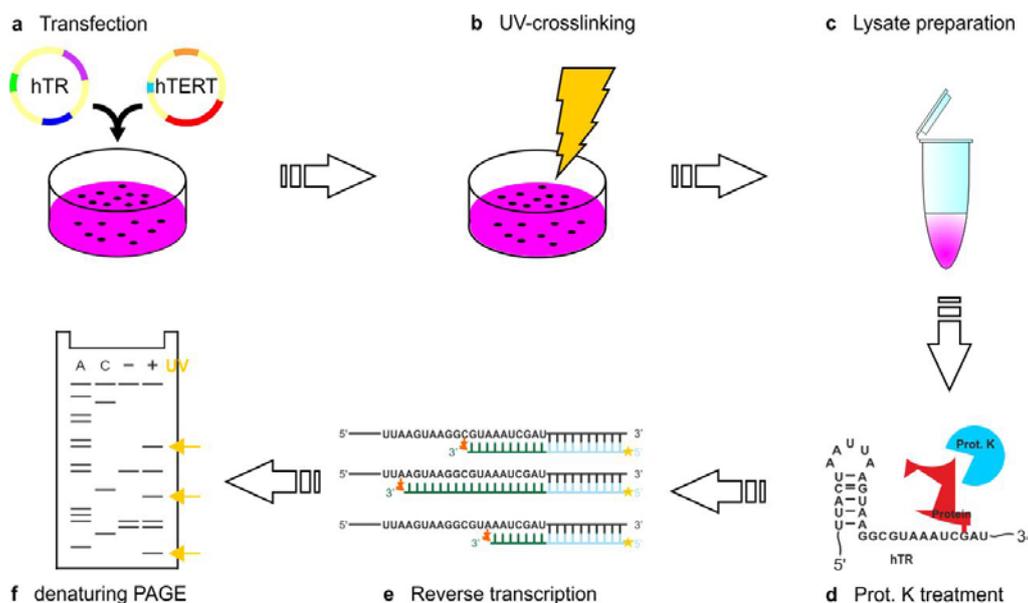
### 4.1. Experimental setup – In vivo UV cross-linking of RNA

Back in the 1960s it was discovered that ultraviolet light is able to induce the formation of covalent bonds between two spatial close residues and can therefore be used to identify RNA-RNA or RNA-protein interactions<sup>173</sup>. In contrast to chemical cross-linking agents such as formaldehyde, UV-light introduces considerable fewer disturbances by interfering with the *in vivo* structure of the cross-linked complex resulting in fewer artifacts<sup>174</sup>. UV light at a wavelength of 254 nm can easily penetrate a single cell layer and acts as a zero-length cross-linking agent, thereby forming a covalent bond between two atoms which are in close proximity to each other. Mechanistically, the formation of cyclobutane pyrimidine dimers (CPD) is the most often observed phenomena and also the most frequent reason for lesions in RNA/DNA strands<sup>175</sup>. Upon excitation with UV light at 254 nm, pyrimidines, such as cytosine and uracil, form new covalent bonds between them using the C5-C6 double bond<sup>176</sup>. Another possibility is the formation of 6-4 photoproducts with the C6 atom of either cytosine or uracil and the O4 of uridine<sup>177</sup>. The probability of bond formation is also dependent on the sequence and the orientation of the bases towards each other<sup>176</sup>. Of course, C5, C6 and O4 can also form a cross-link to atoms not being part of a neighbouring base, when these are in close proximity. The photoreaction itself is an extremely fast process, happening almost instantaneously upon UV irradiation (pico seconds time scale)<sup>175</sup>.

To investigate the structural organization of the human telomerase RNA, we decided to use *in vivo* UV cross-linking on telomerase-expressing HEK293 cells (figure 4.1). In brief, HEK293 cells are co-transfected with plasmids coding for hTR and hTERT. Notably, RNA from irradiated and non-treated cells is harvested in parallel and total RNA is extracted. The UV cross-linking sites are mapped by reverse transcription with hTR-specific primers, followed by separating the pool of cDNA on a denaturing PAGE. Sites of UV cross-links can be identified due to inducing an RT stop at the site of the cross-link. The comparison of cDNA derived from RNA isolated from untreated cells with cDNAs reverse transcribed from RNA isolated from UV-irradiated cells allows for discrimination between natural RT stops and UV-induced stops. The identified cross-link sites are then plotted onto a secondary structure map of hTR, revealing novel structural information.

Several intrinsic features of human telomerase RNA result in major challenges to overcome. For example, hTR is very low abundant in somatic cells and its high GC content makes it extremely challenging to design gene-specific primers for down-stream reverse transcription. By using an *in vivo* over-expression system based on HEK293 cells, we have successfully overcome the low abundance of hTR in normal cells<sup>178</sup>. However, we tested for the contribution of endogenous levels of hTR and hTERT to telomerase activity<sup>178</sup>. As such, activity assays (direct telomerase assays, section 3.15) were performed on lysates derived from three control setups: (a) untransfected cells; b) cells transfected with hTR only and c) cells transfected with hTERT only). No telomerase elongation products were detected, while lysates from cells transfected with both hTR and hTERT showed considerable telomerase activity<sup>178</sup>. Human TR and hTERT are transfected in a 1:5 ratio to ensure that the majority of hTR is bound in the telomerase complex. This ratio is used for the UV-irradiated cells as well as for the untreated control dishes with the only exception being the experiments with hTR alone. In this case the plasmid coding for hTERT is replaced with the empty backbone vector, while the 1:5 ratio is kept constant.

Based on the observations from our control experiments, we conclude that our experimental setup is capable of producing high amounts of active telomerase complexes, while the amount of unbound hTR within the cell and the endogenous levels of both hTR and hTERT are negligible<sup>178</sup>. 48 hours post transfection the cells are used for *in vivo* UV cross-linking. With 451 nucleotides in length, several primers were used to fully map the hTR molecule. So far, we have identified 70 cross-linked nucleotides (16% of all 451 nts), spread over the entire hTR molecule. In detail, the cross-links consist of 42 uridines (60%), 18 cytosines (26%), six guanines (8%) and four adenosines (6%). In line with previous publications<sup>175,179</sup>, we observed, that pyrimidines are much more likely to be cross-linked than purines (89% versus 11%) and that base pairing-state of the cross-linked nucleotides play an important role as well. 52 out of 63 cross-linked nucleotides do not form a canonical base pair according to the secondary structure map (figure 1.4), while only eleven nucleotides are found within a paired region.



**Figure 4.1. Pipeline for detecting UV cross-linking sites in hTR.** a) HEK293 cells are transiently co-transfected with plasmids coding for hTR and hTERT. b) After 48 hours of expression the cell mono layer is subjected to UV-light at 254 nm. c) The irradiated cells are harvested and lysates are prepared. d) In order to digest the proteins covalently bound to the RNA (red), the lysates are treated with Proteinase K (light blue). e) Total RNA is extracted and used as template for reverse transcription with hTR-specific  $^{32}\text{P}$ -labelled primers (light blue, yellow star). Due to the newly formed covalent bonds at a cross-link site (orange), the reverse transcriptase terminates at n-1 for each cross-link site generating cDNAs of different length (green). f) The pool of cDNA is separated via denaturing PAGE, dried on a vacuum slab gel dryer and exposed. Sequencing lanes allow for orientation along the RNA, while the identification of UV cross-linking sites is done by comparing the – lane (cDNA reverse-transcribed from RNA isolated from untreated cells) with the + lane (cDNA reverse-transcribed from RNA isolated from UV-exposed cells). Cross-linking sites only show up in the + lane, while natural RT stops are found in both lanes and allow for discrimination between those two types of stops. The cross-linking sites found are then plotted onto the secondary structure map for visualization.

## 4.2. The structural organization of human telomerase RNA

### 4.2.1. Organization of the pseudoknot domain

A hallmark of all telomerase RNAs known so far is the presence of a pseudoknot<sup>44</sup>. In human telomerase RNA the pseudoknot domain can be divided into the core pseudoknot, consisting of helices P2b and P3, and the extended pseudoknot, which also includes helices P2a and P2a.1 with the latter being only found in mammals<sup>25,44</sup>. Helix P2a is connected to the core pseudoknot via the five nucleotide internal loop J2a/2b. It is found in all vertebrate telomerase RNAs and usually has a length of five nucleotides in mammals<sup>17,28</sup>. The solution structure of J2a/2b revealed its importance for the overall topology of the pseudoknot domain, as this internal loop introduces a S-shape in the backbone, resulting in large bend between P2a and P2b<sup>28</sup>. We have observed cross-links at nucleotides throughout J2a/2b and the flanking helices, in detail U81, U82, U83, U86, C87, C88 and

C89 (figure 4.2 a). U81 to U83 are located in the P2a helix, forming A-U base pairs (U81-A128, U82-A127 and U83-A126), while U86 to C88 are part of J2a/2b and C89 is the first nucleotide of P2b. Interestingly, the bend in the helical orientation occurs at position U86, whose cross-linking intensity is only moderate, while the intensity is strongest for the cross-link at U83, the last base pair of P2a, and U87 (figure 4.3 a). Previous work in our lab using DMS probing of hTR revealed that *in vivo* all helices in the core pseudoknot and in the extended pseudoknot, are fully formed and not accessible to DMS, except for the 3' end of P2a<sup>178</sup>. Residues A127 and A128, as well as the last base pair of P2a (U83-A126) are modified by DMS. This is due to the special form of the backbone in J2a/2b, which introduces a S-shape, thereby resulting in the partial opening of the flanking helices P2a and P2b<sup>28</sup>, especially U81 to U83 and C89. The cross-links surrounding J2a/2b might therefore reflect this particular backbone geometry.

The core pseudoknot consists mostly of nucleotides belonging to both the conserved region 2 (CR2, nucleotides C92 to G120) and conserved region 3 (CR3, A172 to C183)<sup>17</sup>. An early study has shown that the helical integrity of helices P2a.1, P2a, P2b and P3 is crucial for telomerase activity and disrupting one of the helices leads to a severe reduction in telomerase activity<sup>27</sup>. Notably, respective residues like G107 and C108 in P3 are associated with a disease-related double mutation (G107A/C108G) in patients suffering from dyskeratosis congenita<sup>180</sup>. The mode of action is the destabilization of the pseudoknot by disrupting the extensive network of base triples formed between P2b and parts of J2a.1/3 and P3 and J2b/3<sup>25</sup>. Recently, a triple helix was also found in the telomerase RNA of *S. cerevisiae* and *K. lactis*<sup>29,181</sup>, underlining the importance of these tertiary contacts.

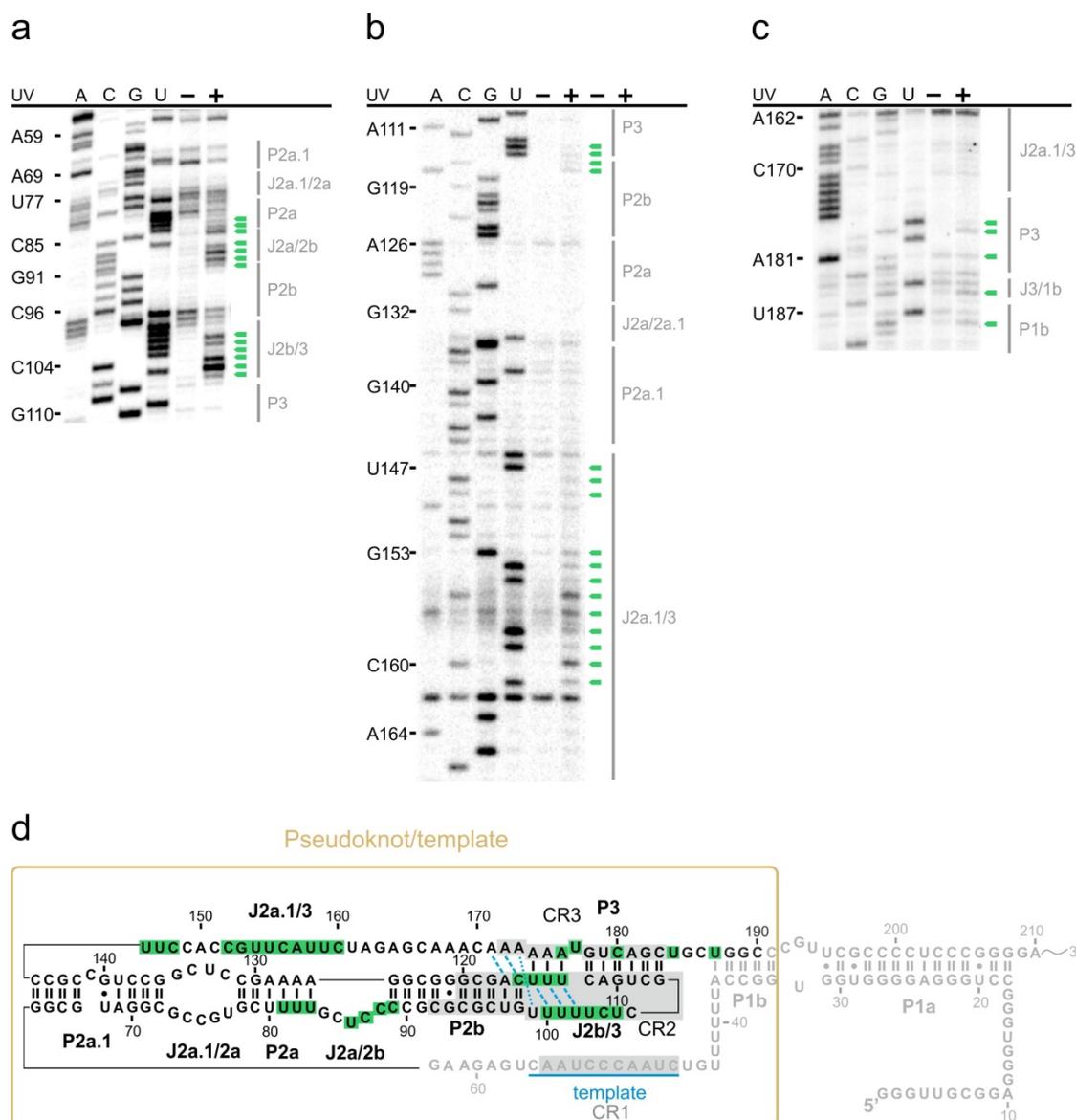
In order to form base triples, the interacting nucleotides must be located with their Hoogsteen edge in close contact to each other. Consequently, it was not surprising to observe cross-links on one or more nucleotides forming the Hoogsteen base pairs of the triple helix (figure 4.2 a, b and d). In detail, cross-links are observed at U101 and U102 (J2b/3), U113, U114 and U115 (P3) and C116 (P2b). These nucleotides are part of four of the five Hoogsteen base triples (U115-A174•U100, U114-A175•U101, U113-A176•U102 and C116-G98•A172). Interestingly, for nucleotide U99, which interacts with A173, no cross-link was observed. However, U99 seems to be a natural RT stop as well and a potential cross-link at this site might be therefore masked. Internal loop J2b/3 is a hot spot for cross-linking, as six out of eight nucleotides are cross-linked. It is noteworthy that the linker J2b/3 is not only conserved in sequence but also in length, indicating its important role in stabilizing the pseudoknot<sup>17</sup>. Aside from aforementioned residues U101 and U102, which are part of the triple helix, the remainder is not known to participate in tertiary interactions. However, the intensities for cross-links at C104 and U103 are the highest within the entire pseudoknot domain, while the signal for flanking

nucleotides U100, U101, U102 and U105 is of average intensity. According to the NMR solution structure of the core pseudoknot (nts 95 to 119 and 170 to 183 from WT hTR), C104 is oriented towards the major groove of P3 and becomes buried by the J2b/3-P3 interaction<sup>25</sup>. In contrast, U103 points to the major groove of P3 between base pairs C112-G178 and A111-U179<sup>25</sup>. The observed cross-links at the residues involved in the formation of the triple helix (U100, U101, U102, U113, U114, U115 and C116) are a good indication that the triple helix is indeed formed *in vivo*. The cross-links at U103 and C104 might reflect tertiary interactions of these nucleotides, but more data is needed to answer this question.

Of the cross-links observed in P3, U177 is of special interest: a single nucleotide bulge in helix P3 (figure 4.2 c), which is a pyrimidine in 83% of all known vertebrate telomerase RNAs<sup>17</sup>. According to the solution structure of the core pseudoknot, U177 prevents base stacking of A176, which is also cross-linked, and G178, by introducing a slight bend in helix P3<sup>25</sup>. This minor change in local geometry was shown to influence the folding dynamics of the pseudoknot. By preventing stabilizing base stacking interactions to take place, stability in this region is no longer dictated by stacking of the canonical A-U base pairs (U113-A176, U114-A175 and U115-A174). Instead, the balance shifts towards the stacking interactions of the Hoogsteen base triples (U115-A174•U100, U114-A175•U101, U113-A176•U102), while in the  $\Delta$ U177 pseudoknot this is not the case<sup>25</sup>. Consequently, in the wildtype pseudoknot the Watson-Crick base pairs upstream of U177 open up, while the Hoogsteen base pairs stay intact. In contrast, in absence of U177 the situation is reversed with Hoogsteen base triples opening up and Watson-Crick base pairs remaining intact<sup>25</sup>. Furthermore, U177 effectively terminates the triple helix with U113-A176•U102 being the last Hoogsteen base triple, while in the  $\Delta$ U177 pseudoknot the triple helix is expanded for another Hoogsteen base triple (G178-C112•U103)<sup>25</sup>. However, the solution structure does not contain hTERT and therefore the exposed position of U177 might be a different one *in vivo*. For the other two cross-links in P3 and J3/1b (C180 and U184, respectively), only limited information is available. While C180 is 80% conserved in vertebrate telomerase RNAs, the identity of U184 varies. However the three nucleotide linker J3/1b is found in every vertebrate and might play an important role despite poor nucleotide conservation<sup>17</sup>.

Previous studies reported a dimerization of hTR *in vitro* via the P3 stem and its nucleotides A174 to C183<sup>58,182</sup>. This results in a dimer with both hTR molecules forming a *trans*-pseudoknot by interacting via their P3 stems. In fact, three of these nucleotides are cross-linked in our case: the aforementioned A176, U177 and C180. Nonetheless, it is important to point out that most of the dimerization studies have been carried out, before the solution structure of the pseudoknot domain was solved in 2005. In fact, Theimer et al. did not observe a correlation between telomerase activity

and the ability to form dimers<sup>25</sup>. This indicates that the preservation of the tertiary interactions forming the triple helix is more important and plays a greater role than a putative dimerization. In addition, dimerization mutants yielded only poor telomerase activity *in vivo*, while performing similar to the wildtype telomerase *in vitro*<sup>58,182</sup>. However, this is in contrast to the last cryo electron microscopy study of *in vivo* reconstituted human telomerase. According to Sauerwald and co-workers human telomerase has a bilobal shape and a weight consistent with a putative dimer<sup>183</sup>. When fitting the telomerase ring from *T. castaneum* into the electron density map, the TERT protein is found at the end of the structure, while the RNA has been placed close to the hinge region<sup>183</sup>. The authors therefore conclude that dimerization of telomerase is most likely carried out via RNA-RNA interactions<sup>183</sup>.



**Figure 4.2.** The important elements J2a/2b and the triple helix are prone to UV cross-linking. Representative primer extension gel showing UV-induced cross-linking sites in human telomerase RNA. Secondary structure elements shown are a) P2a.1, J2a.1/2a, P2a, J2a/2b, P2b, J2b/3, P3, b) P3, P2a, P2b,

J2a/2a.1, P2a.1, J2a.1/3, c) J2a.1/3, P3, J3/1b and P1b. Lanes A, C, G and U are sequencing lanes, the – lane contains cDNA reverse transcribed from RNA isolated from cells not treated with UV light, while the + lane shows cDNA reverse transcribed from RNA isolated from cells exposed to UV light. The sites of cross-linking are identified by comparing the – lane with the + lane. Bands present in both lanes correspond to natural RT stops. Bands only found in the + lanes correspond to a UV cross-linking event and are marked (green arrows). **d)** The cross-linked nucleotides are plotted onto the secondary structure map of hTR (green boxes). The template region is labelled in blue, while the base triples in the pseudoknot are highlighted by blue, dashed lines and the non-canonical base pair U99•A173 is indicated by a blue dotted line. Residues which are highly conserved are shaded in gray. The areas shown in **a)**, **b)** and **c)** are shown in black (base pairs and letters), while the neighbouring regions are depicted in light gray.

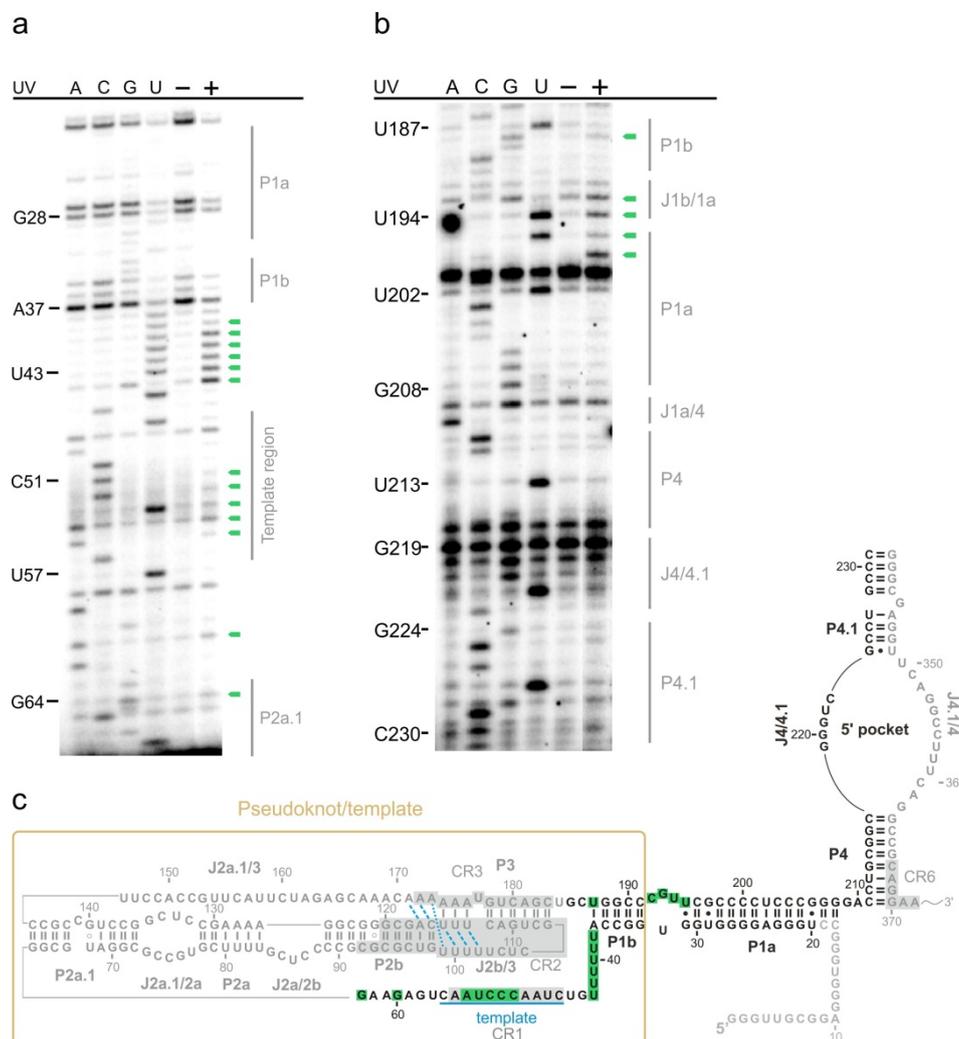
The last cross-links observed in the pseudoknot region are located in the linker J2a.1/3 which connects the mammalian-specific helix P2a.1 with P3. In contrast to the other cross-linked, single-stranded regions, like J2a/2b and J2b/3, the linker J2a.1/3 is neither conserved in sequence nor in length and varies greatly throughout known vertebrate telomerase RNAs<sup>17</sup>. Computational modelling of the human telomerase core (nucleotides 1 to 208) with distance constraints derived from FRET studies places the J2a.1/3 linker region next to the pseudoknot as well as near the template, however the model is only refined to six to eight Ångström<sup>184</sup>. Cross-links were observed at nucleotides U146, U147 and C148 (figure 4.2 b), located at the 5' end of J2a.1/3 and on a continuous stretch of nucleotides C152 to C160 (figure 4.2 b). Interestingly, two out of the three non-pyrimidine nucleotides cross-linked in the pseudoknot are located in this stretch (G153 and A157). So far, the function of J2a.1/3 is not well understood, as earlier studies focused on highly conserved structures, such as the helices of the core pseudoknot and the junctions J2a/2b and J2b/3.

In summary, 34 cross-linked nucleotides were observed in the pseudoknot and the extended pseudoknot. The majority of them, 28, are located in single-stranded regions like the linkers J2a/2b, J2b/3 and J2a.1/3. The remainder are observed in the flanking nucleotides of J2a/2b and on nucleotides involved in forming the triple helix.

#### **4.2.2. Assessing the conformation of the template region**

The template element of the human telomerase RNA contains the conserved region 1 (CR1) and defines the length and sequence of the telomeric repeats, which are added to the telomeres<sup>44</sup>. It is embedded into a long single-stranded stretch of residues connecting helix P1b with helix P2a.1<sup>17</sup>. Cross-links were observed in a stretch of uridines (U38-U43), which is located between the template boundary element (helix P1b) and the template (figure 4.3 a). In human, the exact definition of the template is strongly dependent on correct formation of helix P1b and the length of the linker between P1b and the template<sup>185</sup>. Changes in length in the uridine stretch U38 to U43 result in read-through events and incorporation of non-canonical residues<sup>185</sup>.

For the template itself, cross-links are observed at nucleotides C50, C51, C52, U53 and A54 (figure 4.3 a). While *in vivo* DMS probing of hTR revealed that the entire template is accessible to DMS and therefore not involved in Watson-Crick base pairing<sup>178</sup>, cross-links are only observed at the 3' end of the template. Interestingly, nucleotides C52 to A54 are important for annealing of the telomeric DNA to the template RNA<sup>90,186</sup> and therefore the cross-links at these nucleotides might represent interactions essential for the formation of the RNA-DNA duplex. The last two cross-links are observed at nucleotides G60 and G63 in the single-stranded linker connecting the template to the extended pseudoknot. Both of them might also play a role in template positioning, however they are only 50% conserved in vertebrate telomerase RNAs<sup>17</sup>.



**Figure 4.3 Nucleotides in and adjacent to the template are cross-linked.** Representative gel of primer extension showing UV-induced cross-linking sites in human telomerase RNA. Secondary structure elements shown are **a)** P1a, P1b, the template region, P2a.1, **b)** P1b, J1b/1a, P1a, J1a/4, P4, J4/4.1 and P4.1. **c)** Secondary structure of hTR with cross-links indicated by green boxes. Symbol and color code as in figure 4.2.

In human telomerase RNA, the TBE consists of helices P1 and its substructures P1a and P1b. Within the TBE five cross-links were observed. Two of them are located in helical areas (U187 and U195), while the other three (C192, G193 and U194) are found in single-stranded regions. Both nucleotides U187 and U195 are part of the terminal base pairs of P1b and P1a, respectively (figure 4.3 b). U187 forms a canonical base pair with A38 and U195 pairs with G31 to form a G•U wobble pair. Nucleotides C192, G193 and U194 are part of the internal loop J1b/1a (figure 4.3 b), an element only found in human telomerase RNA, although bulges throughout the length of P1 are quite common in vertebrate telomerase RNAs<sup>17</sup>.

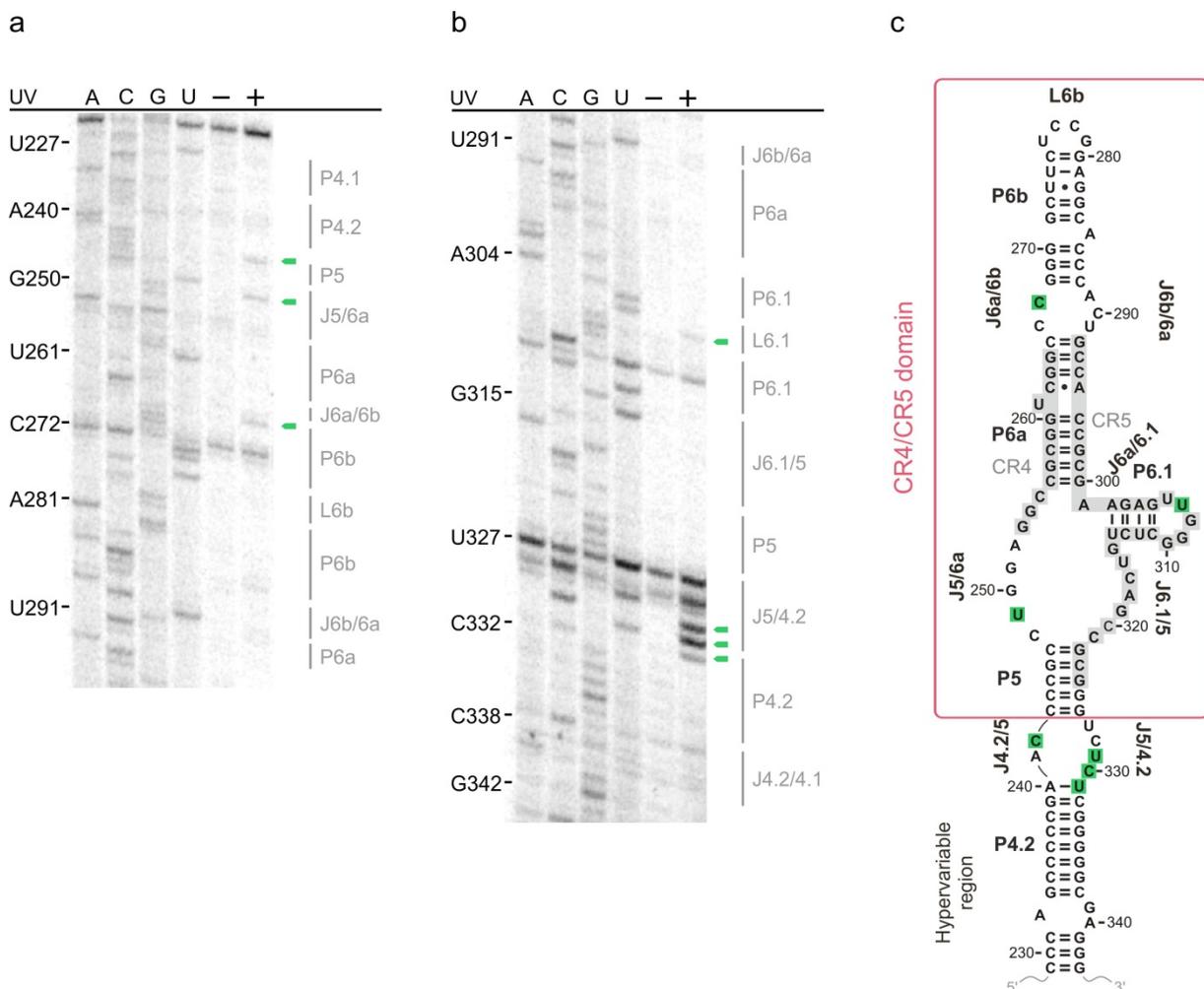
#### **4.2.3. Only a few cross-links are found in the CR4/CR5 domain**

The CR4/CR5 domain has been the target of several studies<sup>17,31,33,75</sup>. It contains the conserved regions four and five and is also thought to be the primary binding site for hTERT<sup>33</sup>. In vertebrates, the structure of the CR4/CR5 domain is dominated by a large three-way junction, which is responsible for binding to the TERT protein<sup>33</sup>. In human telomerase RNA the CR4/CR5 domain consists of the P5-P6a-P6.1 three-way junction, whereby stem P6a is extended by an internal loop (J6a/6a and J6b/6a) and helix P6b closed by the terminal loop L6b (figure 4.4 c). The most conserved nucleotides are located in stems P6a, P6.1 and J6.1/5<sup>17</sup> and form the conserved regions 4 and 5, respectively. Only three cross-linked nucleotides were found in the CR4/CR5 domain (figure 4.4 a, b).

The first nucleotide cross-linked is U249, located in the large junction connecting helices P5 and P6a (figure 4.4 a). In 34 out of 35 vertebrate species, nucleotide 249 is pyrimidine (97%), whereby a U is found considerably more often at this position than a C (75% versus 25%)<sup>17</sup>. The P6a helix does not show any cross-linked nucleotide, however the adjacent internal loop connecting P6a and P6b, harbours the second cross-linked nucleotide: C267 is located in the junction J6a/6b and less conserved than U249. Seven out of 35 vertebrate species lack this nucleotide, however if present, the tendency for a pyrimidine at this position is high (77%)<sup>17</sup>. The solution structure of the P6 element with the internal loop (J6a/6b and J6b/6a) and the terminal loop L6b suggests a base triple forming between C267 and the canonical G268-C288 base pair<sup>187</sup>. Since the O2 of C267 is in unusual close proximity to the O6 of G268, this suggests that the potential base triple is mediated by water<sup>187</sup>. However, a second conformation in which C267 stacks on top of C266 is possible as well<sup>187</sup>. In both cases the Watson-Crick as well as the major groove face of C267 is available for additional interactions<sup>187</sup>. Helix P6b is capped by a UCCG tetraloop, therefore belonging to the UNCG class of tetraloops. Due to their drastically increased thermal stability, UNCG tetraloops often act as nucleation site for correct folding or as protein binding platforms<sup>188</sup>. While GNRA tetraloops are often involved in tertiary interactions, for example the  $\zeta$ - $\zeta'$  interaction in group II introns<sup>189</sup>, UNCG tetraloops are rarely involved in tertiary interactions<sup>188</sup>. This also seems to be the case for the

terminal tetraloop of helix P6b, whose nucleotides were not observed to become cross-linked by UV light (Figure 4.4 a).

Junction J6b/6a, located opposite of the cross-linked nucleotide C267, shows no cross-linked nucleotide. G292, the first nucleotide of the paired region P6a, is also the first nucleotide of the conserved region 5 (CR5), which is involved in formation of helix P6a, the extremely highly conserved P6.1 stem, J6.1/5 and most of helix P5. Interestingly, within this area, a single cross-link at position U307 was identified (figure 4.4 b, c). In vertebrate telomerase RNA, U307 is 100% conserved and was shown to be of utmost importance for telomerase activity<sup>31,190</sup>. Earlier *in vitro* UV cross-linking studies showed a physical interaction of U307 with A54 as well as U306 with C46, U45 and G44, suggesting that L6.1 is in close proximity of the template<sup>32</sup>. Stem P6.1 consists of a total of four A-U and G-C base pairs in alternating order. Interestingly, all of the nucleotides in P6.1, as well as A301 in J6a/6.1, are 100% conserved in vertebrate telomerase RNAs and substitutions of these nucleotides severely affect telomerase activity<sup>17,31,190</sup>. Nonetheless, cross-links were not detected at any nucleotide in P6.1 (figure 4.4 b). The same holds true for J6.1/5; this is interesting with regard to the cross-link at U249 in J5/6a.



**Figure 4.4 The conserved three-way junction in the CR4/CR5 domain shows minor cross-linking.** Representative gel of primer extension showing UV-induced cross-linking sites in human telomerase RNA. Secondary structure elements shown are a) P4.1, P4.2, P5, J5/6a, P6a, J6a/6b/, P6b, L6b, P6b, J6b/6a, P6a, b) P6b, L6b, P6b, J6b/6a, P6a, P6.1, L6.1, P6.1, J6.1/5, P5, J5/4.2, P4.2 and J4.2/4.1. c) Secondary structure of hTR with cross-links indicated by green boxes. Symbol and color code as in figure 4.2.

The CR4/CR5 domain is connected to the pseudoknot/template domain as well as to the scaRNAs domain by a number of helices and loops which are termed hypervariable region (figure 1.4). In contrast to the highly conserved CR4/CR5 domain, the connecting region varies greatly between species and shows only poor sequence conservation<sup>17</sup>. Sequence alignments for several vertebrate telomerase RNAs proposed two paired regions and three large internal loops<sup>17</sup>. The hypervariable region of human telomerase RNA consists of the long P4 element, which is divided into three helices (P4, P4.1 and P4.2), all connected by internal loops. While J4/4.1 and J4.1/4 form a large, asymmetrical internal loop (also referred to as 5' pocket), the remaining J4.1/4.2 is much smaller. Cross-links found in the hypervariable region are mostly located in the junctions connecting helices P4, P4.1 and P4.2 (figure 4.4 a and figure 4.5 a).

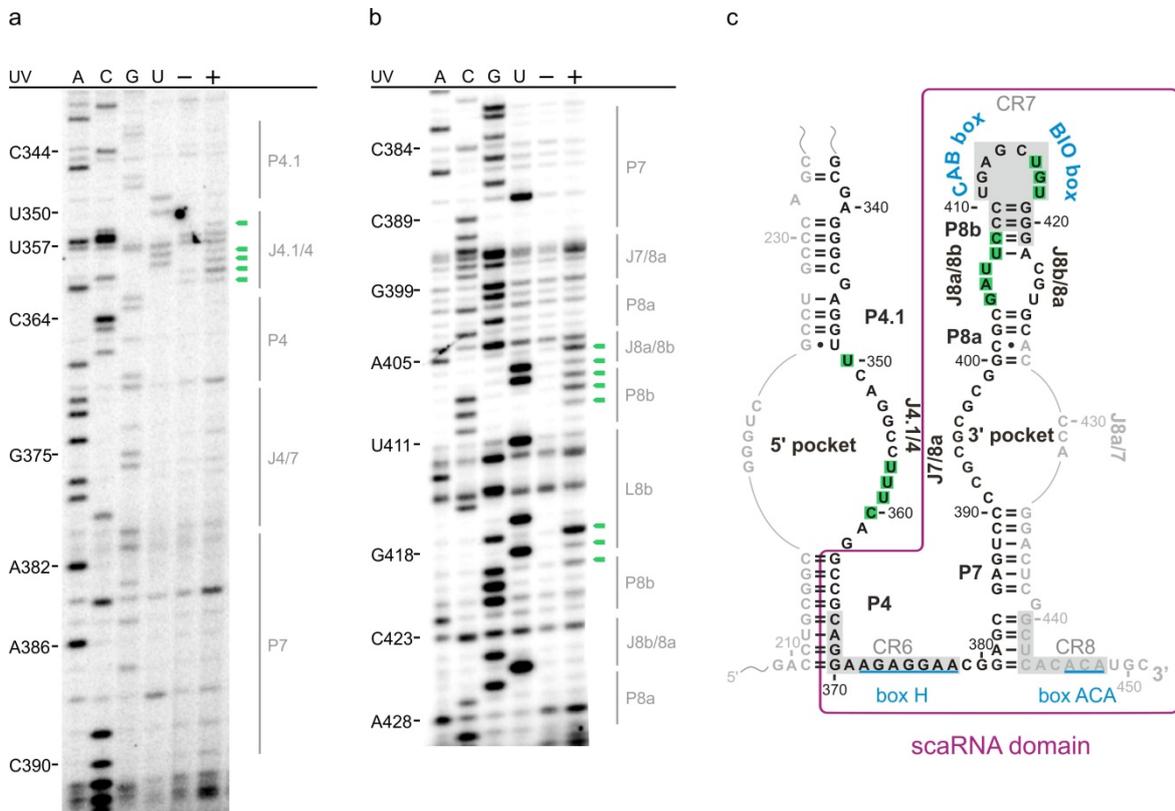
C242 is located in junction J4.2/5, an asymmetrical internal loop connecting helix P4.2 with the first helix of the CR4/CR5 domain, P5. Interestingly, cross-links in J5/4.2 are observed as well; nucleotides U329 and C330 are cross-linked, as well as U331, which forms an A-U base pair with A240 (figure 4.4 a). In addition, cross-links were also found in J4/4.1 at nucleotides U350, U357, U358, U359 and C360 (figure 4.5 a) with U350 being the first nucleotide of J4.1/4, directly downstream of the closing G•U wobble pair of helix P4.1. Due to the small degree of sequence conservation within the hypervariable region, little is known about the role of this region and it is mostly seen as a linker connecting two domains both conserved in sequence and structure.

Taken together, only few cross-links are found in the CR4/CR5 region, however the high degree of sequence conservation of the cross-linked nucleotides might indicate the importance of these cross-links.

#### **4.2.4. The CAB box is cross-linked within the scaRNA domain**

The last functional domain of the human telomerase RNA is the scaRNA domain. It is located at the 3' end of the molecule and harbours the conserved regions 6 to 8 (CR6-8)<sup>17</sup>. In contrast to the pseudoknot domain and the CR4/CR5 domain, the scaRNA domain is dispensable *in vitro* to reconstitute a functional telomerase complex, however the scaRNA domain is needed *in vivo*<sup>35,36,191</sup>. It contains the hallmarks of snoRNAs with the H/ACA box found in J4/7 and the trinucleotide at the distal 3' end<sup>36</sup>. The secondary structure of the region also resembles the characteristic hairpin-hinge-hairpin scheme of snoRNAs which acts as a binding platform for proteins<sup>36</sup>. The scaRNA domain plays

an important role in biogenesis and maturation<sup>19</sup>. Two important features for this task are the CAB box and the BIO box; the first is responsible for correct association of hTR with the Cajal bodies, while the second one increases the incorporation rate of hTR into the telomerase RNP complex<sup>36,40</sup>.



**Figure 4.5 The conserved BIO box is a target of UV cross-linking.** Representative gel of primer extension showing UV-induced cross-linking sites in human telomerase RNA. Secondary structure elements shown are **a)** P4.1, J4.1/4, P4, J4/7, P7, **b)** P7, J7/8a, P8a, J8a/8b, P8b, L8b, P8b, J8b/8a and P8a. **c)** Secondary structure of hTR with cross-links indicated by green boxes. Symbol and color code as in figure 4.2.

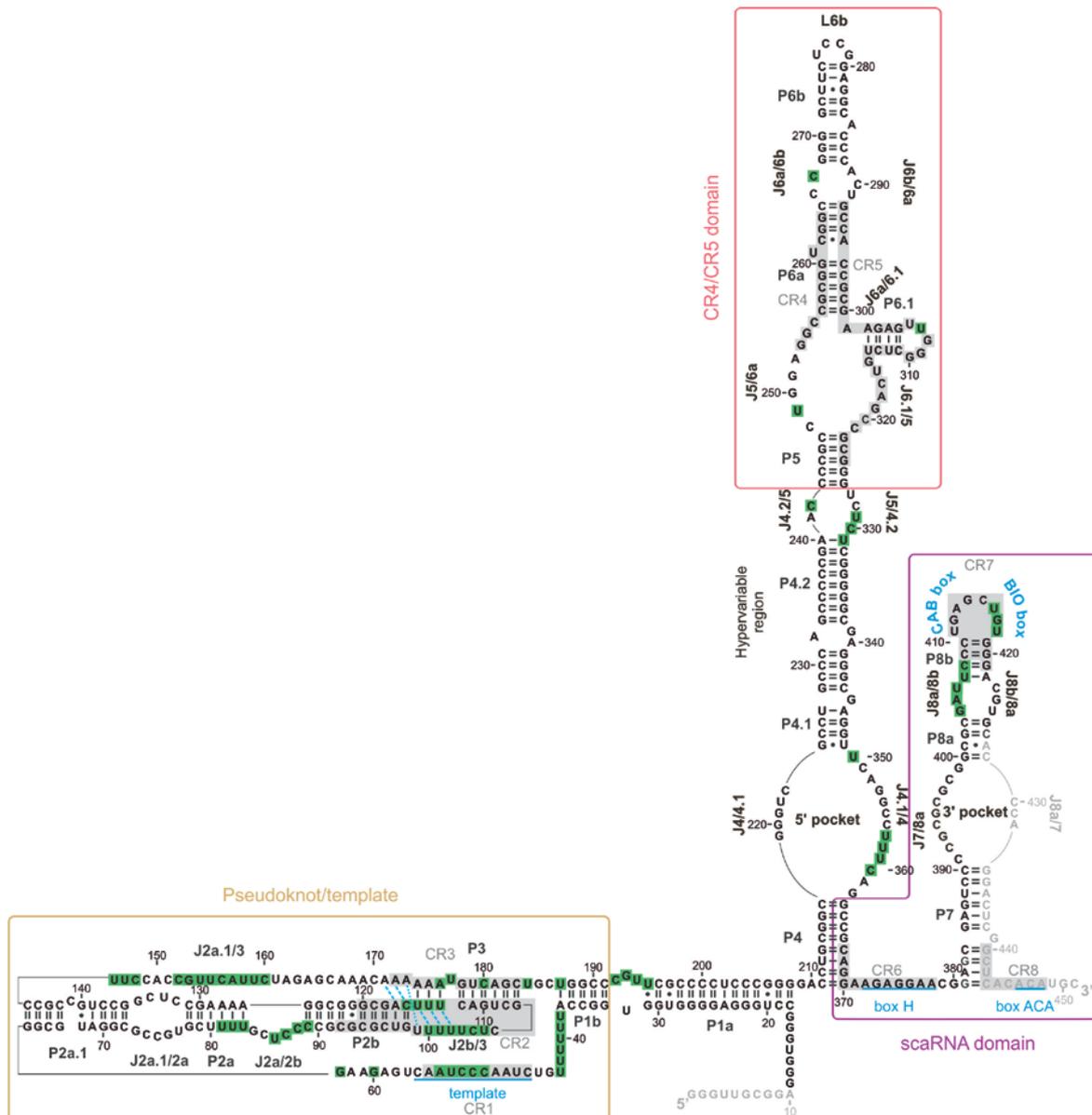
Inside P4, the conserved region 6 is found which contains the box H located in the single-stranded region between P4 and P7, the first helix of the scaRNA domain. Although the box H motif serves as protein binding platform<sup>192,193</sup>, cross-links were not detected at any nucleotide in J4/7. However, this could also be explained by the fact that eleven out of twelve nucleotides of the junction are purines, which are much less likely to be cross-linked than pyrimidines.

For helices P7 and P8a as well as J7/8a, the situation is similar to that of J4/7, as cross-links were not identified at respective residues. In contrast, the internal loop connecting P8a and P8b contains cross-linked nucleotides. G404, A405 and U406 form junction J8a/8b and show considerable UV cross-links (figure 4.5 b). In addition, G404 is one of the four guanosines found to be cross-linked in the entire hTR molecule (the others being G153 (J2a.1/3) and G193, both in the pseudoknot domain, and G417 in L8b). The next two nucleotides are already located in helix P8b and preceding conserved

region 7 (CR7). Both of them, U407 and C408 are cross-linked (figure 4.5 b), thus a complete stretch of five subsequent nucleotides spanning J8a/8b and P8b are intensely cross-linked within the scaRNA domain. Interestingly, the pyrimidine to purine mutation C408G was found to be linked to dyskeratosis congenita and is thought to disrupt stem P8b as well as the terminal loop L8b and impairs hTR localization to the Cajal bodies<sup>35,41,180</sup>. In line with this, DMS probing of the C408G mutant was impossible due to extremely low levels of hTR<sup>178</sup>.

The sequence element showing a high degree of conservation among vertebrate telomerase RNAs is the conserved region 7 (CR7) consisting of nucleotides C409 to G421. It spans the entire loop L8b and the adjacent base pairs in stem P8b (figure 4.5 c). Two elements important for hTR localization and maturation are located within L8b; the CAB box and the BIO box (section 1.3.3). While nucleotides U411 to G414, forming the CAB box, show no cross-links, the situation for the BIO box is entirely different. Here, cross-links were found at positions U416, G417 and U418 (figure 4.5 b). Especially U418 seems to play an important role, judged by its 100% conservation among vertebrate telomerase RNAs<sup>17</sup>. G417 is 90% conserved, while U416 is the least conserved with 80%<sup>17</sup>, but exhibits the strongest cross-linking signal of all three nucleotides (figure 4.5 b). The BIO box is thought to determine hTR's fate by tipping the balance towards RNP assembly rather than exonucleolytic degradation<sup>40</sup>. In this light, these three cross-links are highly interesting. Due to the requirement for primer binding during reverse transcription, it was impossible to map the distal 3' end of hTR, including nucleotides C429 to C451. The binding sites for the utmost 3' RT primer hTR\_433 are nucleotides G433 to C451 which allows for reliable mapping starting around nucleotide A428 (figure 4.5 c).

Plotting the observed UV cross-links onto the secondary structure map of hTR (figure 4.6) reveals that the majority of the cross-linked nucleotides are found in the pseudoknot/template domain: 43 out of 64 cross-linked nucleotides. The hypervariable region contains nine cross-linked nucleotides and only three cross-links have been observed in the CR4/CR5 domain. This was surprising, as we would have expected more cross-links in this highly conserved domain due to its proposed role in hTERT binding. In the scaRNA domain, eight nucleotides were found to be cross-linked. In addition, the vast majority of cross-linked nucleotides are located in single-stranded regions of hTR (51 out of 64) and only twelve nucleotides are located in paired regions (figure 4.6). It is also noteworthy that solely eight cross-link sites are composed of a single nucleotide, while the remainder (56 residues) consist of stretches of neighbouring nucleotides.



**Figure 4.6 UV-induced cross-links within human telomerase RNA in the presence of hTERT.** Overview of all UV-induced cross-linking sites (green boxes) within hTR in the presence of the protein partner hTERT. The conserved regions 1-8 are shaded in grey. Important sequence elements are shown in blue like the template region, the box H, the CAB and the BIO box as well as the ACA trinucleotide. The base triples within the triple helix in the pseudoknot are connected via dashed lines in blue and the non-canonical base pair U99•A173 is indicated by a blue dotted line. Symbol and color code as in figure 4.2.

### 4.3. hTERT induces structural changes within hTR

A fully functional telomerase holoenzyme relies upon two core components: the telomerase RNA and the corresponding protein TERT. In order to assemble into a mature telomerase complex, both subunits must be transcribed and, in the case of TERT, translated. For most of the time during the cell cycle, those two compounds are physically separated from each other in different nuclear compartments, joining only during S phase and being assembled to form the functional telomerase

complex, which is then transported to the telomeres<sup>19,134</sup>. Transcription and subsequent processing steps required to obtain mature human telomerase RNA are tightly regulated and were shown not to be influenced by hTERT<sup>19</sup>. However, when hTERT binds to hTR, discrete changes within the conformation of hTR might take place as previously reported<sup>178</sup>. In order to elucidate the differences hTERT impose on the conformation of human telomerase RNA, hTR was cross-linked in the absence of hTERT as well. In brief, HEK293 cells were co-transfected with the empty backbone vector pCDNA6 (referred to as pCW62 when containing the coding sequence for hTERT) and the hTR-encoding plasmid, while the remaining steps of the protocol are performed identically. We then compare the lanes of cDNA derived RNA extracted from cells transfected both with hTR and hTERT with the cDNA derived from RNA extracted from cells transfected with hTR alone. Since the endogenous levels of hTERT are too low to observe telomerase activity in our experimental setup<sup>178</sup>, we concluded that the influence of endogenous hTERT (if present at all) on the fold of the overexpressed hTR is negligible. As such, it was possible to compare the UV cross-linking pattern of hTR in the presence of hTERT with the pattern in absence of hTERT in order to infer the hTERT-induced conformational changes in hTR. After careful evaluation, we decided to set a threshold at  $\geq 1.5$  for the observed differences in cross-linking intensities. The majority of the observed changes resulted in stronger cross-links in hTR in absence of hTERT with some cross-links found exclusively in absence of hTERT. Only a single nucleotide showed a stronger cross-linking intensity in the presence of hTERT. The entirety of the changes is then visualized by plotting onto the secondary structure map of hTR (figure 4.16).

#### **4.3.1. The triple helix is influenced by hTERT binding**

Correct folding of the triple helix in the pseudoknot domain was shown to be essential in order to obtain a functional telomerase holoenzyme<sup>25</sup>. Disruption of the integral Hoogsteen base pairs has drastic effects on the stability of the pseudoknot and therefore leads to decreased or even abolished telomerase activity, both in human and yeast<sup>25,181,182</sup>. In fact, changes in cross-linking intensity were observed exclusively in the core pseudoknot, while the extended regions of the pseudoknot, like P2a.1, J2a.1/2a, P2a and J2a/2b, were not affected by the absence of hTERT.

In particular, nucleotides involved in forming the non-canonical Hoogsteen base triples show a significant increase in cross-linking intensity in the absence of hTERT (figure 4.7 a, b and figure 4.8). U113, which is part of the U113-A176•U102 base triple, shows an increase (1.9 fold change) in cross-linking intensity in the absence of hTERT (figure 4.8 b). Interestingly, A176 in P3 shows as well an increase in cross-linking intensity. The same is true for U114 and U115, part of U114-A175•U101 and U115-A174•U100, which show an even higher fold change with 2.2 and 1.7, respectively. C116,





**Figure 4.8 UV cross-linking pattern of the pseudoknot in presence and absence of hTERT.** Bar diagram shows the normalized ratio of hTR/hTR+hTERT for nucleotides G66 to U184. The bands from samples derived from cells transfected with hTR alone (lanes hTERT –) and from cells transfected with both hTR and hTERT (lanes hTERT +) are compared (figure 4.7). The mean value and standard deviation of the hTR/hTR+hTERT ratio for each nucleotide is shown from three independent experiments. A value of 1 represents no change in intensity in the presence and absence of hTERT (green line). The blue line indicates cross-links that were  $\geq 1.5$  more intense in the absence of hTERT, while the red line marks residues whose cross-linking efficiency was reduced in the absence of hTERT. The remaining nucleotides represent those which neither showed a cross-link in presence or absence of hTERT. Note that nucleotides U133 to G136 are not represented in the chart. These nucleotides could not be separated to single-nucleotide resolution making it impossible to assign an intensity value.

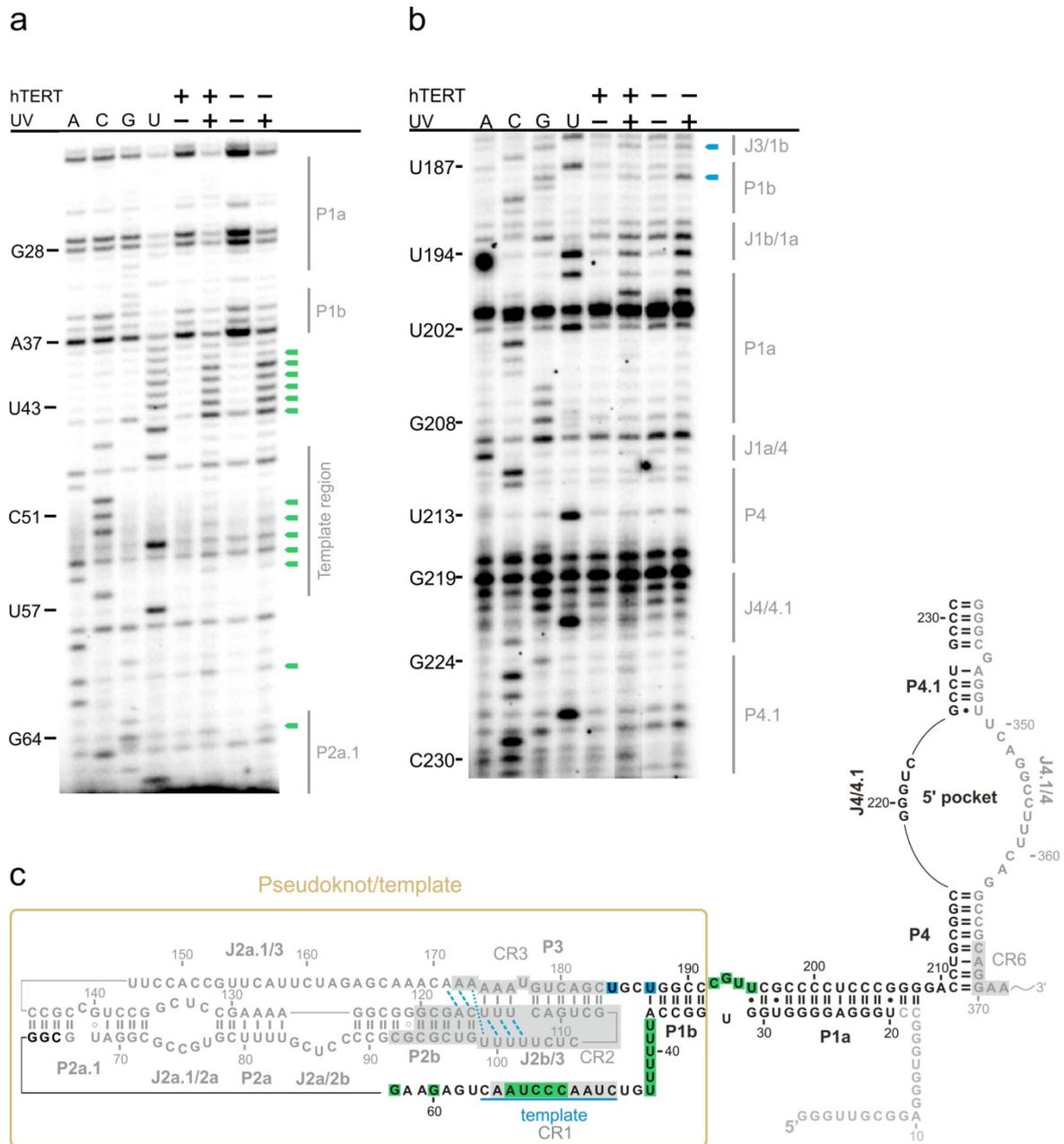
The remaining nucleotide A177, which is also involved in base triples, shows no increase in cross-linking intensity. According to the solution structure of the core pseudoknot every nucleotide showing an increase in cross-linking intensity uses its Watson-Crick edge to form the canonical base pair while the sugar edge is exposed and accessible via the minor groove of the helix<sup>25</sup>. If hTERT interacts with hTR via the minor groove with some of these nucleotides, small rearrangements in the local geometry might explain the changes in cross-linking intensity. An interaction mechanism via the minor groove would also explain the fact, that we did not observe changes in accessibility for C116 and A117 when performing DMS probing in presence – and absence of hTERT *in vivo*<sup>178</sup>.

The most striking changes, however, were observed in the linker J2b/3 connecting helices P2b and P3 in the pseudoknot. Here, U103 shows an increase in cross-linking intensity of 1.6, while the subsequent residue C104 displays a drop in intensity to 0.5 in absence of hTERT (figure 4.7 a and figure 4.8). These significant changes at two neighbouring nucleotides are unique within our results; however the implications are yet unclear. The altered cross-linking potential might indicate a cross-link to the protein or an intra-molecular cross-link. To our surprise, U103 and C104 are the least conserved nucleotides in J2b/3 with less than 80%, while the remainder is 100% conserved<sup>23</sup>. Importantly, the solution structure of the core pseudoknot showed that loop J2b/3 exhibits a high degree of dynamics, especially for nucleotides U103, C104 and U105<sup>23,25</sup>. Thus, our findings that hTERT influences the conformation of U103 and C104 are in good agreement with these reports. The reason for the increased flexibility of J2b/3 seems to be the bulged U177, which alters the position of junction J2b/3 and weakens the interactions of U103 and C104 with the nucleotides in stem P2b<sup>23</sup>. The change in the cross-linking intensity at nucleotide U177 could be due to its bulged nature, making U177 easily accessible for interactions with other residues. According to the solution structure the base of U103 is pointing away from J2b/3 and is facing towards the major groove of P2b, while C104 is completely buried in the J2b/3 loop<sup>25</sup>. How hTERT influences the spatial conformation of U103 and C104 remains enigmatic.

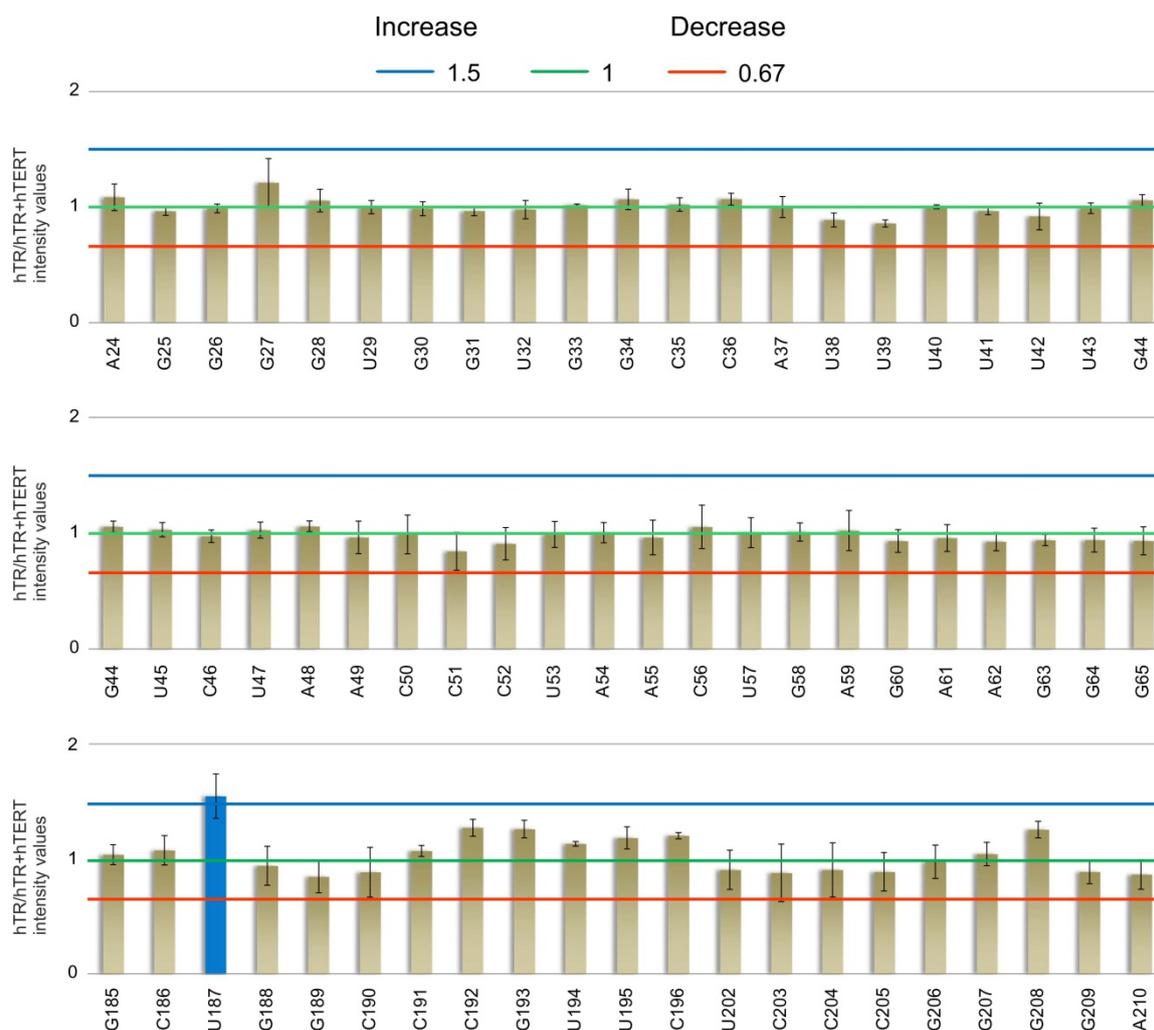
The last residues whose cross-linking intensity changes in absence of hTERT are C180 and U187. The intensity of the cross-link at nucleotide C180 increases more than 2-fold and is therefore together with U115 (2.1) and C104 (0.5) among the residues which show the largest change within the pseudoknot region. C180 forms a canonical base pair with G110, however its Hoogsteen edge is facing towards sugar edge of C104 as well as the phosphate group from the backbone<sup>23</sup>, which are both within hydrogen bonding distance (3.1 Å for C180:H5 with the 2'OH from C104 and 2.1 Å for C180:H42 to C104:O2)<sup>25</sup>. U187 is part of the first base pair of the template boundary element and shows moderate change with a 1.5-fold increase in cross-linking intensity (figure 4.7 c and figure 4.8). The increase might reflect changes in the TBE upon hTERT binding and template positioning. To our surprise none of the cross-linked nucleotides in the linker regions J2a/2b and J2a.1/3 showed a change in cross-linking intensity. This would indicate the majority of the pseudoknot conformation is already formed and act as a pre-organized scaffold for hTERT binding.

#### **4.3.2. The template region is not altered upon hTERT binding**

In contrast to the pseudoknot, in which several strong changes in cross-linking intensity were observed, the template region is not influenced by hTERT binding. No changes in cross-linking intensity were observed for nucleotides C50, C51, C52, U53 and A54 (figure 4.9 a and figure 4.10). The same holds true for the adjacent nucleotides located both 3' and 5' of the template region (G60 and G63 as well as U38, U39, U40, U41, U42 and U43), which also showed no change in their intensities in the absence of hTERT (figure 4.9 a and figure 4.10). As previously shown by *in vivo* DMS probing, the template region of human telomerase RNA becomes highly protected in absence of hTERT<sup>178</sup>, one would have expected a difference in cross-linking intensity within this region. However, while DMS probing tests for the accessibility of the N1 of adenines and the N3 of cytidines, the residues in the template could also form cross-links via the Hoogsteen and sugar edge, respectively, resulting in a change in DMS accessibility but could explain the presence of cross-links both in presence and absence of hTERT. Our observations are also interesting with regard to a recently published report from Brown and co-workers, in which they identified nucleotide A49 as a sequence-inherent pausing site<sup>91</sup>. We did not observe a cross-link at A49, however, the neighbouring nucleotides C50 to A54 are cross-linked. The implications still remain unclear. Similarly, the cross-links observed in J1b/1a were not affected by hTERT. A potential explanation for this might be found in the computational model of hTR derived from FRET distance constraints<sup>184</sup>. According to the model the template region itself is positioned in proximity to the pseudoknot, while the whole P1 element is relatively flexible and without a potential interaction partner<sup>184</sup>. However, it is necessary to point out that the model is based on information derived from *in vitro* experiments in absence of hTERT and therefore does not take into account the more complex situation *in vivo*.



**Figure 4.9** The template region and the template boundary element do not change in presence of hTERT. Representative primer extension gels showing regions **a**) P1a, P1b, the template region, P2a.1, **b**) J3/1b, P1b, J1b/1a, P1a, J1a/4, P4, J4/4.1 and P4.1. **c**) Secondary structure of hTR with nucleotides of the regions from **a**) and **b**) depicted in black. Neighbouring nucleotides not shown in **a**) and **b**) are shaded in gray for simplicity. Symbol and color code as in figure 4.2 and figure 4.7.



**Figure 4.10 UV cross-linking pattern of the template/TBE in presence and absence of hTERT.** The bar diagram shows the normalized ratio of hTR/hTR+hTERT for nucleotides A24 to G65 and G185 to A210. Note that nucleotides G197 to U201 are not represented in the chart. These nucleotides could not be resolved at single nucleotide resolution making it impossible to assign an intensity value. Color code as in figure 4.8.

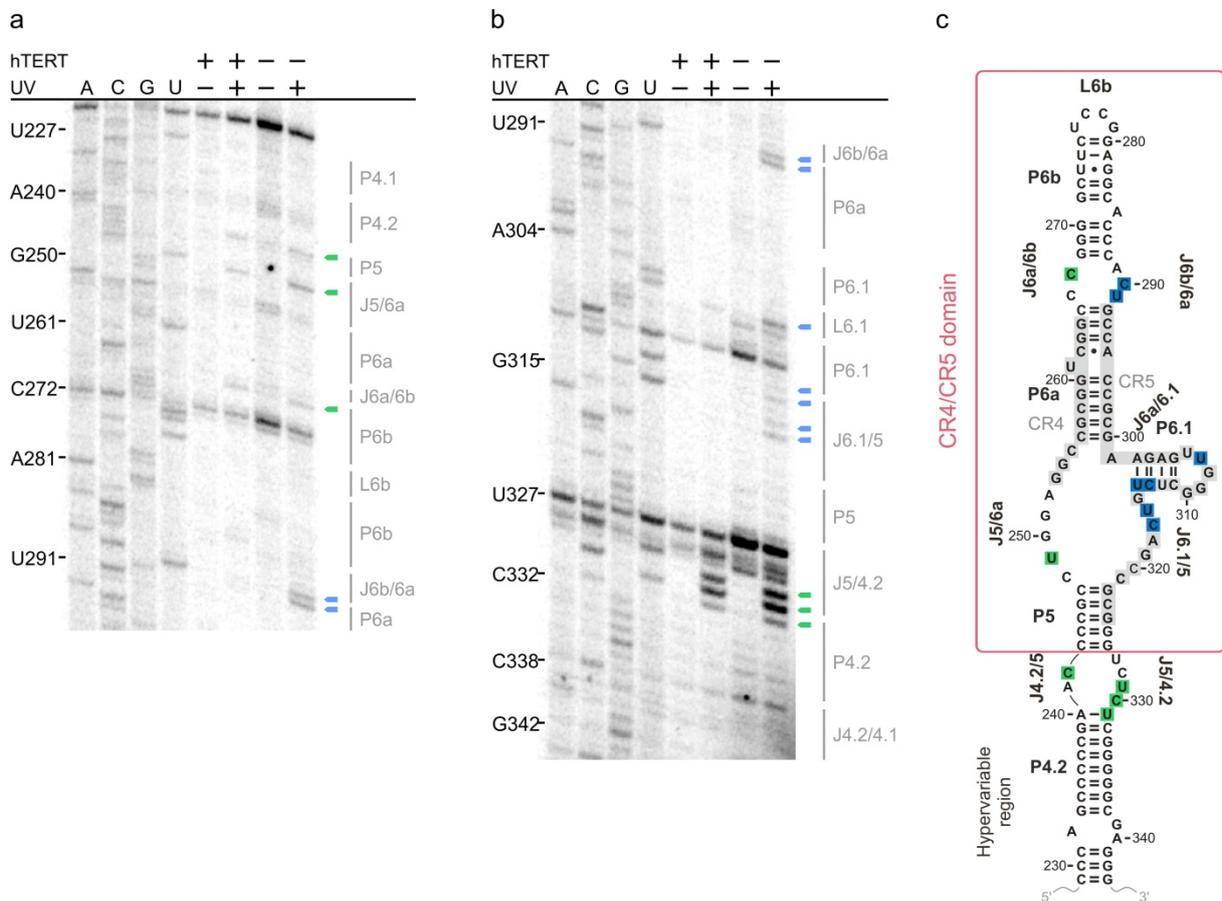
### 4.3.3. The CR4/CR5 domain is strongly influenced by hTERT binding

The CR4/CR5 domain is considered the main binding platform for hTERT, a hypothesis which is strongly supported up by work done on the smallest vertebrate telomerase RNA, mTR from *O. latipes*. In brief, upon mTERT binding major rearrangements in base pairing take place in the three-way junction of CR4/CR5, resulting in a conformational switch of the P6.1 stem-loop from one side of P6 to the other<sup>33</sup> (figure 1.8 and section 1.3.2). In light of this study and the fact that only three cross-links were observed in the presence of hTERT, it was of utmost interest to reveal the influence of hTERT on the CR4/CR5 element of hTR. Indeed, significant differences in UV cross-linking intensity in the absence of hTERT were identified in the CR4/CR5 region. Seven nucleotides display changes in their cross-linking pattern whereby six of the seven are novel cross-links not seen in presence of hTERT. This suggests that this region undergoes tremendous changes upon hTERT binding.

The first two nucleotides, which are cross-linked only in the absence of hTERT, are C290 and U291, both located upstream of helix P6a in the junction J6b/6a (figure 4.11 a and c). The intensity changes for C290 and U291 are 2.1x and 2.2x, respectively, and are among the highest we have observed. C290 is in 74% of vertebrate telomerase RNAs a pyrimidine (46% for C and 54% for U), while U291 is conserved to a higher degree (89% for U and pyrimidine as well). Several lines of evidence support a key role of the CR4/CR5 domain in hTERT binding<sup>31,33,43,187</sup>. First, the solution structure of P6a, the internal loop and P6b with the terminal loop L6b suggests a prominent role of J6b/6a in hTERT binding via nucleotides found in J6<sup>187</sup>. This is supported by *in vitro* studies reporting that a deletion of J6b/6a tremendously decreases hTERT binding to hTR and impairs catalytic activity as well<sup>13</sup>. The final proof, however, is the crystal structure of the medaka CR4/CR5 domain bound to the TRBD of mTERT<sup>33</sup>. Due to the smaller size of medaka telomerase RNA, no homology residues for the human C290 and U291 are found<sup>33</sup>. Interestingly, C290 is both accessible to DMS both in presence and absence of hTERT, suggesting that at least the N3 of C290 is not involved in hTERT binding<sup>178</sup>. U291 is stacked between C290 and G293 and could either form a C•U base pair with C266 or is bulged out as the canonical secondary structure of hTR implies<sup>187</sup>. Together with fact that C267 is also cross-linked both in absence and presence of hTERT, the available data suggests that a complex network of hydrogen bonds exist in the internal loop separating P6a and P6b. In contrast, no cross-links were again observed in P6a and J6a/6.1 harbouring the 100% conserved junction nucleotide A301. On the other hand, hTERT appears to influence P6.1 and adjacent residues in J6.1/5. First of all, the cross-link at U307 (figure 4.11 b and figure 4.12), which is 100% conserved in vertebrates is stronger in the absence of hTERT. When mutated telomerase activity is abolished, but mutating U307 does not influence binding of hTERT to the CR4/CR5 domain<sup>31</sup>. The important role of U307 for catalysis was also supported by the fact that *in vitro* cross-linking studies place this residue in proximity to the template and suggest interactions with nucleotide A54 in the template<sup>32</sup>. Indeed, a cross-link at A54 was observed (figure 4.9). The solution structure of P6.1 reports that U307 is solvent exposed and therefore able to form tertiary interactions<sup>43</sup>.

With hTR being a non-coding RNA the possibility for post-transcriptional modifications must be taken into account as well. Pseudouridines provide an additional hydrogen acceptor in the major groove edge, thus increasing the interaction possibilities of  $\Psi$  compared to Us. So far, a single study addressed this topic and identified six potential pseudouridines within human telomerase RNA<sup>42</sup>. Strikingly, two of them are located in the P6.1 loop:  $\Psi$ 306 and  $\Psi$ 307. Kim et al. also compared the solution structure of P6.1 with  $\Psi$ 6.1 and observed major differences in the loop topology<sup>42</sup>. For instance, when replacing the canonical Us with  $\Psi$ s, the loop closing G•U wobble pair forms only one hydrogen bond instead of two found in canonical G•U wobble pairs. Further, the thermal stability of

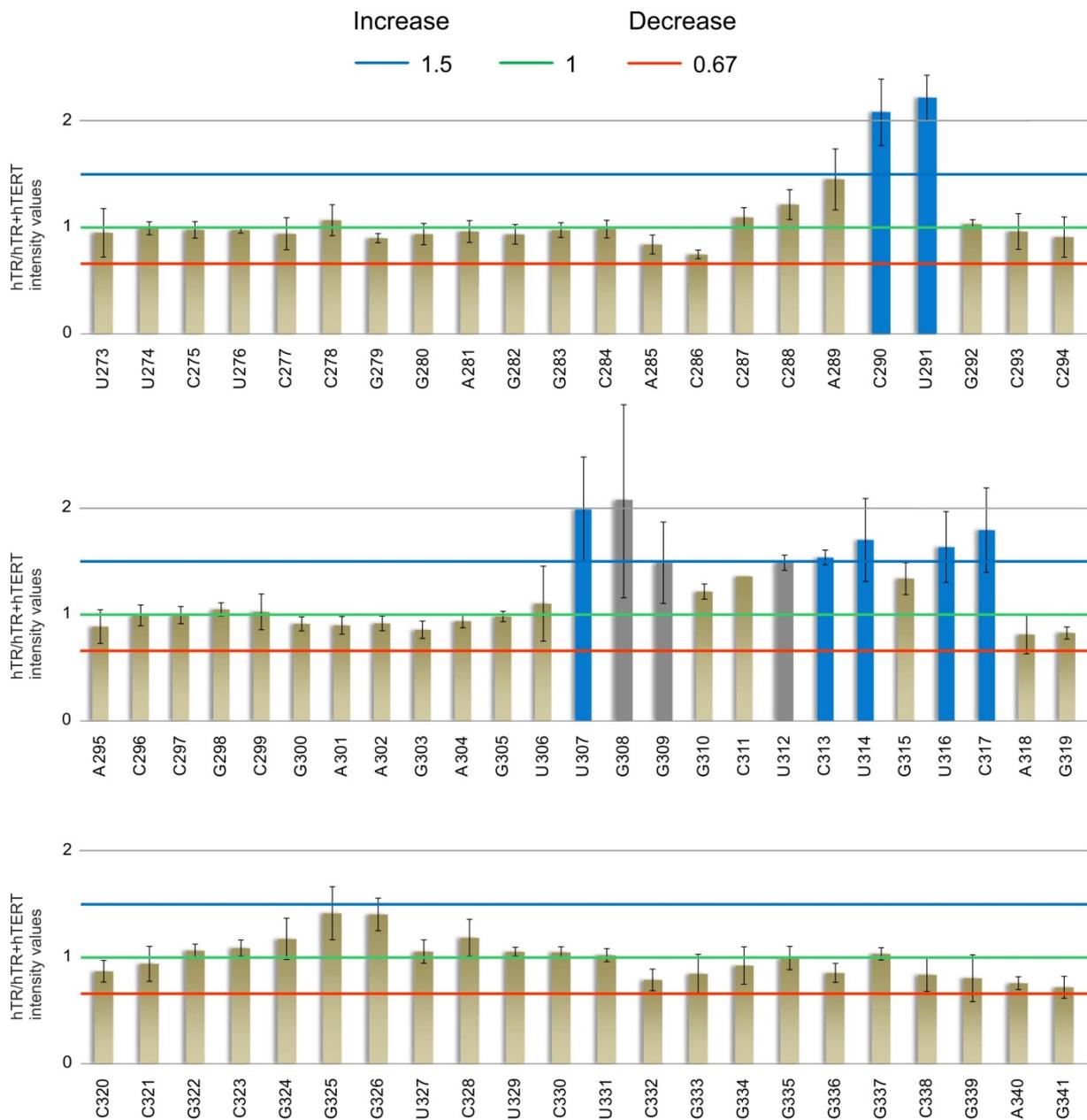
the loop is increased by conferring additional base stacking interactions ( $\Psi$ 306 stacks on G305 and  $\Psi$ 307 on  $\Psi$ 306) which in turn allows for additional hydrogen bonds to be formed<sup>42,44</sup>.



**Figure 4.11 The CR4/CR5 cross-linking pattern shows significant changes upon hTERT binding.** Representative primer extension gels showing regions **a)** P4.1, P4.2, P5, J5/6a, P6a, J6a/6b, P6b, L6b, P6b, J6b/6a, P6a, **b)** J6b/6b, P6a, P6.1, L6.1, P6.1, J6.1/5, P5, J5/4.2, P4.2 and J4.2/4.1. **c)** Secondary structure of hTR with nucleotides of the regions from **a)** and **b)** depicted in black. Neighbouring nucleotides not shown in **a)** and **b)** are shaded in gray for simplicity. Nucleotides which show changes in cross-linking intensity are depicted in blue boxes. Symbol and color code as in figure 4.2 and figure 4.7.

Notably, we have been able to only confirm that residue 307 is also a pseudouridine in the overexpressed hTR (Zemora, Handl & Waldsich, unpublished data), therefore the actual *in vivo* organization of the P6.1 stem could be an intermediate between the reported structure for P6.1 and  $\Psi$ 6.1. Aside from the change in cross-linking efficiency at U307, four novel cross-links were observed in P6.1 (C313 and U314) and J6.1/5 (U316 and C317) in the absence of hTERT (figure 4.11 b and 4.12). The significance of these residues is highlighted by the fact that the stem P6.1 consists only of nucleotides which are 100% conserved in vertebrates<sup>17,43</sup>. The nucleotides located in the junction J6.1/5 are less conserved with 90% for U316 and 80% for C317<sup>17</sup>. Furthermore, the crystal structure of the medaka CR4/CR5 domain bound to the TRBD domain of the corresponding mTERT protein revealed an interaction between the absolutely conserved J6a/6.1 A199 (A301 in hTR) and G213

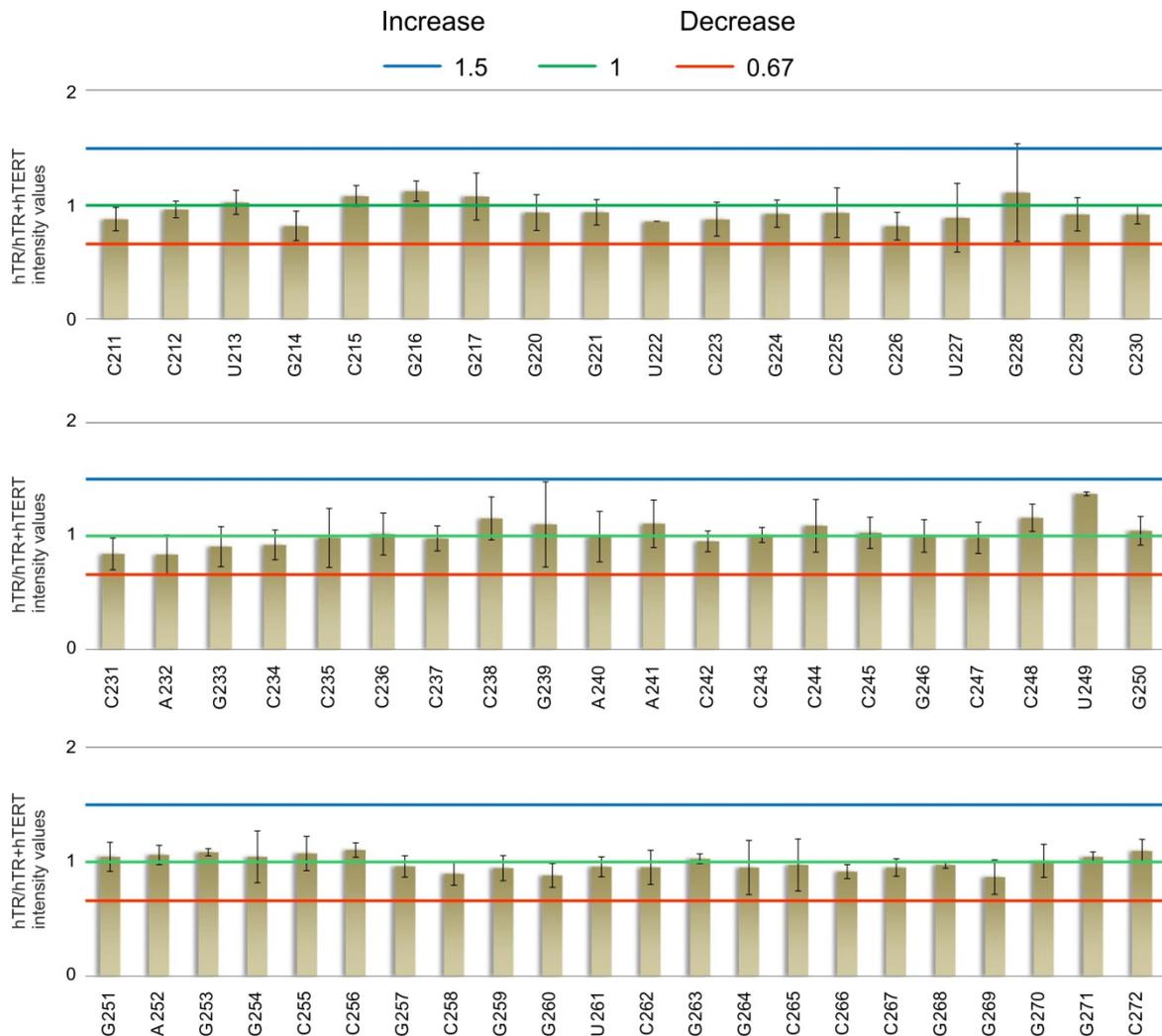
(G315 in human), the first nucleotide downstream of the P6.1 stem<sup>33</sup>. Its importance is also shown by the fact that both residues are 100% conserved among vertebrates as well as in *S. pombe* and *N. grassa*<sup>33</sup>. This specific non-canonical base pair opens up upon mTERT binding and A199 in medaka TR forms a direct contact to phenylalanine 496 and is thus a key residue for binding of mTERT<sup>33</sup>.



**Figure 4.12 UV cross-linking pattern of the CR4/CR5 domain in presence and absence of hTERT.** The bar diagram shows the normalized ratio of hTR/hTR+hTERT for nucleotides U273 to G341. The bars for nucleotides G308, G309 and U312 are strong RT stops and are therefore labelled in gray. Nucleotides which show a change in cross-linking intensity are depicted as filled blue bars. Color code as in figure 4.8.

While no cross-links were detected at A301 and G315, residues flanking G315 (C313-U314 and U316-C317) are cross-linked in the absence of hTERT (figure 4.11 and figure 4.12). Therefore the conserved

A in J6/6.1 and its interactions might represent a universal mechanism for TERT binding. Taken together, the CR4/CR5 domain shows significant changes in the cross-linking pattern in absence of hTERT. This most likely reflects the fact that the CR4/CR5 region is thought to be the main binding platform for hTERT.



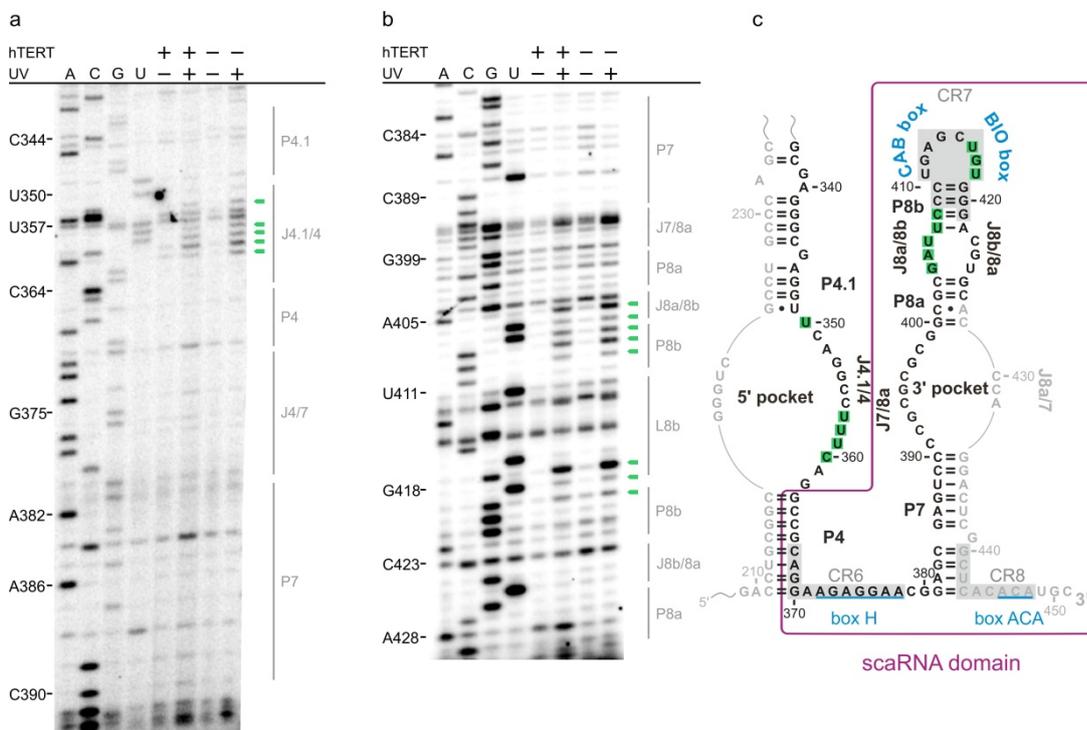
**Figure 4.13 UV cross-linking pattern of the hypervariable region in presence and absence of hTERT.** The bar diagram shows the normalized ratio of hTR/hTR+hTERT for nucleotides C211 to C272. Note that no changes in cross-linking intensity were observed. Color code as in figure 4.8.

The hypervariable region connects the pseudoknot domain with the CR4/CR5 domain (figure 1.4). This domain is made up of three stems (P4, P4.1 and P4.2) separated by internal loops showing little sequence conservation<sup>17,26</sup> and is not needed for activity *in vitro* at all<sup>13,186</sup>. For example, the 5' pocket can tolerate sequence deletions to a great extent without influencing hTR accumulation<sup>194</sup>; the same is true for the P4 stem, which was also shown to be dispensable and not affecting intracellular hTR levels as well as holoenzyme activity<sup>194</sup>. This strengthens the hypothesis that the hypervariable region serves as a physical separator holding apart the pseudoknot domain from the

CR4/CR5 domain. As such, it is not surprising that the UV cross-linking pattern in absence of hTERT is virtually the same as in the presence of hTERT (figure 4.13 and figure 4.14), suggesting that the hypervariable region is in fact not influenced by hTERT in a way that can be analysed with our experimental setup. That hTERT does not alter the conformation of the hypervariable region was also revealed by *in vivo* DMS probing<sup>178</sup>. Given the fact that the hypervariable region does neither contribute to hTERT binding nor telomerase catalysis<sup>13,194</sup> the absence of changes in cross-linking intensity is not surprising.

#### 4.3.4. The scaRNA domain is left unaltered upon hTERT binding

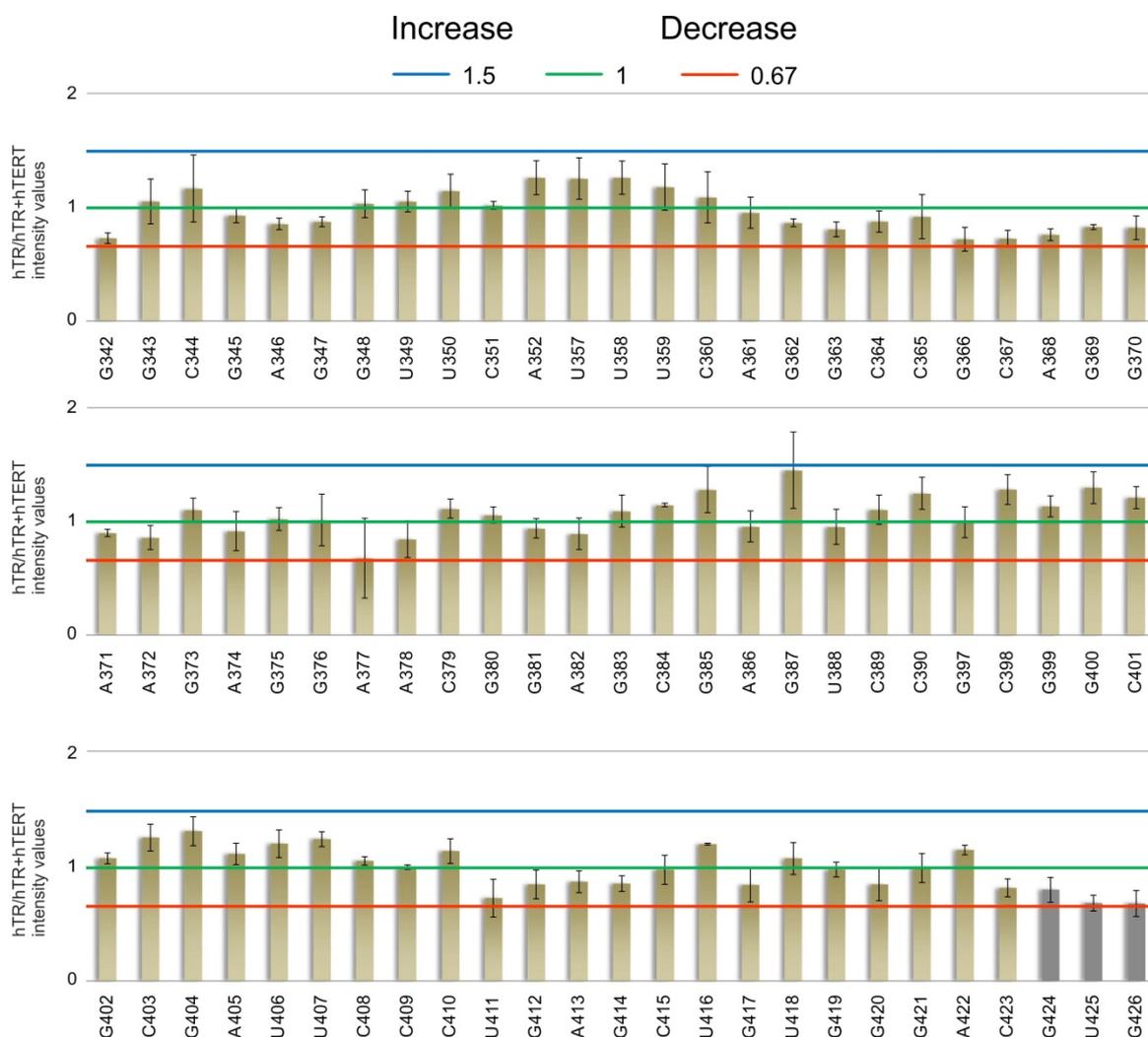
Stems P4 and P4.1 are separated by the large 5' pocket, an asymmetrical internal loop which resembles the pseudouridylation pocket of canonical H/ACA scaRNAs. However, the distance between the conserved box H and the upper end of the 5' pocket differs from the canonical length of 14 to 16 nucleotides<sup>194</sup>. Although this suggests that hTR can only accommodate a single set of H/ACA proteins (dyskerin, Gar1, Nop10 and Nhp2), pull-down studies revealed that hTR binds two sets of H/ACA snoRNA binding proteins<sup>194</sup>. The fact that loading of the H/ACA proteins on hTR is independent with regard to hTERT binding might explain as to why no changes in cross-linking intensity were found throughout the scaRNA domain (figure 4.14 and figure 4.15).



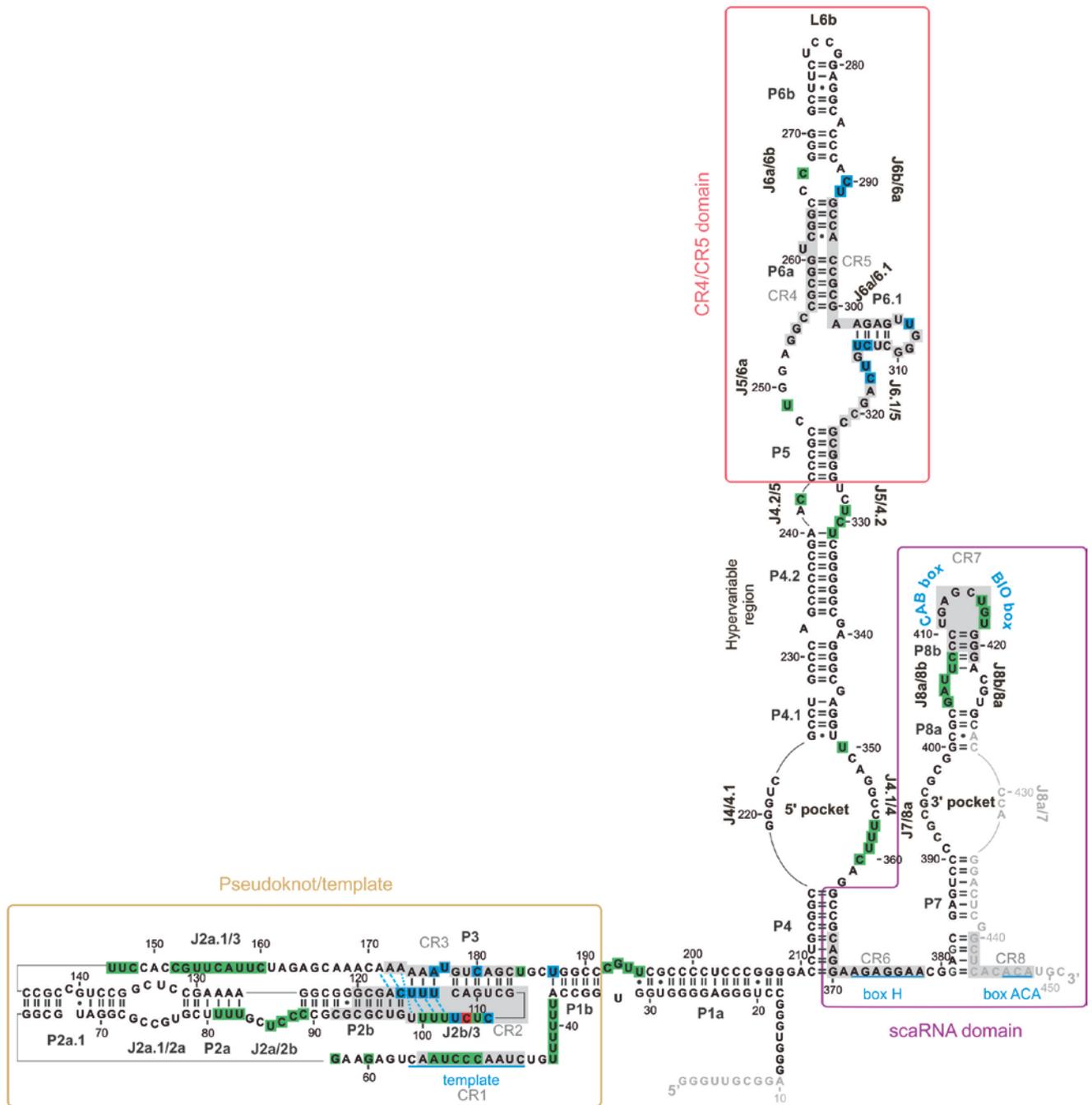
**Figure 4.14** The conserved scaRNA domain is not influenced upon hTERT binding. Representative primer extension gels showing regions **a)** P4.1, J4.1/4, P4, J4/7, P7 **b)** P7, J7/8a, P8a, J8a/8b, P8b, L8b, and J8b/8a. **c)** Secondary structure of hTR with nucleotides of the regions from **a)** and **b)** depicted in black. Neighbouring

nucleotides not shown in **a**) and **b**) are shaded in gray for simplicity. Note that no changes in cross-linking intensity are observed. Symbol and color code as in figure 4.2 and figure 4.7.

That no differences in the cross-linking pattern in L8b harbouring the conserved CAB and BIO boxes were observed (figure 4.14 and figure 4.15), is not surprising, as hTERT is not known to bind to any of these regions. Likewise, the entire scaRNA domain is associated with biogenesis and maturation of hTR, which is also reflected by the fact that telomerase activity can be reconstituted *in vitro* with the pseudoknot and the CR4/CR5 *in trans* omitting the complete scaRNA domain<sup>194</sup>. Finally, our findings are also in agreement with the data obtained from hTR DMS probing *in vivo*, revealing that hTERT does not induce conformational changes in the scaRNA domain of hTR<sup>178</sup>.



**Figure 4.15 UV cross-linking pattern of the scaRNA domain in presence and absence of hTERT.** Bar diagram showing the normalized ratio of TR alone/TR+hTERT for nucleotides G342 to A428 (shown on the gels in figure 4.14, a and b). Note that nucleotides G353 to C356 and C391 to C396 are not represented in the chart. These nucleotides could not be separated to single nucleotide resolution making it impossible to assign an intensity value. Color code as in figure 4.8.



**Figure 4.16** hTERT binding induces changes in the pseudoknot as well as in the CR4/CR5 domain. Differential map showing nucleotides whose UV cross-linking intensities change in an hTERT-dependent manner. Nucleotides with increasing cross-linking intensity are boxed in blue, whereas a decrease in cross-linking intensity is depicted by a red box. Cross-links which do not change are boxed in green. Symbol and color code as in figure 4.6.

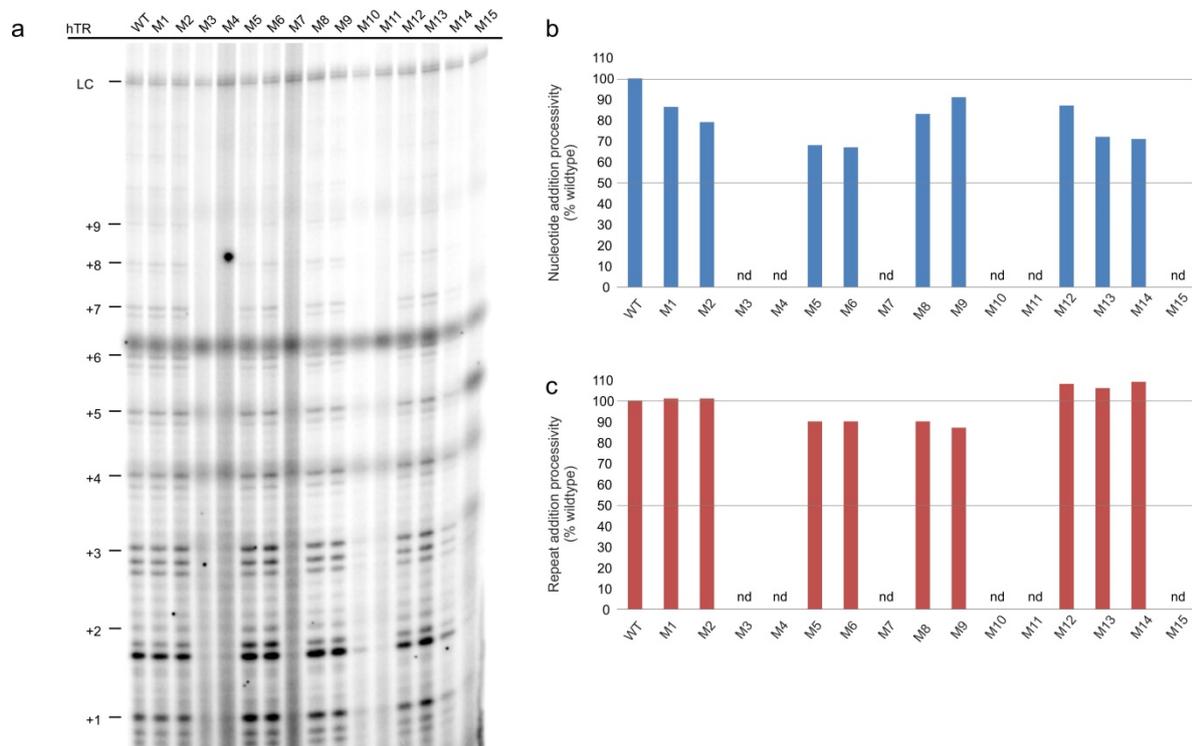
#### 4.4. Identification of intramolecular RNA-RNA cross-links

Based on the UV cross-linking pattern obtained both in presence and absence of hTERT, we set out to identify corresponding cross-linking partners. This allows us to obtain spatial constraints for the conformation of human telomerase RNA and might provide new insights into telomerase organization. In order to achieve this goal hTR mutants with base changes at the sites of the previously identified cross-linked nucleotides were created. Most of the cross-linked nucleotides are located in the pseudoknot region and the CR4/CR5 domain of human telomerase RNA. Overall, we have selected seven sites of cross-linked residues in the pseudoknot/template domain, four in the CR4/CR5 domain and two each in the hypervariable region as well as in the scaRNA domain (figure 4.17). The desired mutations were introduced and the obtained plasmids were used for co-transfection of HEK293 cells according to our established experimental setup (figure 4.1). By mutating the nucleotides which were cross-linked in our initial experiments, we expect to observe a significant change in intensity or loss of the cross-link at the respective partner, thus suggesting that these nucleotides must be in close proximity. Notably, all point mutations were transitions and not transversions (table 3-3). With the majority of the cross-links being located in single-stranded regions we were less concerned about potential distortions in the helical regions, furthermore almost the entirety of cross-links found in helices occur at U residues of canonical A-U base pairs and the substitution of a U to C would therefore induce a A•C base pair which has only minor effect on helix topology<sup>195</sup>.

So far 15 mutants within important cross-linking regions were created (figure 4.17 and table 3-3). Interestingly, our PCR-based approach for introducing the mutations of choice was not successful for three areas of cross-links. The three uridines at the 3' end of P2a (U81 to U83), the nucleotides of J1b/1a (C192 to U195) and the last nucleotide of P4.2 together with two nucleotides from J5/4.2 (U329 to C332), all of them gave positive PCR results and accurate lengths however subsequent sequencing showed that repetitive sequences had been introduced instead of the chosen mutations. For all three sites, several primer pairs with different characteristics (e.g.  $T_M$ , length, binding position, and GC content) as well as altered PCR programs have been tested, however no positive clone could be obtained. The reasons are most probably to be found in adjacent areas consisting of repeats of a single base and are therefore detrimental to primer binding. The cross-linking pattern of the 15 successfully generated hTR mutants were mapped in complex with hTERT. First, the mutant hTR variants were tested for their effect on telomerase activity (figure 4.18). Therefore, cell lysates of each construct were incubated with a telomere mimicking oligo which can be extended if a functional telomerase complex was able to form.



activity. The remaining mutants had slightly decreased telomerase activity compared to wildtype (figure 4.18)

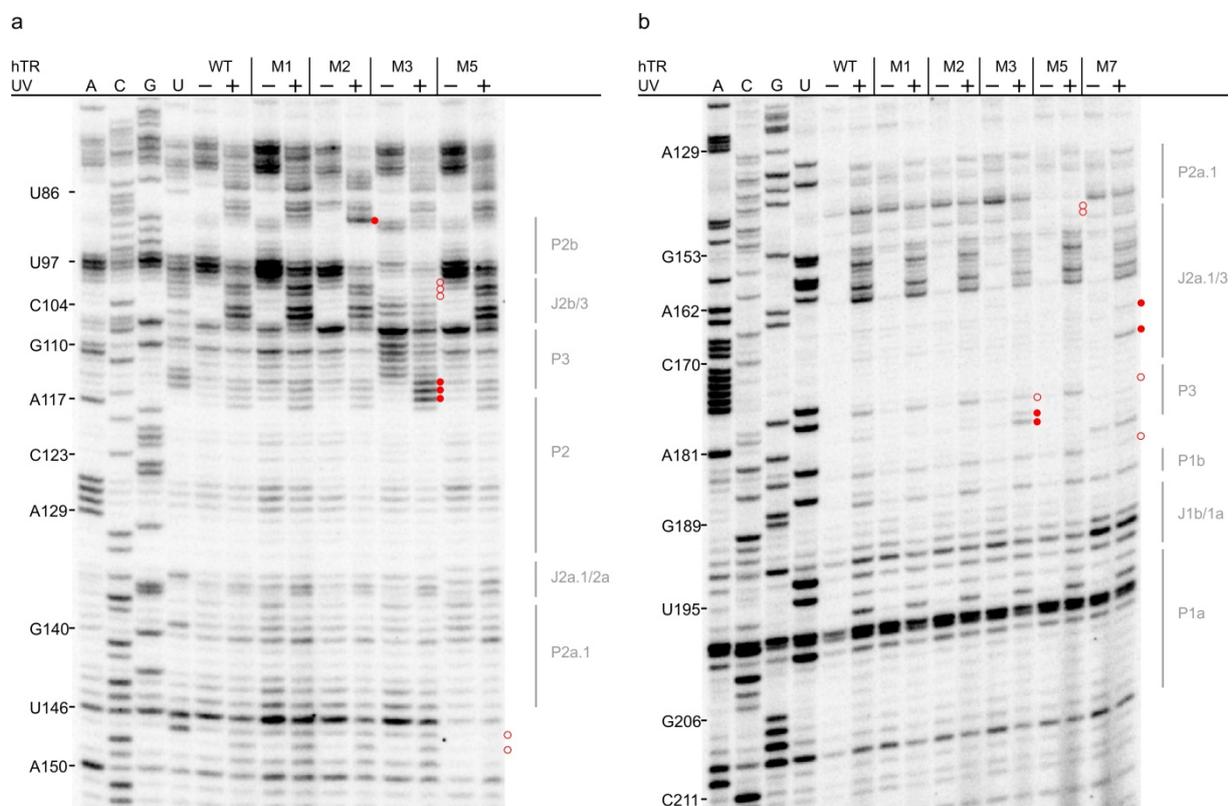


**Figure 4.18 Telomerase activity of hTR mutants.** All available hTR mutants were tested for telomerase activity. A primer mimicking the telomere can be elongated by a functional telomerase complex. **a)** Representative gel showing the direct telomerase assay. Roman numerals on the left side depict the number of repeats added to the primer. LC stands for a  $^{32}\text{P}$ -labelled 100mer which is used as loading control. Quantification of **b)** NAP and **c)** RAP for every single mutant. Mutants for which no numerical value could be determined are labelled nd.

#### 4.4.1. The structural organization of the pseudoknot domain

The pseudoknot and the conserved triple helix contained the majority of cross-linked nucleotides and showed strong changes in intensity when hTERT is absent. Mutants M1 to M7 are located in the pseudoknot domain of human telomerase RNA. By comparing the UV cross-linking pattern of each of these mutants to the wildtype pattern, we have identified several changes in cross-linking intensity. Mutant M1 (U39C/U40C/U41C/U42C/U43C) shows similar telomerase activity compared to wildtype (figure 4.18) and, interestingly, new RT stops at nucleotides U113, U114 and U115 (figure 4.19). These residues are part of the base triples and therefore crucial for the overall topology of the pseudoknot domain. The loss of these cross-links might indicate that the template region is positioned in close proximity of the triple helix. Given the fact, that residues located in the triple helix, like A176 and U177, were shown to play a role in catalysis<sup>29</sup>, close contacts between the template and the pseudoknot seem highly likely. Changes in cross-linking intensity were also

observed at nucleotides U329 and C330 in J5/4.2 in the CR4/CR5 domain (figure 4.20 a), while the remainder of the nucleotides did not change. Construct M2, containing the mutations U86C/C87U/C88U/C89U, performed similar to wildtype telomerase, indicating that the mutations introduced do not affect telomerase activity. The observed changes in cross-linking intensity were only minor, which is surprising, given the importance of J2a/2b for the overall topology of the pseudoknot<sup>28</sup>. We report an increase of intensity at nucleotide C90, which is adjacent to the site of mutation. This, and the fact that we do not observe other changes in the cross-linking pattern throughout the molecule (figure 4.20) might indicate that the initial cross-link in this area occurs between the nucleotides in the J2a/2b due to their unique conformation<sup>28</sup>. For mutants M3 and M4, both containing changes to nucleotides in the triple helix, telomerase activity was not detectable (figure 4.18). However, this is not surprising, as it was shown that the formation of the triple helix is crucial for telomerase activity<sup>23,29</sup>. From all constructs analysed M3 (U100C/U101C/U102C/U103C/C104U/U105C) gave the most interesting cross-linking pattern. The initial cross-links in J2b/3 were completely absent, and natural RT stops for almost every single nucleotide in helix P3 appeared (figure 4.19 a). The cross-linking intensity at nucleotides U113 to U115 as well as U179 and C180 increased while the intensity at U177 decreased compared to wildtype (figure 4.19 b). The remainder of the nucleotides did not change (figure 4.20).

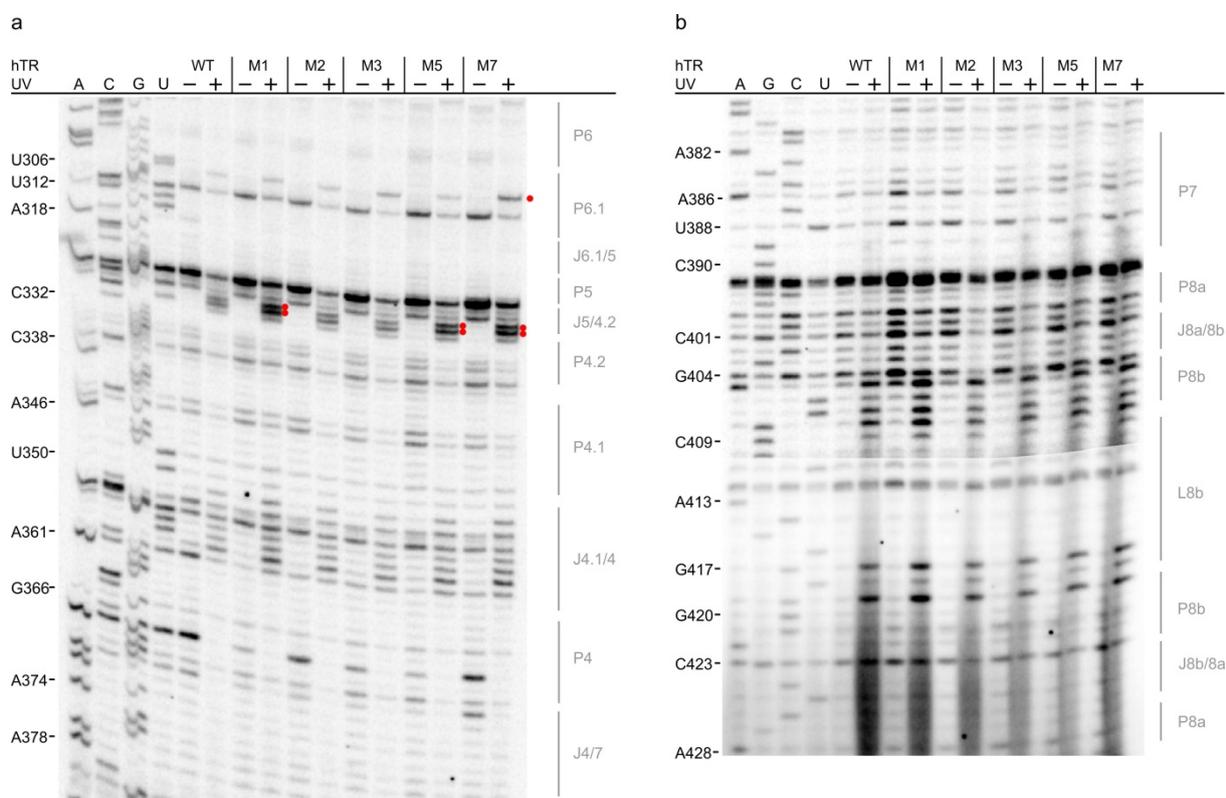


**Figure 4.19 The core pseudoknot changes in the mutant hTR variants.** Representative primer extension gel for mutants a) M1, M2, M3 and M5 showing P2, J2b/3, P3, P2, P2a.1 and J2a.1/3 b) M1, M2, M3, M5 and M7 showing P2a.1, J2a.1/3, P3, P1b, J1b/1a and P1a. Changes in the UV cross-linking pattern compared to wildtype

are indicated by filled red circles for increased intensity and open red circles for decreased intensity. Note that M7 is not shown in **a**) due to the mutation being within the primer binding site. Symbol and color code as in figure 4.2.

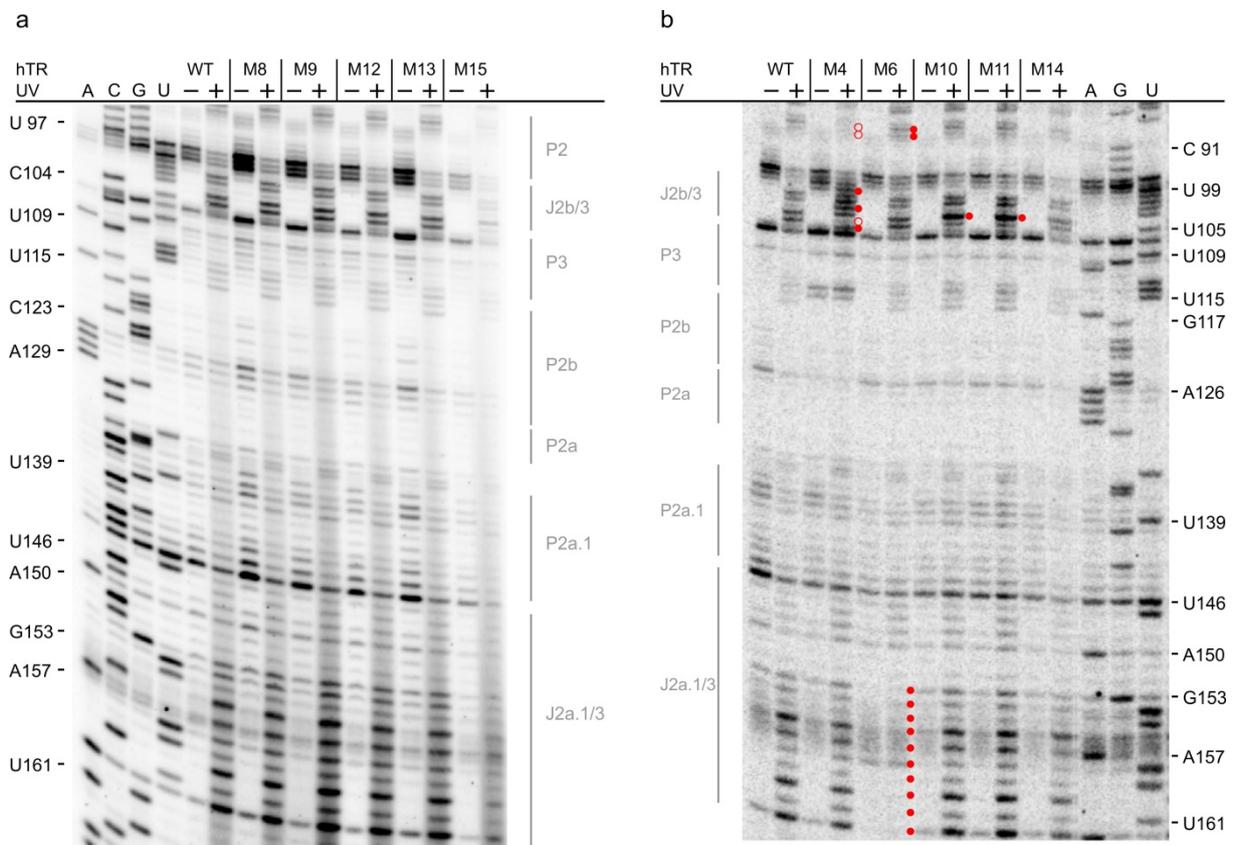
Since the M3 mutant is inactive, the increase in intensity may reflect either an enhanced propensity for cross-linking due to the fact that hTERT may not be bound to the M3 hTR variant or structural rearrangements affecting the physical proximity of respective residues.

A similar result was observed for M4 (U113C/U114C/U115C/C116U). Additional RT stops appeared at the sites of mutation at U114 and U115. Furthermore, a completely different cross-linking pattern was observed in the J2b/3 linker connecting stems P2b and P3 (figure 4.21 b). A new cross-link is observed at position U99, while the cross-link at C104 disappears. In addition, an increase in intensity is observed for U102 and U105, while U100, U101 and U103 do not change, compared to the wildtype pattern. In line with the disruption of stem P3 and associated base triples, M4 is catalytically inactive (figure 4.18). The absence of the triple helix in M4 might provide additional flexibility to the molecule and thereby altering the cross-linking pattern in J2b/3. Together with our observation that C104's cross-linking intensity drastically decreases in absence of hTERT, this hints to a role of C104 in hTERT binding.



**Figure 4.20 J5/4.2 and P6.1 in the CR4/CR5 domain show changes in cross-linking intensities.** Representative primer extension gel for mutants M1, M2, M3, M5 and M7 showing **a**) P6, P6.1, J6.1/5, P5, J5/4.2, P4.2, P4.1, J4.1/4 and P4 **b**) P7, J7/8a, P8a, J8a/8b, P8b, L8b and J8b/8a. Changes in the UV cross-

linking pattern compared to wildtype are indicated by filled red circles for increased intensity and open red circles for decreased intensity. Note that for **b**) nucleotides U411 to A428 are shown with different contrast setting. Symbol and color code as in figure 4.2.

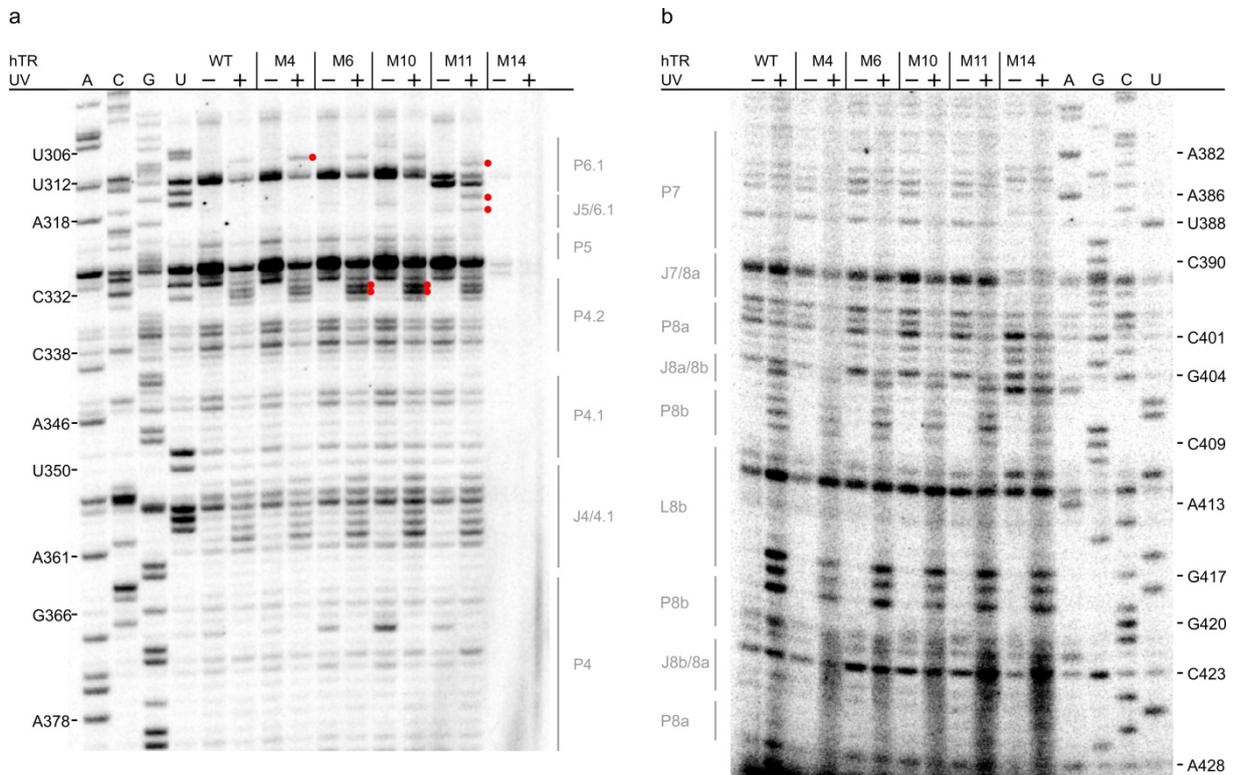


**Figure 4.21 Linker J2b/3 and J2a.1/3 show different cross-linking pattern.** Representative primer extension gel for mutants **a**) M8, M9, M12, M13 and M15 showing P2, J2b/3, P3, P2b, P2a, P2a.1 and J2a.1/3 **b**) M4, M6, M10, M11 and M14 showing J2b/3, P3, P2b, P2a, P2a.1 and J2a.1/3. Changes in the UV cross-linking pattern compared to wildtype are indicated by filled red circles for increased intensity and open red circles for decreased intensity. Symbol and color code as in figure 4.2.

A decrease in intensity is also observed for the cross-links at U86, C87, C88 and C89 in J2a/2b, indicating that the S-shape of the backbone of J2a/2b might be distorted. Finally, the cross-link at U307 in the CR4/CR5 domain also shows an increase in intensity (figure 4.22 a). This is interesting with regard to the similar increase of intensity at U307 which was observed when hTR was cross-linked in absence of hTERT (figure 4.11). For the remainder of the nucleotides, no changes were observed (figure 4.22). Nonetheless, additional experiments are needed to examine the exact role of the involved residues more carefully.

The last construct with mutations in the triple helix is mutant M7. Here, nucleotides in P3 have been changed (A176G/U177C/C180U/U184C), which led to changes in the cross-linking pattern (figure 4.19 b) and abolished telomerase activity (figure 4.18). First, two new cross-links emerge at positions

C166 and C170 in the J2a.1/3 linker while the cross-links at the mutated residues A176, U177, C180 and U184 disappear (figure 4.19 b). This potentially indicates increased flexibility of residues C166 and C170 in J2a.1/3 due to the mutation in stem P3.



**Figure 4.22 The CR4/CR5 domain shows changes in the cross-linking pattern.** Representative primer extension gel for mutants M4, M6, M10, M11 and M14 showing **a)** P6.1, J5/6.1, P5, P4.2, P4.1, J4/4.1 and P4 **b)** P7, J7/8a, P8a, J8a/8b, P8b, L8b, P8b, J8b/8a and P8a. Changes in the UV cross-linking pattern compared to wildtype are indicated by filled red circles for increased intensity and open red circles for decreased intensity. Note that M14 gave only a weak signal due to mutations at the primer binding site. Symbol and color code as in figure 4.2.

As no other cross-links disappeared, hTERT might be the cross-linking partner of A176, U177, C180 and/or U184. In addition, M7 shows an increase in cross-linking intensity at nucleotides U307 in P6.1 and U329 as well as C330 in J5/4.2 (figure 4.20 a). The increase in intensity at U307 is similar to mutant M4 and to the wildtype cross-linking pattern observed in absence of hTERT. For the remaining nucleotides of hTR variant M4 no further changes in the cross-linking pattern were observed (figure 4.20). The last two constructs on the pseudoknot domain are M5 and M6, both containing mutations in the J2a.1/3 linker connecting the extended pseudoknot with the core pseudoknot (U146A/U147A and C152G/G153U/U154A/U155A/C156G/A157U/U158A/U159A/C160G, respectively). Both constructs show a slightly reduced telomerase activity level compared to wildtype and in both cases, the initial cross-links disappear (figure 4.21 and figure 4.19), and new cross-links appear. In both hTR variants M5 and M6 an increase in cross-linking efficiency at nucleotides U329

and C330 in J5/4.2 was observed, while only for M6 an increase in cross-linking intensity was discovered at nucleotides C87 and C88 in J2a/2b. This might indicate that J2a.1/3 influences J5/4.2 in an unknown way. The loss of the initial cross-links might be explained by the fact that both, M5 and M6, contain transversions instead of transitions. Therefore, the decrease of pyrimidine residues might have prevented the formation of cross-links between neighbouring nucleotides. For the remainder of the nucleotides in M5 and M6 no changes in the cross-linking pattern were observed.

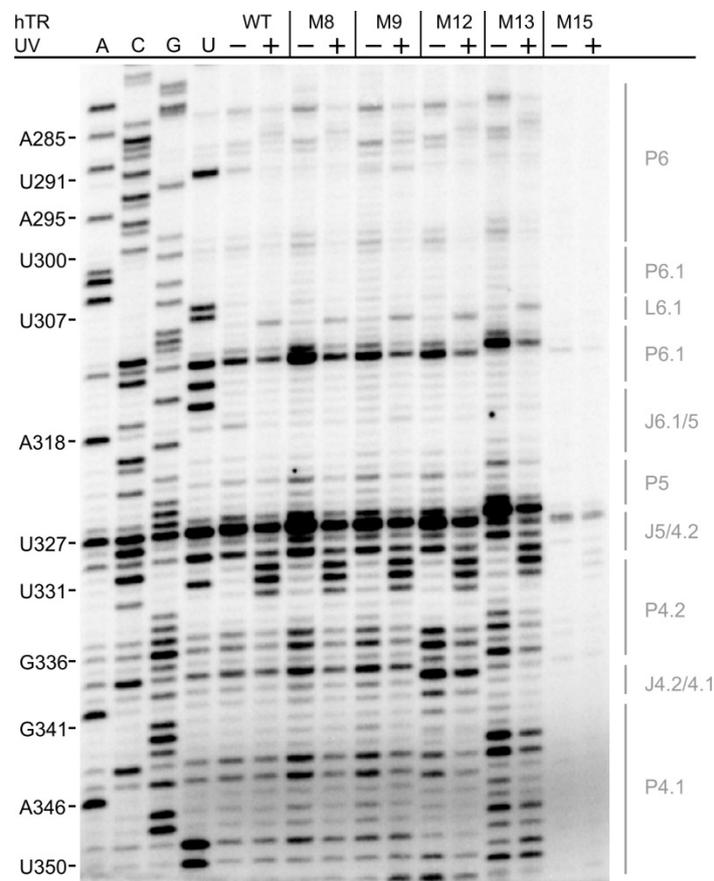
#### **4.4.2. The topology of the important CR4/CR5 domain**

The region with the second-most cross-links is the CR4/CR5 domain of the human telomerase RNA. Four constructs have been created for this element. M8 (U249C) and M9 (C290U/U291C) show both normal telomerase activity compared to wildtype. For both, M8 and M9, no changes in the cross-linking pattern compared to wildtype were observed (figure 4.21 and figure 4.23). The fact that for both constructs telomerase activity is similar to wildtype, might indicate that hTERT is, however, still bound to the hTR variants M8 and M9.

For the constructs M10 (U307C) and M11 (C313U/U314C/U316C/C317U) containing mutations of the highly conserved residues in P6.1, the situation is entirely different. First, both of these constructs show no telomerase activity, neither NAP nor RAP (figure 4.18). Second, M11 gives rise to new cross-links mostly at the positions of the mutated residues (C314 and C316). Interestingly, a novel cross-link appears at position G308 in the terminal loop L6.1 (figure 4.22), together with a new RT stop at position U313. Importantly, G308 is strongly methylated by DMS *in vivo* in the absence of hTERT, but becomes protected upon binding of hTERT<sup>178</sup>. Together, these data may suggest that hTERT binding is abolished in the mutant M11, explaining the loss of activity of this mutant as well. In fact, the observed pattern resembles that obtained in absence of hTERT (figure 4.11 b). While both M10 and M11 do not show an altered UV cross-linking pattern for most of the molecule, one striking similarity is observed for both of them: a strong increase in cross-linking intensity at position U103 in the pseudoknot region (figure 4.21 b). This would place U307 close to the essential triple helix for catalysis. In addition, for M10 an increase in cross-linking efficiency at nucleotides U329 and C330 in J5/4.2 was observed (figure 4.22).

Mutants M12 (U350C) and M13 (U357C/U358C/U359C/C360U) are located in the hypervariable region of hTR, in more detail in the 5' pocket preceding the scaRNA domain. Interestingly, no change in the cross-linking pattern compared to wildtype is observed for both M12 and M13 (figure 4.21 b), suggesting that these cross-links might be directed to H/ACA snoRNA binding proteins, which bind to the 5' pocket of hTR<sup>19</sup>. In addition, for M13 an interesting phenomenon in the CR4/CR5 domain (figure 4.23) was observed, namely the cross-links are shifted by one nucleotide compared to

wildtype and the rest of the mutants. Interestingly, this is not the case for remainder of the nucleotides. The primer binding site and the mutated residues in M12 and M13 are separated by 50 and 60 nucleotides, respectively, thus making problems during primer annealing unlikely. However, the reason for the shift remains yet unclear.

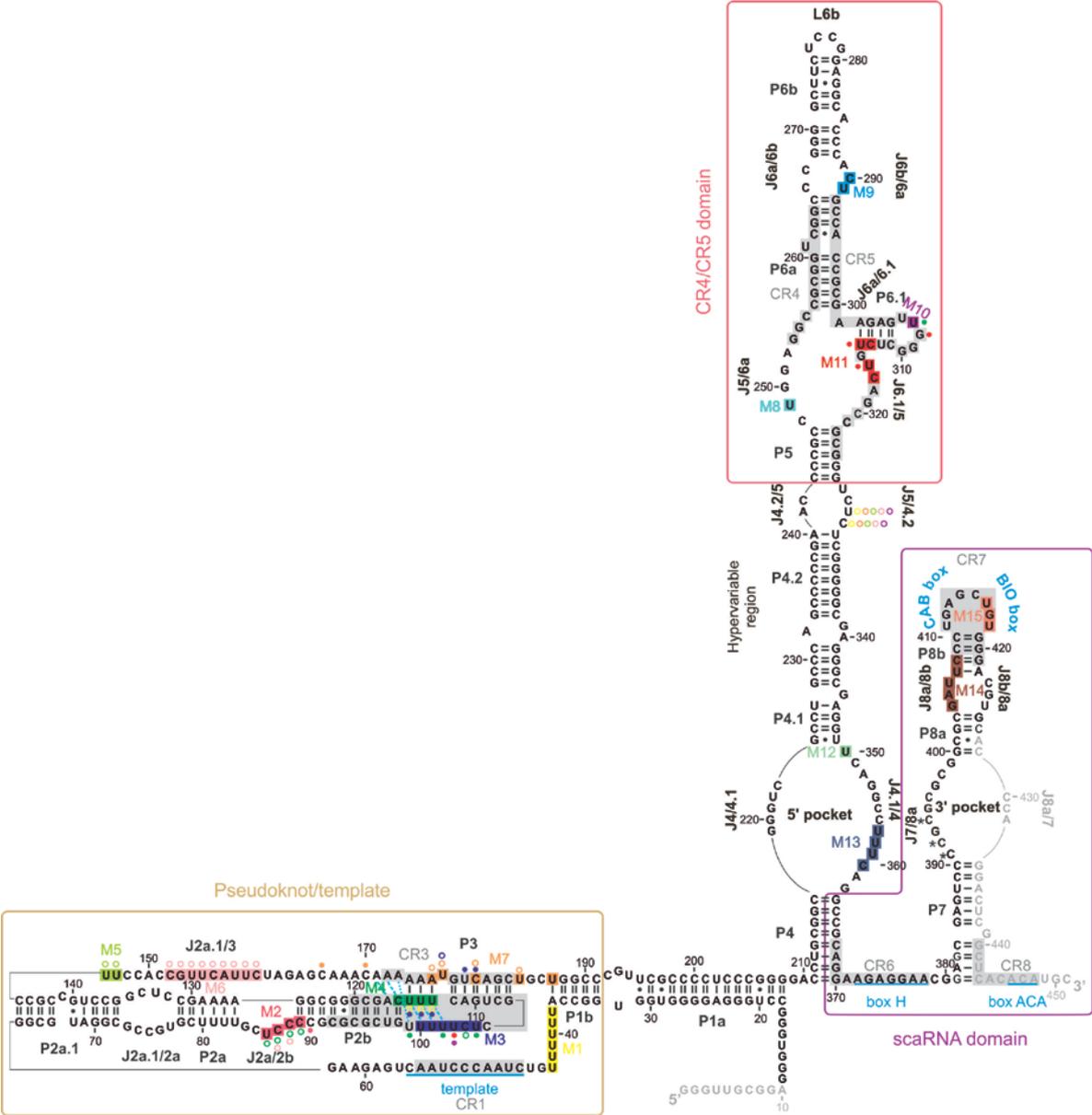


**Figure 4.23 The CR4/CR5 domain remains unaffected.** Representative primer extension gel for mutants M8, M9, M12, M13 and M15 showing P6, P6.1, L6.1, P6.1, J5/6.1, P5, J5/4.2, P4.2, J4.2/4.1 and P4.1. Changes in the UV cross-linking pattern compared to wildtype are indicated by filled red circles for increased intensity and open red circles for decreased intensity. Symbol and color code as in figure 4.2.

#### 4.4.3. The conformation of the scaRNA domain

The remaining two hTR variants are part of the scaRNA domain. M14 contains mutations in J8a/8b as well as in stem P8b (G404A/A405G/U406C/U407C/C408U) while M15 carries the mutated BIO box (U416C/G417A/U418C). The constructs differ significantly in their telomerase activity. M14 shows a slightly reduced activity compared to wildtype, however M15 is catalytically inactive (figure 4.18). Furthermore, the signal strength for mutant M15 is drastically reduced throughout all primer extension gels (figure 4.21 and figure 4.23). This observation is consistent with data from an *in vivo* DMS study, which as well reported a large decrease in signal strength when studying the C408G mutant<sup>178</sup>. The reason is likely to be found in impaired biogenesis of hTR and subsequent degradation

of the precursor molecule, resulting in extremely low levels of cellular hTR. This also explains the lack of telomerase activity observed for this mutant. For M14, signal strength was comparable to wildtype levels; however the cross-linking pattern showed no differences (figure 4.21 and figure 4.22). A possible explanation for this observation might be found in the H/ACA snoRNA binding proteins, which are supposed to bind within the scaRNA domain. Therefore, one of these proteins (Gar1, Nop10, Nhp2 and Dyskerin) could be the corresponding cross-linking partner in wildtype telomerase RNA and M14, suggesting a sequence-independent interaction.



**Figure 4.24 Spatial information for hTR obtained from UV cross-linking.** Secondary structure map of human telomerase RNA depicting elements in proximity to each other. The hTR variants are depicted as boxed nucleotides in different colors. An increase in cross-linking intensity is shown as a filled circle next to the nucleotide, while a decrease in intensity is depicted by an open circle. The color of the circles corresponds to the hTR variants which are responsible for the observed changes. Symbol and color code as in figure 1.4.

## 5. Discussion

The maintenance and protection of the cellular telomeres is an essential requirement for most living organism. The evolutionary most common mechanism to achieve this goal is the use of a specialized RNP complex referred to as telomerase<sup>196</sup>. The telomerase holoenzyme is a RNA-dependent DNA polymerase which catalyzes the addition of repeating DNA sequences to the telomeres. Human telomerase consists of two major compounds: first, the human telomerase reverse transcriptase, called hTERT and second, the human telomerase RNA (hTR), which contains the essential RNA template in order to add the 5' GGTTAG 3' repeats<sup>9,17</sup>. Human telomerase RNA must adopt a specific conformation in order to successfully serve as platform for assembling the mature telomerase complex. We aimed to shed more light onto the structural organization of hTR and the interplay between hTR and hTERT. Therefore, we have established an *in vivo* UV cross-linking setup allowing us to investigate which elements within hTR are in close contact and how the conformation of hTR changes upon binding of its protein partner hTERT. In order to compensate for the low endogenous hTR levels within HEK293 cells, we transiently co-transfected plasmids coding for either component, hTR and hTERT. We then checked for telomerase activity via direct telomerase activity assays in cell lysates to exclude any contribution of endogenous hTR or hTERT levels. Indeed, only upon overexpression of hTR and hTERT, the telomerase complexes present in the lysates efficiently elongated a primer mimicking the telomeres, thus ensuring that our system works well. In addition it appears that the majority of hTR molecules adopt their native conformation in the cell based on the data available from DMS footprinting<sup>178</sup>. We then performed primer extension on UV cross-linked hTR to identify the cross-linked nucleotides. By comparing the cross-linking pattern between wildtype (hTR and hTERT present) and an hTR-only setting, it was possible to assess the conformational changes of hTR upon hTERT binding. Our results provide new information about the structural organization of human telomerase RNA and hTERT-induced changes in the hTR architecture.

### 5.1. The structural organization of hTR in vivo

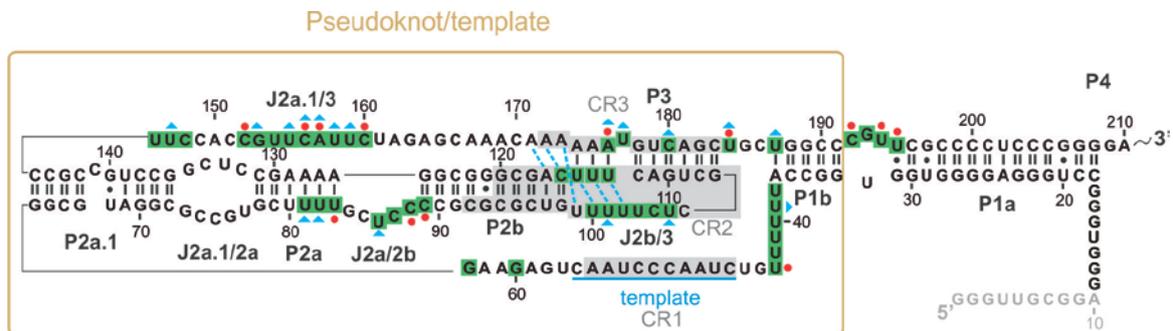
Human telomerase RNA can be divided into three separate domains: the pseudoknot domain containing the template is located in the 5' half; the CR4/CR5 domain is located in the middle of the 451 nucleotide-long RNA while the scaRNAs domain is found within the 3' part of the molecule (figure 1.4). The pseudoknot domain also contains the template which is necessary to for repeat synthesis and is flanked by the template boundary element (helix P1b)<sup>17,20</sup>. Telomerase needs this element to exert its task without read-through errors or miss-incorporation of non-canonical template nucleotides<sup>185</sup>. The linker between P1b and the template is strongly conserved in length but not in sequence<sup>185</sup>. The cross-links we observed at nucleotides U38 to U43 could therefore be related

to the task of template boundary definition (figure 4.3). However, neither *in vivo* DMS nor *in vitro* SHAPE probing for these nucleotides report high accessibility or flexibility, respectively<sup>178,197</sup>. Therefore the possibility that the cross-links were simply caused by several cyclobutane pyrimidine dimers (CPD) forming via their C5-C6 bonds must also be taken into consideration. In line with the *in vivo* DMS data, we observed cross-links at the template nucleotides C50 to A54. These nucleotides are also responsible for template alignment, therefore forming the DNA-RNA duplex before elongation takes place<sup>91,186</sup>. The observed cross-links might therefore represent a structure that is crucial for the formation of the DNA-RNA duplex. For the 3' flanking sequence of the template, fewer cross-links are observed; most probably due to the purine residues, which is in good agreement with the reported fact of decreased cross-linking efficiency for purines<sup>175,176,198</sup> (figure 4.6).

### 5.1.1. The conformation of the pseudoknot domain *in vivo*

The pseudoknot domain of human telomerase RNA contains several unique features which strongly contribute to the topology of this unique pseudoknot. A remarkable network of Hoogsteen base pairs which lead to the formation of a triple helix<sup>25</sup> and the strong bend in the backbone due to the unconventional conformation of junction J2a/2b<sup>28</sup> are just two examples. The observed cross-links on U81, U82 and U83 next to J2a/2b could be explained by the fact that the respective U residues do not form U-A Watson-Crick base pairs. Importantly, A126-A128 are also accessible to DMS modification, suggesting that the respective U and A residues form non-Watson-Crick U•A pairs *in vivo*<sup>178</sup>. Furthermore, *in vitro* SHAPE analysis also indicated a backbone flexibility higher than observed for canonical base pairs<sup>197</sup>. The nucleotides within the loop itself are all exposed to the solvent with their Watson-Crick face, a fact supported up by the strong DMS modifications on C87 and C88<sup>178</sup>. This suggests that the cross-links at C87 and C88 are formed without involving of their N3 atoms. When also taking the SHAPE data into account, which shows average backbone flexibility for C87 to C89<sup>197</sup>, this points to a very rigid nature of residues within J2a/2b, although the interaction partners are not yet known. A second hot spot for UV-induced cross-links is found in the pyrimidine-rich stretch of the linker J2a.1/3 which correlates well with the reported SHAPE data which shows the highest backbone flexibility for the nucleotides in this area<sup>197</sup>. In general, the majority of cross-links were observed at nucleotides that are modified by DMS *in vivo* and SHAPE *in vitro*<sup>178,197</sup>(figure 5.1). Taken together, our results reflect the importance of these residues for the overall hTR topology. Notable exceptions are nucleotides U113, U114, U115 and C116 in stem P2b and U100 to U102 in J2b/3 which form the triple helix in the pseudoknot domain. None of them is modified by DMS and backbone flexibility is also low in this area (figure 5.1). According to the solution structure of the pseudoknot, the triple helix is the predominant structure for this area<sup>23,25</sup>. Earlier studies proposed a conformational switch between the pseudoknot and a hairpin structure which might affect telomerase translocation<sup>199,200</sup>.

In this hairpin residues of J2b/3 (U99 to C104) extend the P2b stem by pairing with nucleotides G110 to U115, while U105 to U109 would form a tetraloop at the end of the extended P2b stem.



**Figure 5.1 Chemical modifications observed in the pseudoknot domain.** Secondary structure map showing the pseudoknot domain of hTR. Sites of UV-induced cross-links are boxed in green, while DMS modifications are shown as red circles. Nucleotides which show increased reactivity to SHAPE are labelled with blue triangles. For simplicity residues which are reactive to DMS and/or SHAPE but have not been cross-linked are not shown. Symbol and color code as in figure 4.6. DMS and SHAPE data are taken from Zemora, 2014<sup>178</sup> and Niederer et al., 2014<sup>197</sup>.

Thus, such a hairpin leads to the disruption of the Hoogsteen base triples U115-A174•U100, U114-A175•U101 and U113-A176•U102 which severely decreases telomerase activity<sup>23,25</sup>. This proposed model is in line with older *in organello* DMS footprinting of hTR which reported that the pseudoknot is not formed<sup>201</sup>. However, *in vivo* DMS footprinting data from our lab decisively show that all stems important for pseudoknot as well as the triple helix formation are stably formed *in vivo*<sup>178</sup>. The differences between these two studies might be the result of two different experimental setups: Antal et al. extracted cell nuclei and performed DMS modification subsequently, while we performed the DMS modification on a mono layer of cells followed by hTR isolation. The fact that UV cross-links at nucleotides which are involved in the Hoogsteen base triples are observed, further hints to formation of the pseudoknot and the associated triple helix *in vivo*. As such, the cross-links are likely to go to the corresponding interaction partner of the base triples.

### 5.1.2. The CR4/CR5 domain harbours only few cross-links

The second important domain besides the pseudoknot/template domain is the CR4/CR5 domain. *In vitro* complementation assays showed that telomerase activity can be reconstituted with the pseudoknot and the CR4/CR5 domain *in trans*<sup>186</sup>. This suggests that the connecting hypervariable domain acts as a linker in the first place and is not involved in telomerase catalytic activity. This is also in line with the high tolerance for mutations within the 5' pocket that do not affect telomerase activity and accumulation<sup>19,194</sup>. Thus, it is possible that the cross-links observed in the 5' pocket might be directed to additional proteins such as the H/ACA RNA binding proteins like Gar1, Nop10, Nhp2

and Dyskerin. This is also supported by the fact that the nucleotides on the opposite strand, J4/4.1, of the 5' pocket, are not cross-linked, therefore reducing the possibility of a hydrogen bond network spanning throughout the pocket. For the smaller internal loop separating the last stem P4.2 in the hypervariable region from the first stem P5 of the CR4/CR5 domain, cross-links are found on both sides of the loop. Such an observation might be indicative of intra-loop contacts rather than long-range tertiary contacts. A conformation like this could confer a specific backbone topology necessary for the domain to carry out its function. Examples for such specific junctions which shape the topology of an entire domain are J2a/2b and the internal loop separating P6a and P6b. While J2a/2b in the pseudoknot domain introduces a sharp bend between the two connecting stems, the P6a-P6b internal loop alters the backbone geometry in way resembling the Greek letter  $\zeta$ <sup>187</sup>. Nucleotide deletions in the P6a-P6b loop resulted both in a strong decrease of telomerase activity as well as lower binding affinity to hTERT<sup>13</sup>. Interestingly, the UCG tetraloop at the end of the CR4/CR5 domain and its 5' flanking nucleotides are also composed of only pyrimidines, however none of these residues was cross-linked in our study (figure 4.4 a and c). This example and other non-cross-linked pyrimidine stretches (stems P1a, P4.1 and P4.2) are a good indicator that the identified cross-links are not only cyclobutane pyrimidine dimers (CPD) but represent substructures whose formation is important for telomerase activity.

Stem P6.1 has been shown to be of utmost importance for telomerase activity<sup>43</sup>. Mutation of U307 drastically reduces telomerase activity, however it does not have an influence on hTERT binding<sup>31,190</sup>. The cross-link observed at U307 and the fact that it is also strongly modified by DMS is noteworthy. DMS modifies the N3 of cytosines and the N1 of adenosines; however pseudouridines ( $\Psi$ ) are often present as a keto-enol tautomer with an N3 atom accessible for methylation. Indeed, U307 was found to be a  $\Psi$  in two independent studies<sup>42,178</sup>. According to the solution structure of a P6.1 stem containing two  $\Psi$  at positions 306 and 307,  $\Psi$ 307 points towards the stem and coordinates a complex hydrogen network which confers additional thermal stability to  $\Psi$ 6.1 compared to the canonical P6.1 stem. We propose that the observed cross-link at U307 is therefore more likely to be an RNA-RNA interaction rather than a RNA-protein contact. This is also in line with the fact that U307 does not play a role in hTERT binding and was found to be close to A54 in an *in vitro* cross-linking study<sup>31,32,190</sup>. Overall we were surprised to observe only very few cross-linked nucleotides within the CR4/CR5 domain despite its importance for telomerase activity as main hTERT binding site.

### 5.1.3. The topology of the scaRNA domain

The last functional domain of human telomerase RNA is the scaRNA domain located in the 3' part of the molecule. The secondary structure shows strong similarities with the canonical hairpin-hinge-hairpin motif of H/ACA snoRNAs and contains the conserved box H as well as the ACA trinucleotide<sup>16</sup>.

In addition, it harbours the CAB box, making hTR an H/ACA scaRNA<sup>16</sup>. Human telomerase RNA biogenesis is similar to scaRNAs. As such, hTR is dependent on H/ACA proteins like Naf1, Nop10, Nhp2 and Dyskerin<sup>19,128</sup>. A sets of these proteins bind to each stem and the entire RNP complex is transported to the Cajal bodies<sup>19,194</sup>. These steps are mediated via the CAB and BIO boxes located in the terminal loop of stem P8b<sup>35,36</sup>. In contrast to the 5' pocket, cross-links were not observed in the corresponding 3' pocket. In addition, the nature of the observed cross-links within the scaRNA domain differs with respect to the high number of conserved purine residues (3 out of 8 nucleotides). The cross-links observed in J8a/8b are in close proximity to the CAB box, which is responsible for Cajal body translocation<sup>202</sup>. The close proximity to this element and the fact that both hairpins within the scaRNAs are loaded with H/ACA RNA binding proteins suggest that these cross-links are going to one of the corresponding H/ACA proteins. For the cross-links in the BIO box, it is also possible that these involve a protein; however, due to the nature of the loop, the cross-links might also reflect intra-loop interactions.

In summary, we have identified cross-links in the functionally important areas of human telomerase RNA, for example the triple helix and J2a/2b in the pseudoknot domain, J6a/6b and P6.1 in the CR4/CR5 domain and at potential protein binding sites in stem P8b as well as L8b of the scaRNA domain. The cross-links detected at the nucleotides forming the triple helix strongly suggest that the triple helix is formed *in vivo*.

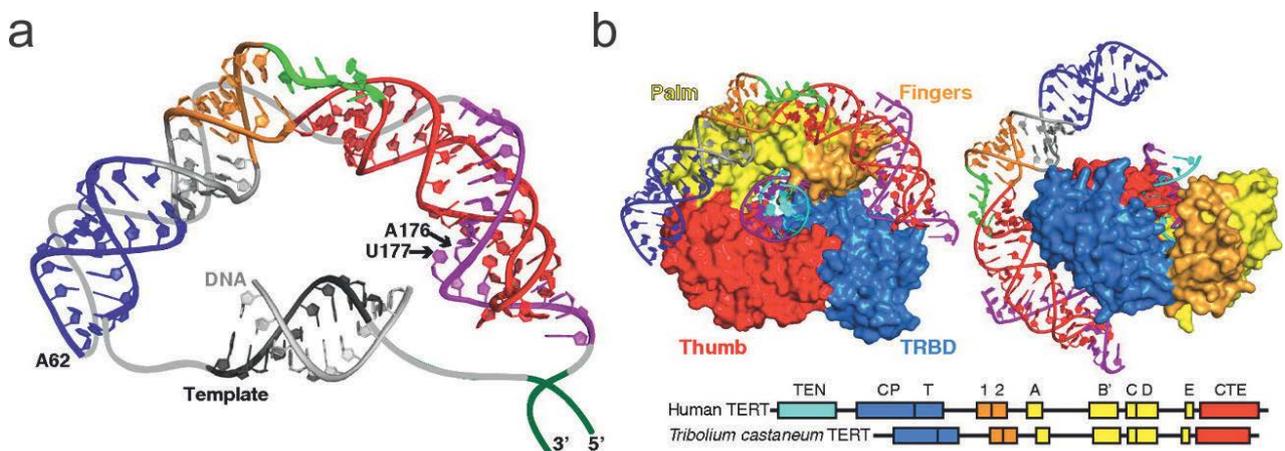
## 5.2. hTERT-dependent conformational changes within hTR

In order to obtain a functional holoenzyme both human telomerase RNA and human reverse transcriptase must interact with each other. Several studies investigated the nature of this interaction and in brief two independent hTERT binding sites were proposed: **a)** the pseudoknot domain and **b)** the CR4/CR5 domain. For example, hTR fragments consisting of nucleotides 33 to 147 and 164 to 330 can both independently bind to hTERT, however with different affinities<sup>203</sup>. In line with this, co-immunoprecipitation studies could identify significantly more RNA in hTERT/hTR 164-330 complexes than in hTERT/hTR 33-147 complexes<sup>11</sup>. For the CR4/CR5 region, the solution structure of both the P6 element as well as the P6.1 stem have confined the binding site of hTERT to nucleotides 255 to 320<sup>43,187</sup>. This was confirmed by the crystal structure of the medaka CR4/CR5 domain bound to the TRBD of mTERT<sup>33</sup>. As for the pseudoknot it has not been achieved yet to narrow down the hTERT binding site due to the lower binding affinity to hTERT. Mutations in the 3' strand of the P3 helix lower telomerase activity in a manner independent of hTERT binding<sup>182</sup>. By mapping the UV cross-linking sites within the native telomerase complex, we have identified regions of structural importance; however their precise nature and interacting partner remain elusive. To address this question hTR was cross-linked in absence of hTERT and the observed changes in the cross-linking

pattern were summarized in a differential map (figure 4.16). Our observations are in excellent agreement with previous published data which suggest two independent binding sites for hTERT<sup>44,203</sup>. Both in the proposed pseudoknot domain as well as in the CR4/CR5 domain significant changes in cross-linking intensity upon hTERT binding were detected.

### 5.2.1. hTERT-induced conformational changes in the pseudoknot

The triple helix in the pseudoknot domain is essential for telomerase activity<sup>29</sup>. It relies on an extensive network of Hoogsteen base triples to stabilize the pseudoknot and mutation of any of these residues abolishes telomerase activity<sup>23,25</sup>. The overall topology of the pseudoknot domain is defined by the aforementioned triple helix and the junction J2a/2b, which induces a strong bend and places the helices P2a and P2b in inversed V shape orientation to each other<sup>28</sup> (figure 5.2 a). The distance between the ends of the expanded pseudoknot spans approximately 70 Å, which is enough to accommodate the template-DNA duplex and its flanking regions<sup>23</sup>. The nucleotides, which show a change in cross-linking intensity in absence of hTERT, are located at the intersection of helices P2b (C116) to P3 (U113, U114, U115, C116 and C180) or in the linker J2b/3 (U100-U105). A computational model based on the TERT structure from *T. castaneum* positions the pseudoknot on the TERT protein either in parallel or perpendicular to the template-DNA duplex located in the central cavity of the TERT ring<sup>44,55</sup> (figure 5.2 b). Both of these positions would allow for the cross-linked nucleotides to interact with the protein. The pseudoknot is both times facing the TRBD domain of TERT which is in general responsible for hTR-hTERT binding<sup>47</sup>. It is tempting to speculate that the cross-linked nucleotides in P2b, P3 and J2b/3 might mediate the interactions between hTR and hTERT. However, with our current experimental setup we cannot identify RNA-protein contacts, and therefore we cannot exclude the possibility that the changes in cross-linking efficiency might be of indirect nature. In addition, the fact that cross-linking intensity increases in most cases when hTERT is absent, might indicate additional flexibility of hTERT. Thus, further studies on the exact role of these residues are of utmost importance.



**Figure 5.2 The core domain of hTR and possible interactions with tcTERT.** Computational models showing **a)** the pseudoknot domain of hTR based on the available solution structure<sup>25</sup> showing P2a.1 (blue), P2b (yellow), J2a/2b (green), P2b (red) and P3 (purple). A template-DNA duplex is positioned in the middle of the inverse V-shaped model. **b)** Possible orientations of the pseudoknot with regard to the *Tribolium castaneum* TERT protein<sup>55</sup>. The domains are coloured according to the legend below. The template-DNA duplex (purple and cyan, respectively) is located in the central cavity of the TERT ring. Figure adapted from Zhang et al., 2011<sup>44</sup>.

### 5.2.2. The influence of the CR4/CR5 domain on hTERT binding

The second site of hTERT-induced changes in cross-linking intensity is located in the conserved CR4/CR5 domain. As already mentioned earlier, this domain shows a much higher binding affinity to hTERT<sup>11,203</sup>. *In vitro* cross-linking studies between TERT and TR in *O. latipes* (medaka) identified several RNA residues which form a direct contact to the protein (U182-Tyr503, U187-Phe355 and U205-Trp477)<sup>75</sup> however the crystal structure of the TRBD of mTERT in complex with the CR4/CR5 element of mTR confirmed only U182 (U261 in hTR) to directly contact the protein and revealed new contacts as well (G189 and A199 in mTR, G268 and A301 in hTR)<sup>33,75</sup>. The crystal structure showed as well that stem P6.1 plays a major role in TERT recognition and binding. It undergoes a large conformational change upon TERT binding and therefore acts like a switch<sup>33</sup>. In respect to the change in P6.1 the entire three-way junction becomes rearranged and several new base pairs form upon TERT binding<sup>33</sup>. In good agreement with this study<sup>33</sup>, we observe the strongest changes in cross-linking intensities for nucleotides located in the CR4/CR5 domain (figure 4.16). C290 and U291, which are located in J6b/6a, face the human homolog (C266 in J6a/6b) of medaka U187, and show a two-fold increase in intensity in the absence of hTERT. This is in line with the proposed role of nucleotides from the P6b-P6a internal loop playing a role in hTERT binding<sup>187</sup>. P6b-P6a is not present in medaka and the key residue mediating RNA-protein contact is the bulged U182<sup>33,187</sup>. The human homolog of U182 is U261 or C262. U261 being the bulged nucleotide is supported by sequence alignment and NMR<sup>33,187</sup>, while C262 is favoured by phylogenetic studies<sup>17</sup>. Available DMS data shows no change in methylation both in presence and absence of hTERT, thus hinting for a non-canonical C•A pair with A295<sup>178,201</sup>. In contrast, the human homologs C266 and G268 of medaka U187 and G189 both show changes in DMS modification in presence of hTERT, as these residues become more or less accessible to DMS modification<sup>178</sup>. This suggests that hTERT bind to the CR4/CR5 element of hTR in a slightly different manner than described for medaka telomerase RNA<sup>33</sup>.

Stem P6.1 has been shown to undergo drastic conformational changes upon hTERT binding which also leads to the rearrangement of base pairs throughout the three-way junction<sup>33</sup>. In brief, the three base pairs C174-G216, C176-G213 and G198•U212 open up (potential human homologues are C247-G321, C255-G315 and G300•U314, however it is unclear if those interactions do exist) and pave the way for new interactions to be formed upon hTERT binding (A199-A200, U212-G231 and the two

base triples C174•G198-C177 and C176•U217-A173)<sup>33</sup> (figure 1.8). In human telomerase RNA, cross-linking intensities in absence of hTERT increase for the last two nucleotides of the 3' stem of P6.1, C313 and U314 as well as for U316 and C317 in J6.1/5, all of them being 100% conserved in vertebrates<sup>17</sup> (figure 4.16). U314, the human homolog for medaka U212 therefore could be involved in similar rearrangements (e.g. a potential pairing with G300 upon hTERT binding). In our opinion, this would be a solid suggestion given the high sequence conservation in this area. Furthermore, nucleotide U249 in J5/6a is also cross-linked, however its cross-linking intensity is independent of hTERT. Nonetheless it might participate in interactions with other residues of the three-way junction (e.g. A318 which becomes protected from DMS upon hTERT binding<sup>178</sup>). Based on our data we therefore propose a similar mechanism for hTERT binding to the one recently found in medaka<sup>33</sup>. We suggest rearrangements in base pairing in the three-way junction to accommodate the protein and allow a stable formation of the mature telomerase complex. Taken together, the major binding platform for hTERT is indeed located in the CR4/CR5 domain.

### **5.3. Identification of cross-linking partners**

The ability of telomerase to efficiently elongate the cellular telomeres is utterly dependent on the successful formation of the mature telomerase RNP complex. This requires correct folding of both hTR and hTERT in order to successfully bind to each other. Human telomerase RNA has three major domains important for catalytic activity, hTERT binding and maturation which are linked by a hypervariable region. Based on our initial UV cross-linking experiments we were interested in the structural elements that are spatially close. Therefore mutated variants of hTR were created based on the previously identified cross-linked nucleotides. By mapping the cross-linking sites within these constructs in presence of hTERT and comparing the obtained UV cross-linking pattern with that of the wildtype telomerase, we were able to gain new insights into the structural conformation of hTR.

#### **5.3.1. The triple helix is formed *in vivo***

The core element of the pseudoknot domain is the conserved triple helix. Its formation is crucial for telomerase activity and seems to be a general feature of telomerase RNAs, as a triple helix was also found in yeast as well<sup>23,25,29,181</sup>. Underlining the importance of the triple helix, the most profound changes in the cross-linking pattern were indeed observed on nucleotides, which are involved in the formation of said triple helix (figure 4.24). Mutants M3, M4 as well as M7 contain mutations in J2b/3, P2b stem and stem P3 and beautifully show the dynamics in this area. For instance, mutation of the J2b/3 linker nucleotides U100 to U105 in M3 lead to an increase in cross-linking intensity at U113 to U115 (figure 4.21). Interestingly, U113 to U115 are part of the base triples with their interaction partner being U100 to U103. Construct M4, in which U113 to C116 were mutated, showed similar results; here, the cross-linking intensity changed drastically for residues U99, U102, C104 and U105

which are part of the J2b/3 linker region. The same is true for constructs M3 and M7 as well. Mutations in J2b/3 of M3 had an impact on stem P3, in detail on the cross-linking intensity of nucleotides U177, U179 and C180, while mutations in M7 resulted in a change in intensity for U103 in J2b/3. This strongly suggests that the triple helix is indeed formed *in vivo*. Furthermore, construct M7 displays changes in the cross-linking pattern in J2a.1/3 with new cross-links at C166 and C170. This is in good agreement with the reported solution structure which placed the linker J2a.1/3 into the minor groove of helix P2b<sup>23,25</sup>. Unfortunately, all structures available<sup>23,25</sup> (e.g. PDB 2K95) are based on truncated pseudoknot variants that have G95-C119 as the last base pair of stem P2b; on the one hand the length of J2a.1/3 as well as P2b is drastically shortened, but on the other hand increases the likelihood of obtaining a good solution structure. Furthermore, construct M6 also shows an increase in cross-linking intensity for nucleotides C87 and C88 in J2a/2b. Based on our observations, it seems plausible that the nucleotides in J2a.1/3 form base triples with P2b and extended stems.

Constructs M1, M2, M5 and M6 displayed more subtle changes in the cross-linking pattern compared to wildtype. M1, which contains mutations C38 to C43, leads to a decrease in intensity at nucleotides U113 to U115 and might therefore indicate that the template region is in close proximity to the triple helix. Considering the fact, that A176 and U177 of the triple helix are known to be involved in catalysis<sup>29</sup>, this spatial constraint is in good agreement. The mutated nucleotides in J2a/2b, the five-nucleotide bulge defining the overall topology of the pseudoknot domain<sup>28</sup>, have very limited effect on the cross-linking pattern. So far, only the adjacent C90 shows increased intensity. However, hTR construct M4 (C113 to U116) influenced the cross-linking intensity of the nucleotides in J2a/2b. This might indicate that while the triple helix is not formed, flexibility for J2a/2b is greatly increased. Finally, constructs M5 and M6 in the linker J2a.1/3 are interesting with respect to as the initial cross-links disappeared, while the cross-linking intensity increased for nucleotides U329 and C330 in J5/4.2, indicating that J2a.1/3 might be in proximity of the CR4/CR domain. Interestingly, the intensity of U329 and C330 was also increased for constructs M1 and M7, therefore suggesting that the pseudoknot is in proximity of J5/4.2. However, for M5 and M6, it is noteworthy that the mutations in M5 and M6 are transversions rather than transitions as it is the case for the other constructs.

Taken together, we have shown that mutation of the residues involved in the triple helix scaffold influences other nucleotides of this element. This indicates that these residues are in close proximity and therefore react to changes of their binding partners, which leads us to the conclusion that the triple helix is formed *in vivo*.

### 5.3.2. The CR4/CR5 domain is in proximity of the triple helix

For the mutants in the CR4/CR5 region the observed changes are much more subtle than those in the pseudoknot. Of particular interest is the increase in cross-linking intensity at nucleotide U307 in constructs M4 and M7. Together with the observation that U329 and C330 in J5/4.2 are influenced by nucleotides in the pseudoknot (as seen with constructs M1, M5, M6 and M7), the increase at U307 further strengthens the hypothesis that the pseudoknot and the CR4/CR5 domain are in close proximity. The similarity in the cross-linking pattern of M11 with those obtained in absence (figure 4.11 b and figure 4.22) of hTERT is striking and might therefore indicate that this mutant is not able to bind hTERT. This would be in agreement with previous studies showing a high affinity of the CR4/CR5 domain for hTERT<sup>13,31,194</sup>. Strikingly, a new cross-link was observed at G308, a nucleotide which is strongly modified by DMS in absence of hTERT but protected when hTERT is present<sup>178</sup>. This might explain the complete loss of activity of hTR variant M11, as hTERT might not be able to interact with G308 anymore. In addition, both M10 and M11 show a strong increase in cross-linking intensity on U103 in the pseudoknot (figure 4.21 b). This suggests close proximity between the CR4/CR5 domain and the triple helix. Given the fact that residues from both regions have been shown to be important for catalysis (e.g. U307 and G309 in CR4/CR5<sup>31,190</sup> and A176 and U177 in the pseudoknot<sup>23</sup>) it is plausible that the regions containing these nucleotides must be in close physical proximity to each other. Interestingly, neither M10 nor M11 showed a change in cross-linking intensity for nucleotides in the template region. This is surprising, given the fact that A54 was shown to cross-link to U307 *in vitro*<sup>32</sup>. Further investigation is needed to find out which residues mediate the contacts between the pseudoknot domain and the CR4/CR5 domain. Our observations of telomerase activity and comparison to WT confirmed the importance of the triple helix and the CR4/CR5 region as most of the mutants in the triple helix as well as in the CR4/CR5 domain showed a complete lack of telomerase activity (figure 4.18). The decrease in activity of mutant M14, and the loss of activity in the case of M15 can be explained by interfering with regions required for hTR processing, maturation and localization<sup>40,202</sup>. Therefore, hTR levels of M15 are drastically decreased, which prevented the analysis of the cross-linking pattern for this mutant.

Taken together, our results provide first evidence that the pseudoknot and the triple helix are formed *in vivo* with the template region being in close proximity to the triple helix. In addition, our UV cross-linking data strongly suggests that the CR4/CR5 domain is also in close proximity to the triple helix. Upon binding of hTERT, we propose a rearrangement of the three-way junction in the CR4/CR5 domain, similar to binding mechanism of mTERT to mTR<sup>33</sup>. We also identified a potential RNA-protein cross-link at residue C104 which might be directed to hTERT. Based on the observations of the cross-linking pattern for the mutants in the scaRNA domain, we suggest that the cross-links

observed in this region might be directed to H/ACA RNA binding proteins. Summarizing these observations, we have gained new insights into the structural organization of the human telomerase RNA.

## 6. References

1. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How many species are there on Earth and in the ocean? *PLoS Biol.* **9**, e1001127 (2011).
2. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–57 (1983).
3. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147–57 (1982).
4. Nobel Media AB. The Nobel Prize in Chemistry 1989. *Nobelprize.org* Web (2014). at <[http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1989](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1989)>
5. McClintock, B. The Stability of Broken Ends of Chromosomes in *Zea Mays*. *Genetics* **26**, 234–82 (1941).
6. Muller, H. J. The remaking of chromosomes. *Collect. Net* **13**, 1181–1198 (1938).
7. Hayflick, L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* **37**, 614–36 (1965).
8. Olovnikov, A. M. [Principle of marginotomy in template synthesis of polynucleotides]. *Dokl. Akad. Nauk SSSR* **201**, 1496–9 (1971).
9. Greider, C. W. & Blackburn, E. H. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337**, 331–337 (1989).
10. Savage, S. A. Human telomeres and telomere biology disorders. *Prog. Mol. Biol. Transl. Sci.* **125**, 41–66 (2014).
11. Bachand, F. & Autexier, C. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. *Mol. Cell. Biol.* **21**, 1888–97 (2001).
12. Tesmer, V. M. *et al.* Two inactive fragments of the integral RNA cooperate to assemble active telomerase with the human protein catalytic subunit (hTERT) in vitro. *Mol. Cell. Biol.* **19**, 6207–16 (1999).
13. Mitchell, J. R. & Collins, K. Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Mol. Cell* **6**, 361–71 (2000).
14. Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236–41 (1995).
15. Zaug, A. J., Lingner, J. & Cech, T. R. Method for determining RNA 3' ends and application to human telomerase RNA. *Nucleic Acids Res.* **24**, 532–3 (1996).
16. Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol. Cell. Biol.* **19**, 567–76 (1999).

17. Chen, J.-L., Blasco, M. a & Greider, C. W. Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503–14 (2000).
18. Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr. Opin. Struct. Biol.* **16**, 307–18 (2006).
19. Egan, E. D. & Collins, K. Biogenesis of telomerase ribonucleoproteins. *RNA* **18**, 1747–59 (2012).
20. Berman, A. J., Akiyama, B. M., Stone, M. D. & Cech, T. R. The RNA accordion model for template positioning by telomerase RNA during telomeric DNA synthesis. *Nat. Struct. Mol. Biol.* **18**, 1371–5 (2011).
21. Akiyama, B. M., Gomez, A. & Stone, M. D. A conserved motif in *Tetrahymena thermophila* telomerase reverse transcriptase is proximal to the RNA template and is essential for boundary definition. *J. Biol. Chem.* 0–19 (2013). doi:10.1074/jbc.M113.452425
22. Chen, J. J.-L. Telomerase database. (2013). at <<http://telomerase.asu.edu/sequences.html>>
23. Kim, N.-K. *et al.* Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J. Mol. Biol.* **384**, 1249–61 (2008).
24. Kirwan, M. & Dokal, I. Dyskeratosis congenita: a genetic disorder of many faces. *Clin. Genet.* **73**, 103–12 (2008).
25. Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* **17**, 671–82 (2005).
26. Xie, M. *et al.* Structure and function of the smallest vertebrate telomerase RNA from teleost fish. *J. Biol. Chem.* **283**, 2049–59 (2008).
27. Ly, H., Blackburn, E. H. & Parslow, T. G. Comprehensive structure-function analysis of the core domain of human telomerase RNA. *Mol. Cell. Biol.* **23**, 6849–56 (2003).
28. Zhang, Q., Kim, N.-K., Peterson, R. D., Wang, Z. & Feigon, J. Structurally conserved five nucleotide bulge determines the overall topology of the core domain of human telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18761–8 (2010).
29. Qiao, F. & Cech, T. R. Triple-helix structure in telomerase RNA contributes to catalysis. *Nat. Struct. Mol. Biol.* **15**, 634–40 (2008).
30. Dandjinou, A. T. *et al.* A Phylogenetically Based Secondary Structure for the Yeast Telomerase RNA. *Curr. Biol.* **14**, 1148–1158 (2004).
31. Chen, J.-L., Opperman, K. K. & Greider, C. W. A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic Acids Res.* **30**, 592–7 (2002).
32. Ueda, C. T. & Roberts, R. W. Analysis of a long-range interaction between conserved domains of human telomerase RNA. *RNA* **10**, 139–47 (2004).
33. Huang, J. *et al.* Structural basis for protein-RNA recognition in telomerase. *Nat. Struct. Mol. Biol.* (2014). doi:10.1038/nsmb.2819

34. Kim, N.-K., Zhang, Q. & Feigon, J. Structure and sequence elements of the CR4/5 domain of medaka telomerase RNA important for telomerase function. *Nucleic Acids Res.* **42**, 3395–408 (2013).
35. Fu, D. & Collins, K. Distinct Biogenesis Pathways for Human Telomerase RNA and H/ACA Small Nucleolar RNAs. *Mol. Cell* **11**, 1361–1372 (2003).
36. Jády, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *J. Cell Biol.* **164**, 647–52 (2004).
37. Jády, B. E. *et al.* Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J.* **22**, 1878–88 (2003).
38. Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J.* **21**, 2746–56 (2002).
39. Verheggen, C. *et al.* Mammalian and yeast U3 snoRNPs are matured in specific and related nuclear compartments. *EMBO J.* **21**, 2736–45 (2002).
40. Egan, E. D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Mol. Cell. Biol.* **32**, 2428–39 (2012).
41. Theimer, C. A. *et al.* Structural and functional characterization of human telomerase RNA processing and cajal body localization signals. *Mol. Cell* **27**, 869–81 (2007).
42. Kim, N.-K., Theimer, C. A., Mitchell, J. R., Collins, K. & Feigon, J. Effect of pseudouridylation on the structure and activity of the catalytically essential P6.1 hairpin in human telomerase RNA. *Nucleic Acids Res.* **38**, 6746–56 (2010).
43. Leeper, T., Leulliot, N. & Varani, G. The solution structure of an essential stem-loop of human telomerase RNA. *Nucleic Acids Res.* **31**, 2614–21 (2003).
44. Zhang, Q., Kim, N.-K. & Feigon, J. Architecture of human telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20325–32 (2011).
45. Baltimore, D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1209–11 (1970).
46. Temin, H. M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211–3 (1970).
47. Autexier, C. & Lue, N. F. The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* **75**, 493–517 (2006).
48. Lingner, J. *et al.* Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–7 (1997).
49. Osanai, M., Kojima, K. K., Futahashi, R., Yaguchi, S. & Fujiwara, H. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* **376**, 281–9 (2006).

50. Malik, H. S., Burke, W. D. & Eickbush, T. H. Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*. *Gene* **251**, 101–8 (2000).
51. Vítková, M., Král, J., Traut, W., Zrzavý, J. & Marec, F. The evolutionary origin of insect telomeric repeats, (TTAGG)<sub>n</sub>. *Chromosome Res.* **13**, 145–56 (2005).
52. Sahara, K., Marec, F. & Traut, W. TTAGG telomeric repeats in chromosomes of some insects and other arthropods. *Chromosome Res.* **7**, 449–60 (1999).
53. Frydrychová, R., Grossmann, P., Trubac, P., Vítková, M. & Marec, F. Phylogenetic distribution of TTAGG telomeric repeats in insects. *Genome* **47**, 163–78 (2004).
54. Wyatt, H. D. M., West, S. C. & Beattie, T. L. InTERTpreting telomerase structure and function. *Nucleic Acids Res.* **38**, 5609–22 (2010).
55. Gillis, A., Schuller, A. P. & Skordalakes, E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633–7 (2008).
56. Rouda, S. & Skordalakes, E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure* **15**, 1403–12 (2007).
57. Mason, M., Schuller, A. & Skordalakes, E. Telomerase structure function. *Curr. Opin. Struct. Biol.* **21**, 92–100 (2011).
58. Moriarty, T. J., Marie-Egyptienne, D. T. & Autexier, C. Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. *Mol. Cell. Biol.* **24**, 3720–33 (2004).
59. Sekaran, V. G., Soares, J. & Jarstfer, M. B. Structures of telomerase subunits provide functional insights. *Biochim. Biophys. Acta* **1804**, 1190–201 (2010).
60. Steczkiewicz, K. *et al.* Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9443–8 (2011).
61. Jacobs, S. a, Podell, E. R. & Cech, T. R. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nat. Struct. Mol. Biol.* **13**, 218–25 (2006).
62. Jurczyk, J. *et al.* Direct involvement of the TEN domain at the active site of human telomerase. *Nucleic Acids Res.* **39**, 1774–88 (2011).
63. Romi, E. *et al.* High-resolution physical and functional mapping of the template adjacent DNA binding site in catalytically active telomerase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8791–6 (2007).
64. Lue, N. F. & Li, Z. Modeling and structure function analysis of the putative anchor site of yeast telomerase. *Nucleic Acids Res.* **35**, 5213–22 (2007).
65. Wyatt, H. D. M., Tsang, A. R., Lobb, D. A. & Beattie, T. L. Human telomerase reverse transcriptase (hTERT) Q169 is essential for telomerase function in vitro and in vivo. *PLoS One* **4**, e7176 (2009).

66. Xia, J., Peng, Y., Mian, I. S. & Lue, N. F. Identification of functionally important domains in the N-terminal region of telomerase reverse transcriptase. *Mol. Cell. Biol.* **20**, 5196–207 (2000).
67. Harrington, L. A. & Greider, C. W. Telomerase primer specificity and chromosome healing. *Nature* **353**, 451–4 (1991).
68. Collins, K. & Greider, C. W. Tetrahymena telomerase catalyzes nucleolytic cleavage and nonprocessive elongation. *Genes Dev.* **7**, 1364–76 (1993).
69. Zaug, A. J., Podell, E. R. & Cech, T. R. Mutation in TERT separates processivity from anchor-site function. *Nat. Struct. Mol. Biol.* **15**, 870–2 (2008).
70. Armbruster, B. N., Banik, S. S., Guo, C., Smith, A. C. & Counter, C. M. N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo. *Mol. Cell. Biol.* **21**, 7775–86 (2001).
71. Wyatt, H. D. M., Lobb, D. A. & Beattie, T. L. Characterization of physical and functional anchor site interactions in human telomerase. *Mol. Cell. Biol.* **27**, 3226–40 (2007).
72. Hammond, P. W., Lively, T. N. & Cech, T. R. The anchor site of telomerase from *Euplotes aediculatus* revealed by photo-cross-linking to single- and double-stranded DNA primers. *Mol. Cell. Biol.* **17**, 296–308 (1997).
73. Yang, Y., Chen, Y., Zhang, C., Huang, H. & Weissman, S. M. Nucleolar localization of hTERT protein is associated with telomerase function. *Exp. Cell Res.* **277**, 201–9 (2002).
74. Lai, C. K., Mitchell, J. R. & Collins, K. RNA binding domain of telomerase reverse transcriptase. *Mol. Cell. Biol.* **21**, 990–1000 (2001).
75. Bley, C. J. *et al.* RNA-protein binding interface in the telomerase ribonucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20333–8 (2011).
76. Xie, M., Podlevsky, J. D., Qi, X., Bley, C. J. & Chen, J. J.-L. A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic Acids Res.* **38**, 1982–96 (2010).
77. Harkisheimer, M., Mason, M., Shuvaeva, E. & Skordalakes, E. A Motif in the Vertebrate Telomerase N-Terminal Linker of TERT Contributes to RNA Binding and Telomerase Activity and Processivity. *Structure* 1–9 (2013). doi:10.1016/j.str.2013.08.013
78. Mitchell, M., Gillis, A., Futahashi, M., Fujiwara, H. & Skordalakes, E. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nat. Struct. Mol. Biol.* **17**, 513–8 (2010).
79. Wang, J. *et al.* Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell* **89**, 1087–99 (1997).
80. Di Marco, S. *et al.* Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **280**, 29765–70 (2005).
81. Sarafianos, S. G. *et al.* Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. *EMBO J.* **21**, 6614–24 (2002).

82. Lue, N. F., Lin, Y.-C. & Mian, I. S. A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. *Mol. Cell. Biol.* **23**, 8440–9 (2003).
83. Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nat. Genet.* **17**, 498–502 (1997).
84. Bolton, E. C., Mildvan, A. S. & Boeke, J. D. Inhibition of reverse transcription in vivo by elevated manganese ion concentration. *Mol. Cell* **9**, 879–89 (2002).
85. Bosoy, D. & Lue, N. F. Functional analysis of conserved residues in the putative ‘finger’ domain of telomerase reverse transcriptase. *J. Biol. Chem.* **276**, 46305–12 (2001).
86. Haering, C. H., Nakamura, T. M., Baumann, P. & Cech, T. R. Analysis of telomerase catalytic subunit mutants in vivo and in vitro in *Schizosaccharomyces pombe*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6367–72 (2000).
87. Kelleher, C., Teixeira, M. T., Förstemann, K. & Lingner, J. Telomerase: biochemical considerations for enzyme and substrate. *Trends Biochem. Sci.* **27**, 572–9 (2002).
88. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–90 (1992).
89. Jacobo-Molina, A. *et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6320–4 (1993).
90. Wu, R. A. & Collins, K. Sequence specificity of human telomerase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11234–5 (2014).
91. Brown, A. F. *et al.* A self-regulating template in human telomerase. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1402531111
92. Bosoy, D. & Lue, N. F. Yeast telomerase is capable of limited repeat addition processivity. *Nucleic Acids Res.* **32**, 93–101 (2004).
93. Morin, G. B. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521–9 (1989).
94. Sun, D., Lopez-Guajardo, C. C., Quada, J., Hurley, L. H. & Von Hoff, D. D. Regulation of catalytic activity and processivity of human telomerase. *Biochemistry* **38**, 4037–44 (1999).
95. Maine, I. P., Chen, S. F. & Windle, B. Effect of dGTP concentration on human and CHO telomerase. *Biochemistry* **38**, 15325–32 (1999).
96. Hardy, C. D., Schultz, C. S. & Collins, K. Requirements for the dGTP-dependent repeat addition processivity of recombinant *Tetrahymena* telomerase. *J. Biol. Chem.* **276**, 4863–71 (2001).
97. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–34 (2008).

98. De Lange, T. *et al.* Structure and variability of human chromosome ends. *Mol. Cell. Biol.* **10**, 518–27 (1990).
99. Hastie, N. D. *et al.* Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–8 (1990).
100. Lejnine, S., Makarov, V. L. & Langmore, J. P. Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 2393–7 (1995).
101. Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–2 (1990).
102. Damm, K. *et al.* A highly selective telomerase inhibitor limiting human cancer cell proliferation. *EMBO J.* **20**, 6958–68 (2001).
103. De Lange, T. How telomeres solve the end-protection problem. *Science* **326**, 948–52 (2009).
104. Baird, D. M., Rowson, J., Wynford-Thomas, D. & Kipling, D. Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nat. Genet.* **33**, 203–7 (2003).
105. Capper, R. *et al.* The nature of telomere fusion and a definition of the critical telomere length in human cells. *Genes Dev.* **21**, 2495–508 (2007).
106. Makarov, V. L., Hirose, Y. & Langmore, J. P. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657–66 (1997).
107. McElligott, R. & Wellinger, R. J. The terminal DNA structure of mammalian chromosomes. *EMBO J.* **16**, 3705–14 (1997).
108. Klobutcher, L. A., Swanton, M. T., Donini, P. & Prescott, D. M. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci.* **78**, 3015–3019 (1981).
109. Henderson, E. R. & Blackburn, E. H. An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell. Biol.* **9**, 345–348 (1989).
110. Yuan, X. *et al.* Presence of telomeric G-strand tails in the telomerase catalytic subunit TERT knockout mice. *Genes Cells* **4**, 563–72 (1999).
111. Hemann, M. T. & Greider, C. W. G-strand overhangs on telomeres in telomerase-deficient mouse cells. *Nucleic Acids Res.* **27**, 3964–9 (1999).
112. Sfeir, A. J., Chai, W., Shay, J. W. & Wright, W. E. Telomere-end processing the terminal nucleotides of human chromosomes. *Mol. Cell* **18**, 131–8 (2005).
113. Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–14 (1999).
114. O'Sullivan, R. J. & Karlseder, J. Telomeres: protecting chromosomes against genome instability. *Nat. Rev. Mol. Cell Biol.* **11**, 171–81 (2010).
115. Nandakumar, J. & Cech, T. R. Finding the end: recruitment of telomerase to telomeres. *Nat. Rev. Mol. Cell Biol.* (2013). doi:10.1038/nrm3505

116. Yanez, G. H., Khan, S. J., Locovei, A. M., Pedroso, I. M. & Fletcher, T. M. DNA structure-dependent recruitment of telomeric proteins to single-stranded/double-stranded DNA junctions. *Biochem. Biophys. Res. Commun.* **328**, 49–56 (2005).
117. Pedroso, I. M., Hayward, W. & Fletcher, T. M. The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures. *Nucleic Acids Res.* **37**, 1541–54 (2009).
118. Zaug, A. J., Podell, E. R. & Cech, T. R. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci.* **102**, 10864–10869 (2005).
119. Ye, J. Z.-S. *et al.* TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J. Biol. Chem.* **279**, 47264–71 (2004).
120. Kubo, Y., Okazaki, S., Anzai, T. & Fujiwara, H. Structural and phylogenetic analysis of TRAS, telomeric repeat-specific non-LTR retrotransposon families in Lepidopteran insects. *Mol. Biol. Evol.* **18**, 848–57 (2001).
121. Okazaki, S., Ishikawa, H. & Fujiwara, H. Structural analysis of TRAS1, a novel family of telomeric repeat-associated retrotransposons in the silkworm, *Bombyx mori*. *Mol. Cell. Biol.* **15**, 4545–52 (1995).
122. Fujiwara, H., Osanai, M., Matsumoto, T. & Kojima, K. K. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. *Chromosome Res.* **13**, 455–67 (2005).
123. Tatsuke, T. *et al.* The telomere-specific non-LTR retrotransposons SART1 and TRAS1 are suppressed by Piwi subfamily proteins in the silkworm, *Bombyx mori*. *Cell. Mol. Biol. Lett.* **15**, 118–33 (2010).
124. Mason, J. M., Reddy, H. M. & Frydrychová, R. in *DNA replication - Current Advances* (ed. Seligmann, H.) (InTech, 2011).
125. Kiss, T., Fayet-Lebaron, E. & Jády, B. E. Box H/ACA small ribonucleoproteins. *Mol. Cell* **37**, 597–606 (2010).
126. Hamma, T. & Ferré-D'Amaré, A. R. The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification. *J. Biol. Chem.* **285**, 805–9 (2010).
127. Wang, C. & Meier, U. T. Architecture and assembly of mammalian H/ACA small nucleolar and telomerase ribonucleoproteins. *EMBO J.* **23**, 1857–67 (2004).
128. Darzacq, X. *et al.* Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J. Cell Biol.* **173**, 207–18 (2006).
129. Richard, P., Kiss, A. M., Darzacq, X. & Kiss, T. Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol. Cell. Biol.* **26**, 2540–9 (2006).
130. Yang, Y. *et al.* Conserved composition of mammalian box H/ACA and box C/D small nucleolar ribonucleoprotein particles and their interaction with the common factor Nopp140. *Mol. Biol. Cell* **11**, 567–77 (2000).

131. Boulon, S. *et al.* PHAX and CRM1 are required sequentially to transport U3 snoRNA to nucleoli. *Mol. Cell* **16**, 777–87 (2004).
132. Tycowski, K. T., Shu, M.-D., Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol. Cell* **34**, 47–57 (2009).
133. Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644–8 (2009).
134. Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M. & Terns, M. P. Cell cycle-regulated trafficking of human telomerase to telomeres. *Mol. Biol. Cell* **17**, 955–65 (2006).
135. Xi, H., Li, C., Ren, F., Zhang, H. & Zhang, L. Telomere, aging and age-related diseases. *Aging Clin. Exp. Res.* **25**, 139–46 (2013).
136. Schroeder, T. M. & Hofbauer, M. Letter: Dyskeratosis congenita Zinsser-Cole-Engman form with abnormal karyotype. *Dermatologica* **151**, 316–8 (1975).
137. Angrisani, A., Vicidomini, R., Turano, M. & Furia, M. Human dyskerin: beyond telomeres. *Biol. Chem.* (2014). doi:10.1515/hsz-2013-0287
138. Armanios, M. Syndromes of telomere shortening. *Annu. Rev. Genomics Hum. Genet.* **10**, 45–61 (2009).
139. Thumati, N. R. *et al.* Severity of X-linked dyskeratosis congenita (DKCX) cellular defects is not directly related to dyskerin (DKC1) activity in ribosomal RNA biogenesis or mRNA translation. *Hum. Mutat.* (2013). doi:10.1002/humu.22447
140. He, J. *et al.* Targeted disruption of Dkc1, the gene mutated in X-linked dyskeratosis congenita, causes embryonic lethality in mice. *Oncogene* **21**, 7740–4 (2002).
141. Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L. & Carbon, J. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol. Cell. Biol.* **19**, 7461–72 (1999).
142. Carroll, K. A. & Ly, H. Telomere dysfunction in human diseases: the long and short of it! *Int. J. Clin. Exp. Pathol.* **2**, 528–43 (2009).
143. Ulaner, G. A., Hu, J. F., Vu, T. H., Giudice, L. C. & Hoffman, A. R. Telomerase activity in human development is regulated by human telomerase reverse transcriptase (hTERT) transcription and by alternate splicing of hTERT transcripts. *Cancer Res.* **58**, 4168–72 (1998).
144. Harle-Bachor, C. & Boukamp, P. Telomerase activity in the regenerative basal layer of the epidermis in human skin and in immortal and carcinoma-derived skin keratinocytes. *Proc. Natl. Acad. Sci.* **93**, 6476–6481 (1996).
145. Hiyama, K. *et al.* Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* **155**, 3711–5 (1995).
146. Chiu, C. P. *et al.* Differential expression of telomerase activity in hematopoietic progenitors from adult human bone marrow. *Stem Cells* **14**, 239–48 (1996).

147. Shay, J. W. & Wright, W. E. Role of telomeres and telomerase in cancer. *Semin. Cancer Biol.* **21**, 349–53 (2011).
148. Wright, W. E., Pereira-Smith, O. M. & Shay, J. W. Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Mol. Cell. Biol.* **9**, 3088–92 (1989).
149. Vinagre, J. *et al.* Telomerase promoter mutations in cancer: an emerging molecular biomarker? *Virchows Arch.* **465**, 119–133 (2014).
150. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–61 (2013).
151. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339**, 957–9 (2013).
152. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Ann. Hum. Biol.* **40**, 463–71
153. Low, K. C. & Tergaonkar, V. Telomerase: central regulator of all of the hallmarks of cancer. *Trends Biochem. Sci.* (2013). doi:10.1016/j.tibs.2013.07.001
154. Draskovic, I. & Londono-Vallejo, A. Telomere Recombination and the ALT Pathway: a Therapeutic Perspective for Cancer. *Curr. Pharm. Des.* (2014). at <<http://www.ncbi.nlm.nih.gov/pubmed/24975611>>
155. Romaniuk, A., Kopczynski, P., Książek, K. & Rubiś, B. Telomerase Modulation in Therapeutic Approach. *Curr. Pharm. Des.* (2014). at <<http://www.ncbi.nlm.nih.gov/pubmed/24975610>>
156. Mocellin, S., Pooley, K. A. & Nitti, D. Telomerase and the search for the end of cancer. *Trends Mol. Med.* **19**, 125–33 (2013).
157. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–6 (1958).
158. Nobel Media AB. The Nobel Prize in Chemistry 1962. *Nobelprize.org* Web (2014). at <[http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1962/index.html](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/index.html)>
159. Riek, R., Hornemann, S., Wider, G., Glockshuber, R. & Wüthrich, K. NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231). *FEBS Lett.* **413**, 282–8 (1997).
160. Purcell, E., Torrey, H. & Pound, R. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.* **69**, 37–38 (1946).
161. Bloch, F., Hansen, W. & Packard, M. Nuclear Induction. *Phys. Rev.* **69**, 127–127 (1946).
162. Williamson, M. P., Havel, T. F. & Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by <sup>1</sup>H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315 (1985).
163. Wildauer, M., Zemora, G., Liebeg, A., Heisig, V. & Waldsich, C. Chemical probing of RNA in living cells. *Methods Mol. Biol.* **1086**, 159–76 (2014).

164. Liebeg, A. & Waldsich, C. Probing RNA structure within living cells. *Methods Enzymol.* **468**, 219–38 (2009).
165. Sachsenmaier, N., Handl, S., Debeljak, F. & Waldsich, C. Mapping RNA structure in vitro using nucleobase-specific probes. *Methods Mol. Biol.* **1086**, 79–94 (2014).
166. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–5 (2014).
167. Spitale, R. C., Flynn, R. A., Torre, E. A., Kool, E. T. & Chang, H. Y. RNA structural analysis by evolving SHAPE chemistry. *Wiley Interdiscip. Rev. RNA* **5**, 867–81
168. Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* **29**, 607–14 (2011).
169. Moulton, J., Pedersen, J. T., Judson, R. & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins* **23**, ii–v (1995).
170. Moulton, J. *et al.* Critical assessment of methods of protein structure prediction-Round VII. *Proteins* **69 Suppl 8**, 3–9 (2007).
171. Li, C. *et al.* FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnol.* **11**, 92 (2011).
172. Cristofari, G. & Lingner, J. Telomere length homeostasis requires that telomerase levels are limiting. *EMBO J.* **25**, 565–74 (2006).
173. Zhang, L., Zhang, K., Prändl, R. & Schöffl, F. Detecting DNA-binding of proteins in vivo by UV cross-linking and immunoprecipitation. *Biochem. Biophys. Res. Commun.* **322**, 705–11 (2004).
174. ALEXANDER, P. & MOROSON, H. Cross-linking of deoxyribonucleic acid to protein following ultra-violet irradiation different cells. *Nature* **194**, 882–3 (1962).
175. González-Ramírez, I. *et al.* On the photoproduction of DNA/RNA cyclobutane pyrimidine dimers. *Theor. Chem. Acc.* **128**, 705–711 (2010).
176. Climent, T., González-Ramírez, I., González-Luque, R., Merchán, M. & Serrano-Andrés, L. Cyclobutane Pyrimidine Photodimerization of DNA/RNA Nucleobases in the Triplet State. *J. Phys. Chem. Lett.* **1**, 2072–2076 (2010).
177. Roca-Sanjuán, D., Olaso-González, G., González-Ramírez, I., Serrano-Andrés, L. & Merchán, M. Molecular basis of DNA photodimerization: intrinsic production of cyclobutane cytosine dimers. *J. Am. Chem. Soc.* **130**, 10768–79 (2008).
178. Zemora, G. Structural insights into the human telomerase RNA. (University of Vienna, 2014).
179. Friedberg, E. C. A brief history of the DNA repair field. *Cell Res.* **18**, 3–7 (2008).
180. Vulliamy, T. *et al.* The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* **413**, 432–5 (2001).

181. Shefer, K. *et al.* A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol. Cell. Biol.* **27**, 2130–43 (2007).
182. Ly, H., Xu, L., Rivera, M. A., Parslow, T. G. & Blackburn, E. H. A role for a novel ‘trans-pseudoknot’ RNA-RNA interaction in the functional dimerization of human telomerase. *Genes Dev.* **17**, 1078–83 (2003).
183. Sauerwald, A. *et al.* Structure of active dimeric human telomerase. *Nat. Struct. Mol. Biol.* **20**, 454–60 (2013).
184. Gavory, G., Symmons, M. F., Krishnan Ghosh, Y., Klenerman, D. & Balasubramanian, S. Structural analysis of the catalytic core of human telomerase RNA by FRET and molecular modeling. *Biochemistry* **45**, 13304–11 (2006).
185. Chen, J.-L. & Greider, C. W. Template boundary definition in mammalian telomerase. *Genes Dev.* **17**, 2747–52 (2003).
186. Wu, R. A. & Collins, K. Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. *EMBO J.* (2014). doi:10.1002/embj.201387205
187. Leeper, T. C. & Varani, G. RNA The structure of an enzyme-activating fragment of human telomerase RNA. *RNA* **11**, 394–403 (2005).
188. Wu, L., Chai, D., Fraser, M. E. & Zimmerly, S. Structural variation and uniformity among tetraloop-receptor interactions and other loop-helix interactions in RNA crystal structures. *PLoS One* **7**, e49225 (2012).
189. Waldsich, C. & Pyle, A. M. A folding control element for tertiary collapse of a group II intron ribozyme. *Nat. Struct. Mol. Biol.* **14**, 37–44 (2007).
190. Robart, A. R. & Collins, K. Investigation of human telomerase holoenzyme assembly, activity, and processivity using disease-linked subunit variants. *J. Biol. Chem.* **285**, 4375–86 (2010).
191. Mitchell, J. R., Cheng, J. & Collins, K. A Box H/ACA Small Nucleolar RNA-Like Domain at the Human Telomerase RNA 3’ End. *Mol. Cell. Biol.* **19**, 567–576 (1999).
192. Ganot, P., Caizergues-Ferrer, M. & Kiss, T. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.* **11**, 941–56 (1997).
193. McMahon, M., Contreras, A. & Ruggero, D. Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip. Rev. RNA* **6**, 173–189 (2015).
194. Egan, E. D. & Collins, K. Specificity and stoichiometry of subunit interactions in the human telomerase holoenzyme assembled in vivo. *Mol. Cell. Biol.* **30**, 2775–86 (2010).
195. Jang, S. B. *et al.* Structure of an RNA internal loop consisting of tandem C-A+ base pairs. *Biochemistry* **37**, 11726–31 (1998).
196. Gomez, D. E. *et al.* Telomere structure and telomerase in health and disease (Review). *Int. J. Oncol.* **41**, 1561–9 (2012).

197. Niederer, R. O. & Zappulla, D. C. Refined secondary-structure models of the core of yeast and human telomerase RNAs directed by SHAPE. *RNA* (2014). doi:10.1261/rna.048959.114
198. Chodosh, L. a. UV cross-linking of proteins to nucleic acids. *Curr. Protoc. Mol. Biol.* **Chapter 12**, Unit 12.5 (2001).
199. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H. & James, T. L. A molecular switch underlies a human telomerase disease. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16998–7003 (2002).
200. Theimer, C. A., Finger, L. D., Trantirek, L. & Feigon, J. Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 449–54 (2003).
201. Antal, M., Boros, E., Solymosy, F. & Kiss, T. Analysis of the structure of human telomerase RNA in vivo. *Nucleic Acids Res.* **30**, 912–20 (2002).
202. Cristofari, G. *et al.* Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation. *Mol. Cell* **27**, 882–9 (2007).
203. Bachand, F., Triki, I. & Autexier, C. Human telomerase RNA-protein interactions. *Nucleic Acids Res.* **29**, 3385–93 (2001).

# Addendum

## I. List of abbreviations

Å	Angström
BIO box	biogenesis promoting box
CAB box	Cajal body box
CR1-8	conserved regions 1-8
CTE	C-terminal extension
D loop	displacement loop
DC	dyskeratosis congenita
DMS	dimethyl sulfate
hTR	human telomerase RNA
IFD	insertion in fingers domain
LTR	long terminal repeats
NAP	nucleotide addition processivity
NHEJ	non-homologous end joining
NMR	nuclear magnetic resonance
PK	pseudoknot
Pol II	DNA polymerase II
RAP	repeat addition processivity
RNP	Ribonucleoprotein
RT	reverse transcriptase
scaRNA	small Cajal body-specific RNA
SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
snoRNA	small nucleolar RNA
STE	stem terminus element
T loop	telomere loop

TBE	template boundary element
TEN domain	telomerase essential N-terminal domain
TER, TERC, TR	telomerase RNA (compound)
TERT	telomerase reverse transcriptase
TRBD	telomerase RNA binding domain
TRE	template recognition element
WT	wild type
X-DC	X-linked dyskeratosis congenita

## I. List of figures

Figure 1.1. Central dogma of molecular biology. ....	8
Figure 1.2. Schematic representation of DNA synthesis. ....	10
Figure 1.3. Comparison of different telomerase RNAs. ....	11
Figure 1.4. Secondary structure of the human telomerase RNA. ....	12
Figure 1.5. Secondary structure map of the hTR core domain. ....	13
Figure 1.6. Geometry of J2a/2b in hTR. ....	14
Figure 1.7. The TRBD of TERT interacts with the CR4/CR5 domain in telomerase. ....	16
Figure 1.8. Crystal structure of medaka CR4/CR5 bound to mTERT. ....	17
Figure 1.9. Solution structure of the terminal hairpin of the scaRNA domain. ....	19
Figure 1.10. Organization of TERT proteins from different species. ....	20
Figure 1.11. Structural organization of the TEN domain. ....	21
Figure 1.12. Alignment of TRBDs from different species. ....	22
Figure 1.13. Structural organization of <i>Tribolium castaneum</i> TERT. ....	23
Figure 1.14. Positioning of the DNA-RNA duplex at the active site of the TERT protein from <i>T. castaneum</i> . ....	25
Figure 1.15. The telomerase catalytic cycle. ....	27
Figure 1.16. Schematic representation of human telomere organization. ....	29
Figure 1.17. Biogenesis and maturation of the human telomerase RNP complex. ....	31
Figure 1.18. Mutations associated with Dyskeratosis congenita. ....	33
Figure 1.19. Disease-related mutations sites within hTR. ....	34
Figure 4.1. Pipeline for detecting UV cross-linking sites in hTR. ....	54
Figure 4.2. The important elements J2a/2b and the triple helix are prone to UV cross-linking. ....	57

Figure 4.3 Nucleotides in and adjacent to the template are cross-linked. ....	59
Figure 4.4 The conserved three-way junction in the CR4/CR5 domain shows minor cross-linking.....	62
Figure 4.5 The conserved BIO box is a target of UV cross-linking. ....	63
Figure 4.6 UV-induced cross-links within human telomerase RNA in the presence of hTERT.....	65
Figure 4.7 The triple helix undergoes discrete changes upon hTERT binding.....	67
Figure 4.8 UV cross-linking pattern of the pseudoknot in presence and absence of hTERT. ....	69
Figure 4.9 The template region and the template boundary element do not change in presence of hTERT.....	71
Figure 4.10 UV cross-linking pattern of the template/TBE in presence and absence of hTERT. ....	72
Figure 4.11 The CR4/CR5 cross-linking pattern shows significant changes upon hTERT binding.....	74
Figure 4.12 UV cross-linking pattern of the CR4/CR5 domain in presence and absence of hTERT. ....	75
Figure 4.13 UV cross-linking pattern of the hypervariable region in presence and absence of hTERT.	76
Figure 4.14 The conserved scaRNA domain is not influenced upon hTERT binding. ....	77
Figure 4.15 UV cross-linking pattern of the scaRNA domain in presence and absence of hTERT. ....	78
Figure 4.16 hTERT binding induces changes in the pseudoknot as well as in the CR4/CR5 domain. ...	79
Figure 4.17 Overview of hTR mutants.....	81
Figure 4.18 Telomerase activity of hTR mutants.....	82
Figure 4.19 The core pseudoknot changes in the mutant hTR variants.....	83
Figure 4.20 J5/4.2 and P6.1 in the CR4/CR5 domain show changes in cross-linking intensities. ....	84
Figure 4.21 Linker J2b/3 and J2a.1/3 show different cross-linking pattern.....	85
Figure 4.22 The CR4/CR5 domain shows changes in the cross-linking pattern. ....	86
Figure 4.23 The CR4/CR5 domain remains unaffected. ....	88
Figure 4.24 Spatial information for hTR obtained from UV cross-linking. ....	89
Figure 5.1 Chemical modifications observed in the pseudoknot domain.....	92
Figure 5.2 The core domain of hTR and possible interactions with tcTERT.....	96

## II. List of tables

Table 3-1. Summary of PCR protocols used for molecular cloning. ....	38
Table 3-2. Pipetting scheme for conventional, overlap and Fastcloning PCR.....	39
Table 3-3. Mutated hTR constructs used for mapping RNA-RNA interaction sites.....	39
Table 3-4. Composition for the stacking and separation gel for SDS PAGE. ....	47
Table 3-5. Workflow for Western blotting. ....	48
Table 3-6. Setup of 5' end-labeling reaction. ....	50
Table 3-7. Reaction setup for direct telomerase assay .....	50

Table III-1. Oligonucleotides used for molecular cloning of hTR mutants. ....v

Table III-2. Oligonucleotides used for mapping human telomerase RNA via primer extension. ....v

### III. List of oligonucleotides and plasmids

Oligo	Sequence (5' to 3')	Construct
CW177F	5' ACGGTGGTTGGCGGCAGGCCGAGGC 3'	pCW117
CW177R	5' GGCCTGCCGCCAACACCGTTCTATTCTAGAGC 3'	pCW117
CW135_3F	5' GTTAGGGTTAGACGGGGGGTGGCCACCACCCCTCC 3'	pCW135
CW135_3R	5' GGTGGCCACCCCCGTCTAACCTAACTGAGAAGGGCG 3'	pCW135
CW137_2F	5' GAAAAACAGCGCGCGAAAGGCAAAGC 3'	pCW137
CW137_2R	5' CCGTGCTTTTGCCTTTCGCGCGCTG 3'	pCW137
CW138_5F	5' GCCCGCTGAAAGTCAGCGGAGGGGACAGCGCGGGGAGCAAAGC 3'	pCW138
CW138_5R	5' CCGCGCGCTGTCCCCTCCGCTGACTTTCAGCGGGCGGAAAAGCCTCG 3'	pCW138
CW139_3F	5' GGCAGGCCGAGGCTTTTCCGCCCGCTCGGGGTCAGCGAG 3'	pCW139
CW139_3R	5' CCCGCGCGCTGTTTTCTCGCTGACCCGAGCGGGCGGAAAAGCC 3'	pCW139
CW145F	5' CGAGGGCGAGGCCAGGCCTTTCAGGCCGCGAG 3'	pCW145
CW145R	5' GAAAGGCCTGGGCCTCGCCCTCGCCCCGAGA 3'	pCW145
CW146_4F	5' CTCTTCTGCGGCCTAGGGGGCTGAACCTCGCCC 3'	pCW146
CW146_4R	5' GCGAGGGCGAGGTTTCAGGCCCTAGGCCGC 3'	pCW146
CW147_2F	5' CCCACAGCTCAGGAGGCTGCGCCGCG 3'	pCW147
CW147_2R	5' GCGCGCGGCGCAGCCTCCTGAGC 3'	pCW147
CW148_4F	5' CCTGGGTGCACGTCCCTGTGCTCAG 3'	pCW148
CW148_4R	5' GCGATTCCCTGAGCACAGGGACGTGC 3'	pCW148
CW163F	5' CGCCTTCCACGCAAGTAAGTAGAGCAA 3'	pCW163
CW163R	5' TTTGTTTGCTCTACTTACTTGCGTGGAAGGCG 3'	pCW163
CW187F	5' CGGGCCAGCGGCTAACGCTTTTTGTTTGCTC 3'	pCW187
CW187R	5' CAAACAAAAAGCGTTAGCCGCTGGCCCGTTTCG 3'	pCW187
CW189F	5' CGACCGCGGCCTCCGGGCGG 3'	pCW189
CW189R	5' CGAACCCCGCCCGGAGGCCGC 3'	pCW189
CW190F	5' ACTCTTCGCGGTGGCGATGGGTG 3'	pCW190
CW190R	5' AGGCACCCATCGCCACCGC 3'	pCW190
CW191F	5' CGGCTGACAGAGCCCGACTC 3'	pCW191
CW191R	5' CGAAGAGTCGGGCTCTGTGAGC 3'	pCW191

CW192F	5' CGAGAGACCCGCGGCTAGCGAAGCCC 3'	pCW192
CW192R	5' AAGAGTTGGGCTTCGCTAGCCGCGGGTC 3'	pCW192
CW62F	5' CAGAATTCGCGGCCGCGTCG 3'	pCW62
CW62R	5' CGGAATTCTTCAAAACTGAAAACTCATATATTCAGTATTTTACTCCCACAGCACC 3'	pCW62

**Table III-1. Oligonucleotides used for molecular cloning of hTR mutants.**

Oligo name	Sequence (5' to 3')	Binding site
hTR_72	5' GGGGAGCAAAAGCACGGCG 3'	C72 to C90
hTR_149	5' GTTTGCTCTAGAATGAACGGTG 3'	C149 to C170
hTR_178	5' GAACGGGCCAGCAGCTGACA 3'	U177 to C196
hTR_238	5' GCCTCCAGGCGGGGTTTCG 3'	C238 to C255
hTR_404	5' GTCCCACAGCTCAGGGAATC 3'	G404 to C423
hTR_433	5' GCATGTGTGAGCCGAGTCC 3'	G433 to C451

**Table III-2. Oligonucleotides used for mapping human telomerase RNA via primer extension.**

## IV. Reagents and Kits

Pierce ECL Plus kit	Thermo Scientific Cat. 32132
PureYield™ Plasmid Midiprep System	Promega Cat. A2495
PureYield™ Plasmid Miniprep System	Promega Cat. A1222
Wizard® SV Gel and PCR Clean-Up System	Promega Cat. A9282

3xFLAG M2 antibody	Sigma Aldrich Cat. F1804
Acryl amide 30% 37.5:1	Applichem Cat. A3626
Acryl amide 40% 19:1	Applichem Cat. A3658
Amersham Hybond ECL	Cat. RPN203D
Amersham Hyperfilm™ ECL	GE Healthcare Cat. 28-9068-36
Ammonium persulfate	Applichem Cat. A2941
Boric acid	Applichem Cat. A2940
CHAPS	Sigma Aldrich Cat. C9426

DMEM	Sigma Aldrich Cat. D6429
DMEM	Sigma Aldrich Cat. D6429
dNTPs (PCR grade)	Roche Cat. 11969064001
dNTPs (RT grade)	Roche Cat. 11277049001
DPBS	Sigma Aldrich Cat. D8537
EDTA	Sigma Aldrich Cat. E5134
FCS	Sigma Aldrich Cat. F7524
FuGENE <sup>®</sup> HD	Promega Cat. E2312
GoTaq polymerase	Promega Cat. M3171
PBS	Sigma Aldrich Cat. D8662
Pfu polymerase	Promega Cat. M7741
PfuUltra II fusion polymerase	Agilent Cat. 600672
Pre-stained protein marker	NEB Cat. P7706S
Proteinase K solution	Applichem Cat. 4392
Recombinant RNAsin <sup>®</sup>	Promega Cat. N2515
SIGMAFAST <sup>™</sup> Protease Inhibitor Cocktail Tablets	Sigma Cat. S8830
Skim milk	Applichem Cat. A0830
T4 polynucleotide kinase	NEB Cat. M0201
TEMED	Applichem Cat. A1148
Transcriptor reverse transcriptase	Roche Cat. 03531287001
Tris-HCl	Applichem Cat. A3452
Trizol solution	Sigma Aldrich Cat. T9424
Trypsin	Sigma Aldrich Cat. 59427C
Urea	Applichem Cat. 1049
Water-saturated phenol	Applichem Cat. A1624
$\alpha$ -IgG/IgM HRP antibody	Thermo Scientific Cat. 31444

## **V. Publications**

### **I. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction**

J. Cruz, M. Blanchet, M. Boniecki, J. Bujnicki, S. Chen, S. Cao, R. Das, F. Ding, N. Dokholyan, S. Flores, L. Huang, C. Lavender, V. Lisi, F. Major, K. Mikolajczak, D. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. Weeks, C. Waldsich, M. Wildauer, N. Leontis, E. Westhof,

RNA, 18:pp 610-25, 2012



# RNA

A PUBLICATION OF THE RNA SOCIETY

## ***RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction***

José Almeida Cruz, Marc-Frédéric Blanchet, Michal Boniecki, et al.

*RNA* 2012 18: 610-625 originally published online February 23, 2012  
Access the most recent version at doi:[10.1261/ma.031054.111](https://doi.org/10.1261/ma.031054.111)

---

**References** This article cites 64 articles, 26 of which can be accessed free at:  
<http://rnajournal.cshlp.org/content/18/4/610.full.html#ref-list-1>

**Open Access** Freely available online through the *RNA* Open Access option.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

EXIQON

RNA Shhhh...



---

To subscribe to *RNA* go to:  
<http://rnajournal.cshlp.org/subscriptions>

---

# RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction

JOSÉ ALMEIDA CRUZ,<sup>1</sup> MARC-FRÉDÉRIC BLANCHET,<sup>2</sup> MICHAL BONIECKI,<sup>3</sup> JANUSZ M. BUJNICKI,<sup>3,4</sup> SHI-JIE CHEN,<sup>5</sup> SONG CAO,<sup>5</sup> RHIJU DAS,<sup>6,7</sup> FENG DING,<sup>8</sup> NIKOLAY V. DOKHOLYAN,<sup>8</sup> SAMUEL COULBOURN FLORES,<sup>9</sup> LILI HUANG,<sup>10</sup> CHRISTOPHER A. LAVENDER,<sup>11</sup> VÉRONIQUE LISI,<sup>2</sup> FRANÇOIS MAJOR,<sup>2</sup> KATARZYNA MIKOLAJCZAK,<sup>3</sup> DINSHAW J. PATEL,<sup>10</sup> ANNA PHILIPS,<sup>3,4</sup> TOMASZ PUTON,<sup>4</sup> JOHN SANTALUCIA,<sup>12,13</sup> FREDRICK SIJENYI,<sup>13</sup> THOMAS HERMANN,<sup>14</sup> KRISTIAN ROTHER,<sup>4</sup> MAGDALENA ROTHER,<sup>4</sup> ALEXANDER SERGANOV,<sup>10</sup> MARCIN SKORUPSKI,<sup>4</sup> TOMASZ SOLTYSINSKI,<sup>3</sup> PARIN SRIPAKDEEVONG,<sup>6,7</sup> IRINA TUSZYNSKA,<sup>3</sup> KEVIN M. WEEKS,<sup>11</sup> CHRISTINA WALDSICH,<sup>15</sup> MICHAEL WILDAUER,<sup>15</sup> NEOCLES B. LEONTIS,<sup>16</sup> and ERIC WESTHOF<sup>1,17</sup>

<sup>1</sup>Architecture et Réactivité de l'ARN, Université de Strasbourg, IBMC-CNRS, F-67084 Strasbourg, France

<sup>2</sup>Institute for Research in Immunology and Cancer (IRIC), Department of Computer Science and Operations Research, Université de Montréal, Montréal, Québec H3C 3J7, Canada

<sup>3</sup>Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland

<sup>4</sup>Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland

<sup>5</sup>Department of Physics and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, USA

<sup>6</sup>Department of Biochemistry, <sup>7</sup>Department of Physics, Stanford University, Stanford, California 94305, USA

<sup>8</sup>Department of Biochemistry and Biophysics, University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27599, USA

<sup>9</sup>Computational & Systems Biology Program, Institute for Cell and Molecular Biology, Uppsala University, 751 05 Uppsala, Sweden

<sup>10</sup>Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

<sup>11</sup>Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599, USA

<sup>12</sup>Department of Chemistry, Wayne State University, Detroit, Michigan 48202, USA

<sup>13</sup>DNA Software, Ann Arbor, Michigan 48104, USA

<sup>14</sup>Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093, USA

<sup>15</sup>Max F. Perutz Laboratories, Department of Biochemistry, University of Vienna, Vienna 1030, Austria

<sup>16</sup>Department of Chemistry and Center for Biomolecular Sciences, Bowling Green State University, Bowling Green, Ohio 43403, USA

## ABSTRACT

We report the results of a first, collective, blind experiment in RNA three-dimensional (3D) structure prediction, encompassing three prediction puzzles. The goals are to assess the leading edge of RNA structure prediction techniques; compare existing methods and tools; and evaluate their relative strengths, weaknesses, and limitations in terms of sequence length and structural complexity. The results should give potential users insight into the suitability of available methods for different applications and facilitate efforts in the RNA structure prediction community in ongoing efforts to improve prediction tools. We also report the creation of an automated evaluation pipeline to facilitate the analysis of future RNA structure prediction exercises.

**Keywords:** 3D prediction; bioinformatics; force fields; structure

## INTRODUCTION

The determination of the atomic structure of any biological macromolecule, RNA molecules being no exception, contributes regularly toward the understanding of the molecular

basis of the underlying biological process. Each of the current experimental methods for determining the three-dimensional (3D) structures of RNA molecules—X-ray crystallography, NMR, and cryo-electron microscopy—requires great expertise and substantial technical resources. Therefore, the ability to reliably predict accurate RNA 3D structures based solely on their sequences, or in concert with efficiently obtained biochemical information, is an important problem and constitutes a major intellectual challenge (Tinoco and Bustamante 1999). Recent decades have seen several

<sup>17</sup>Corresponding author.

E-mail [E.Westhof@ibmc-cnrs.unistra.fr](mailto:E.Westhof@ibmc-cnrs.unistra.fr).

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.031054.111>.

significant theoretical advances toward this goal that include:

1. The development of predictive models for RNA secondary structure, pioneered by the seminal work of Tinoco and coworkers (Tinoco et al. 1973) and made commonly available in a number of tools that perform reasonably well for sequences of moderate size (Hofacker et al. 1994; Zuker 2003; Reuter and Mathews 2010);
2. The ability to meaningfully deduce RNA structures through comparative sequence analysis (Woese et al. 1980; Michel and Westhof 1990);
3. The systematization of the knowledge about RNA architecture and interactions (Leontis and Westhof 2001) to gain a handle on the rapid increase in the number and size of RNA molecules with published structures available in public databases (Berman et al. 2000);
4. The availability of comprehensive sequence alignments (Gardner et al. 2009) permitting the study of the relationship between structure and sequence;
5. The development of improved molecular dynamics force fields and techniques (Ditzler et al. 2010);
6. Finally, the increasing availability of inexpensive computing power and data storage allows for extensive computational searches.

As a consequence, exciting developments in the field of de novo structure prediction have occurred in the last few years: computer-assisted modeling tools (Martinez et al. 2008; Jossinet et al. 2010); conformational space search (Parisien and Major 2008); discrete molecular dynamics (Ding et al. 2008a); knowledge-based, coarse-grained refinement (Jonikas et al. 2009); template-based (Flores and Altman 2010; Rother et al. 2011b); and force-field-based approaches (Das et al. 2010) inspired by proven protein-folding techniques adapted to the RNA field (for review, see Rother et al. 2011a). All these new approaches are pushing the limits of automatic RNA structure prediction from short sequences of a few nucleotides to medium-sized molecules with several dozens. Assuming continuing, steady progress, one can expect that in the near-to-medium future, de novo prediction of RNA 3D structures will become as common and useful as RNA secondary structure prediction is today.

These promising results and the increasing number of available tools raise the need for objective evaluation and comparison. Indeed, the establishment of a benchmark for RNA structure prediction has become essential in order to optimize and improve the current methods and tools for structural prediction. Here, we present the results of a blind exercise in RNA structure prediction. Sequences of RNA structures solved by crystallographers were provided, before publication, to active research groups that develop new methods and perform RNA 3D structure prediction. Comparisons between predicted and experimental X-ray struc-

tures were undertaken once the structures were published. The resulting benchmarks function as a snapshot of the current status of this field. On the basis of this successful first round, we would like to extend to RNA the idea established by the protein structure prediction community (Moult 2006) and to propose a continuous, open, and collective structure prediction experiment, with the essential, active participation of experimentalists.

## RNA-PUZZLES

*RNA-Puzzles* is a collective blind experiment for evaluation of de novo RNA structure prediction. With this initiative, we hope to (1) assess the cutting edge of RNA structure prediction techniques; (2) compare the different methods and tools, elucidating their relative strengths and weaknesses and clarifying their limits in terms of sequence length and structure complexity; (3) determine what has still to be done to achieve an ultimate solution to the structure prediction problem; (4) promote the available methods and guide potential users in the choice of suitable tools for different problems; and (5) encourage the RNA structure prediction community in their efforts to improve the current tools.

The procedure that governs *RNA-Puzzles* is straightforward. Based on the successful first round, we propose the following steps:

- Complete nucleotide sequences will be periodically released to interested groups who agree to keep sequence information confidential. These target sequences correspond to experimentally determined crystallographic or NMR structures, kindly provided by experimental groups, and not yet published in any form. Confidentiality of RNA sequence information is essential to protect the target selection and molecular engineering strategies of participating experimental groups.
- The interested groups will have a specified length of time (usually 4–6 wk) to submit their predicted models to a website in a standard pdb format that respects atom naming and nomenclature conventions.
- The predicted models will be evaluated with regard to stereochemical correctness, topology, and geometrical similarity, relative to the experimental structure.
- After publication of the original experimental structures, all predicted models, experimental results, and comparative data will be made publicly available.

To set up and automate these steps, the *RNA-Puzzles* team has put together a public website for announcing new experiments and publishing results of completed experiments. The website also provides a processing pipeline to carry out model evaluation. The *RNA-Puzzles* site is publicly available at <http://paradise-ibmc.u-strasbg.fr/rnapuzzles/>.

## STRUCTURE ANALYSIS AND COMPARISON

The evaluation of the biological value of a structural model raises many questions. How do you determine if a given model is a meaningful prediction? What is, in fact, a biologically meaningful prediction? Which questions should a structural model answer? Clearly, addressing some questions requires very high precision (1–2 Å or below); whereas, in other cases, important insights may be obtained with residue-level or domain-level precision.

To evaluate the predictive success of the proposed models, we established two general criteria:

1. The predicted model must be geometrically and topologically as close as possible to the experimentally determined structure, used as the reference. It is assumed that the crystal structure or NMR structure is correct within the limitations of the experimental methods.
2. The predicted model must be stereochemically correct (with bond distances and intermolecular contacts close to the experimentally observed values).

To geometrically compare predicted models with the experimental structures, we used the Root Mean Square Deviation (RMSD) measure and the Deformation Index (DI) (Parisien et al. 2009). The RMSD is the usual measure of distance between two superimposed structures defined by the formula:

$$RMSD(A, B) = \sqrt{\frac{\sum_{i=1}^n (a_i - b_i)^2}{n}},$$

in which A and B are the modeled and experimental structures and  $(a_i - b_i)$  represents the distance between the  $i$ -th atoms of the two structures. The DI is given by:

$$DI(A, B) = \frac{RMSD(A, B)}{MCC(A, B)},$$

in which MCC is the Matthews Correlation Coefficient (Matthews 1975) computed on the individual base-pair and base-stacking predictions. The reason for this choice is that the RMSD, as a measure of similarity, does not account for specific RNA features such as the correctness of base-pair and stacking interactions. The DI score complements the RMSD values by introducing those specific features in the metric. Using the DI value, the quality of two models with close RMSDs can be discriminated according to the accuracy of their predictions of the base-pairing and stacking interactions of the experimental structure. As we observed in the first experiments, the ranking of the models is sensitive to the chosen metric (see Tables 1–3). Such observations were also made during the CASP competitions of protein

structure prediction (see Marti-Renom et al. 2002). Use of diverse, complementary metrics should contribute to the design of improved metrics and an understanding of their relative strengths and limitations.

In a recent work (Hajdin et al. 2010), Weeks, Dokholyan, and coworkers showed that when sampling the conformational space of an RNA molecule using discrete molecular dynamics, the RMSD values are distributed normally with a mean related to the length of the molecule by the power law:

$$\langle RMSD \rangle = a \times N^{0.41} - b$$

Here,  $N$  is the number of nucleotides, and  $a$  and  $b$  are constants that depend on whether secondary structure information is provided as input to the molecular dynamic simulation. From this observation, it is possible to compute the significance level ( $P$ -value) of a prediction with a given RMSD with respect to an accepted structure. This  $P$ -value corresponds to the probability that a given structure prediction is better than that expected by chance (Hajdin et al. 2010). Structure models with  $P$ -values  $<0.01$  represent, in general, successful predictions of a global RNA fold. The  $P$ -value is sensitive to the amount of preexisting information available for a given structure prediction problem, especially whether the pattern of base-pairing is known in advance. For most structure prediction problems, much or all of the secondary structure is known and is used as a constraint during structure prediction and refinement. In this round of *RNA-Puzzles*,  $P$ -value analysis was appropriate for Puzzle 3 (below).

The stereochemical correctness of the predicted models was evaluated with MolProbity (Davis et al. 2007), which provides quality validation for 3D structures of nucleic acids. MolProbity performs several automatic analyses, from checking the lengths of H-bonds present in the model to validating the compliance with the rotameric nature of the RNA backbone (Murray et al. 2003). The reduce-build script of MolProbity was used for adding hydrogen atoms to the heavy atoms of the models. As a single measure of stereochemical correctness, we chose the *clash score*, i.e., the number of steric clashes per thousand residues (Word et al. 1999).

All of the computed values are shown in a comparison summary page, which ranks the submitted models according to each of the computed metrics. In addition to the comparison summary, we provide a report for each of the predicted models. The report presents the structural superposition between predicted model and experimental structure, the analysis of the predicted base pairs—correctly predicted (true positives), incorrectly predicted (false positives), and missed (false negatives)—and a complete Deformation Profile matrix (DP) (Parisien et al. 2009), which provides an evaluation of the predictive quality of a model at multiple scales.

## THE PROBLEMS

Two crystallography laboratories sent coordinates for the prediction contest: the laboratories of Thomas Hermann at UC San Diego and of Dinshaw Patel and Alexander Serganov at the Memorial Sloan-Kettering Cancer Center. The three trial experiments were the following.

### Problem 1: Dimer

Predict the structure of the following sequence:



knowing that the crystal structure shows a homodimer that contains two strands of the sequence that hybridize with blunt ends (C-G closing base pairs). The solution structure corresponds to the regulatory element from human thymidylate synthase mRNA (Dibrov et al. 2011a), which, in the crystal, forms a dimer with two asymmetrical internal loops, despite perfect sequence symmetry (Fig. 1A,B). The crystal structure was resolved to 1.97 Å resolution. A total of 14 predicted models were submitted with an RMSD ranging from 3.41 Å to 6.94 Å (mean RMSD of 4.7 Å) (Table 1).

### Problem 2: Square

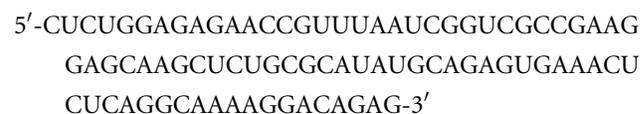
The crystal structure, which was resolved to 2.2 Å resolution, shows a 100-nt square of double-stranded RNA that self-assembles from four identical inner and four identical outer strands (Dibrov et al. 2011b). The secondary structure

shown was used for the design of the square. Actual base-pairing in the crystal may deviate. 3D coordinates of the nucleotides in the inner strands (B, D, F, H) were provided. What are the structures of the outer strands (A, C, E, G)?

The square is formed by four helices connected by four single-stranded loops. All of the helices are identical at the sequence level, and so are all the loops (Fig. 2).

### Problem 3: A riboswitch domain

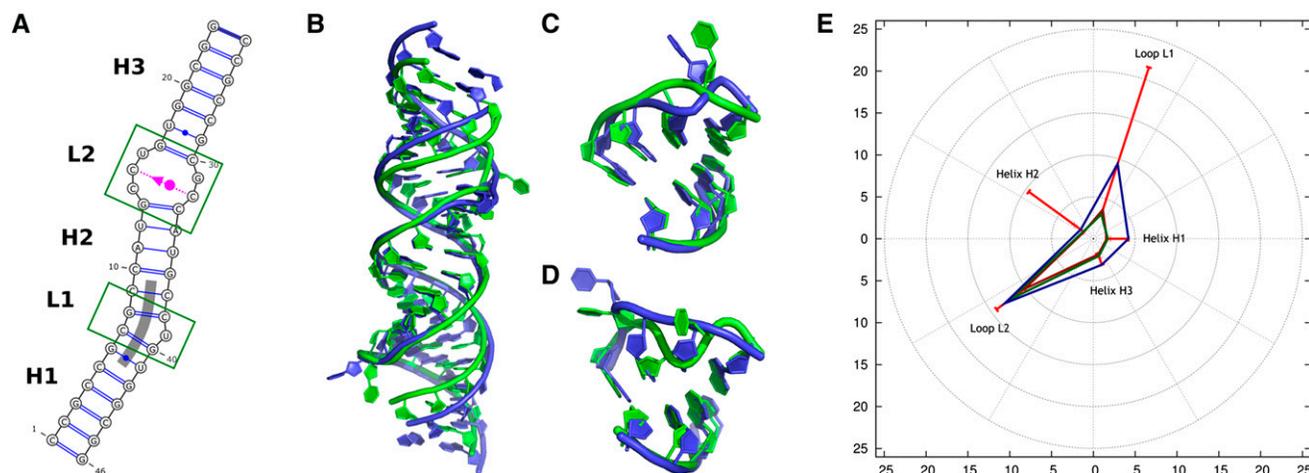
A riboswitch domain was crystallized. The sequence is the following:



The crystallized sequence was slightly different (an apical loop was replaced by a GAAA loop), but this detail of RNA crystal engineering was not disclosed to modelers to protect the crystallographers (Fig. 3A,B; Huang et al. 2010).

## RESULTS

Eight research groups participated in the *RNA-Puzzles* experiments. The Bujnicki group used a hybrid strategy previously developed for protein modeling in the course of the CASP experiment (Kosinski et al. 2003). The Chen lab used a multi-scale, free energy landscape-based RNA folding model (Vfold model) (Chen 2008; Cao and Chen 2011). The Das group used the stepwise assembly (SWA) method for



**FIGURE 1.** Problem 1—An RNA dimer. (A) Secondary structure of the reference RNA molecule. Note that the structure is symmetric on the sequence level but is asymmetric in the crystal, indicating that crystal-packing forces played a significant role in the conformation of this RNA. The interaction (in magenta) was detected with the RNAView (Yang et al. 2003) annotation program but not with MC-Annotated (Gendron et al. 2001). (Thick gray band) Coaxial stacking between helices. X-ray structures of the reference RNA molecule (green) and the lowest RMSD predicted models (blue) for the complete Das model 3 (B), details of loop L1 of Das model 1 (C), and details of loop L2 of Das model 3 (D). (E) Deformation Profile values for each of the five domains of the homodimer. (Colored lines) The DP values for the two predicted models with lowest RMSD, Das model 3 (dark red), and Das model 1 (dark green), and for the predicted model with highest RMSD, Dokholyan model 1 (dark blue). (Radial red lines) The minimum, maximum, and mean DP values for each domain.

**TABLE 1.** Summary of the results for Puzzle 1

Problem 1 Group <sup>a</sup>	Number <sup>b</sup>	RMSD <sup>c</sup>	Rank <sup>d</sup>	DI all <sup>e</sup>	Rank <sup>d</sup>	INF all <sup>f</sup>	Rank <sup>d</sup>	INF wc <sup>g</sup>	Rank <sup>d</sup>	INF stack <sup>h</sup>	Rank <sup>d</sup>	Clash Score <sup>i</sup>	Rank <sup>d</sup>	L1 <sup>j</sup>
Das	3	3.41	1	3.66	1	0.93	1	0.95	2	0.92	1	0.00	5	x
Das	1	3.58	2	3.89	2	0.92	3	0.95	1	0.91	2	0.00	3	x
Das	4	3.91	3	4.31	3	0.91	4	0.91	8	0.91	4	0.00	4	
Major	1	4.06	4	4.57	4	0.89	5	0.95	6	0.87	5	66.40	11	
Chen	1	4.11	5	5.01	6	0.82	9	0.87	11	0.80	8	0.68	6	
Das	2	4.34	6	4.70	5	0.92	2	0.95	4	0.91	3	1.36	7	x
Das	5	4.56	7	5.36	7	0.85	7	0.88	10	0.84	7	0.00	2	
Bujnicki	3	4.66	8	5.75	9	0.81	11	0.95	3	0.74	14	54.73	10	x
Bujnicki	4	4.74	9	6.59	11	0.72	14	0.65	14	0.75	13	83.33	14	
Bujnicki	5	4.89	10	6.26	10	0.78	13	0.78	13	0.80	9	81.98	13	
Bujnicki	1	5.07	11	5.75	8	0.88	6	0.93	7	0.86	6	0.00	1	x
Bujnicki	2	5.43	12	6.75	12	0.80	12	0.90	9	0.77	12	71.57	12	x
Santalucia	1	5.69	13	6.75	13	0.84	8	0.95	5	0.79	11	39.86	9	
Dokholyan	1	6.94	14	8.55	14	0.81	10	0.86	12	0.79	10	31.74	8	
Mean		4.67		5.56		0.85		0.89		0.83				
Standard deviation		0.93		1.34		0.06	N	0.09		0.07				
X-Ray Model													1.35	

Values in each row correspond to a predicted model.

<sup>a</sup>Name of the research group that submitted the model.

<sup>b</sup>Number of the model among all models from the same group.

<sup>c</sup>RMSD of the model compared with the accepted structure.

<sup>d</sup>Columns indicate the rank of the model with respect to the left-hand column metric.

<sup>e</sup>DI<sub>all</sub> is the Deformation Index taking into account all interactions (stacking, Watson-Crick, and non-Watson-Crick).

<sup>f</sup>INF<sub>all</sub> is the Interaction Network Fidelity taking into account all interactions.

<sup>g</sup>INF<sub>wc</sub> is the Interaction Network Fidelity taking into account only Watson-Crick interactions.

<sup>h</sup>INF<sub>stack</sub> is the Interaction Network Fidelity taking into account only stacking interactions.

<sup>i</sup>Clash Score as computed by the MolProbity suite (Davis et al. 2007).

<sup>j</sup>An "x" in this column indicates models that correctly predict base-pair interactions in loop L1. No model correctly predicted all interactions in loop L2.

recursively constructing atomic-detail biomolecular structures in small building steps (Sripakdeevong et al. 2011). The Dokholyan group adopted a multi-scale molecular dynamics approach (Ding and Dokholyan 2012). The Flores group used the RNABuilder program, a computer-assisted RNA modeling tool (Flores and Altman 2010). The Major group applied the fully automated MC-Fold and MC-Sym pipeline they developed (Parisien and Major 2008). The Santalucia group also applied their own program.

The amount of time required to produce the models and the degree of automation varied as a function of approach. One point should be emphasized. Compared with CASP protein targets, an RNA puzzle typically involves multiple "mini-puzzles" such as separate tertiary modules and specific non-Watson-Crick pairs. There are several examples of this from this first round, for example, the four corners and the four helices of the nanosquare. Thus, a single RNA puzzle can provide multiple challenges for testing modeling methods.

### Problem 1: Dimer

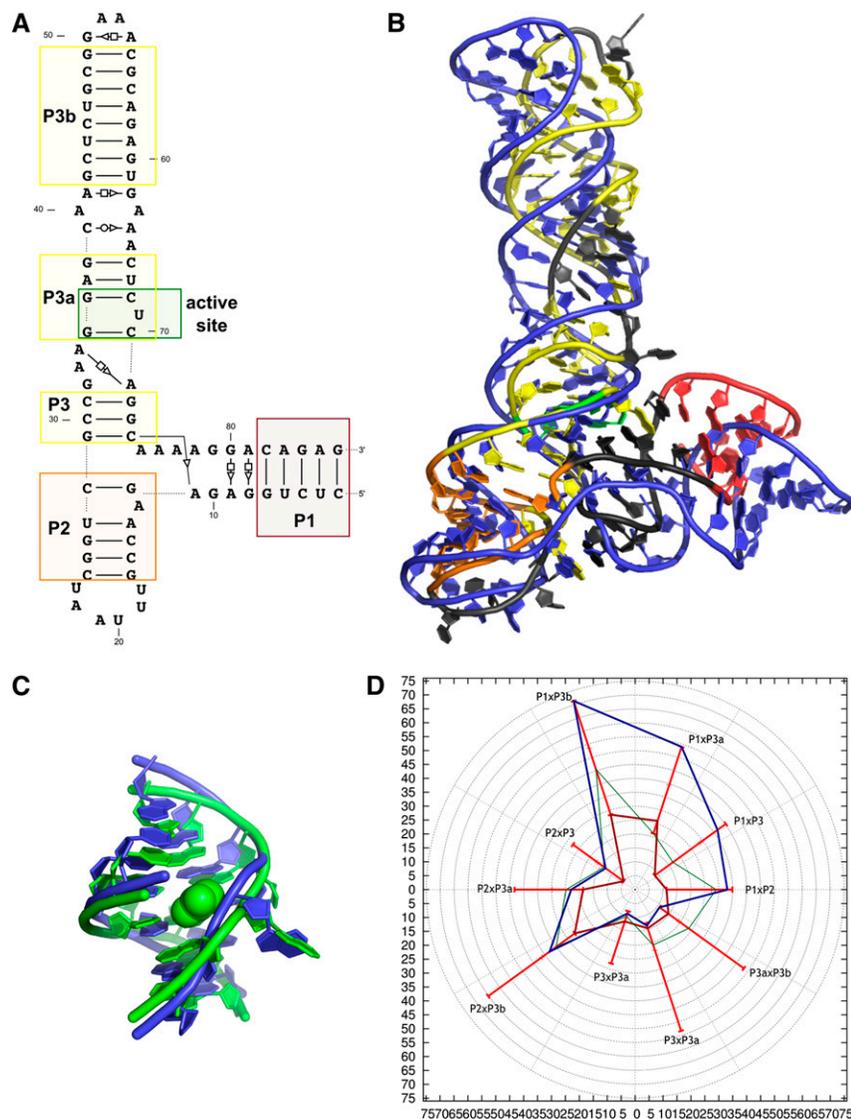
Fourteen predicted models were submitted. The RMSDs range from 3.41 Å to 6.94 Å (with a mean of 4.67 Å). The base-pair interactions were correctly predicted in almost all

models with >85% of WC base pairs correctly predicted in all but two models and >75% of stacking interaction predicted in all but one model. Contrary to the X-ray structure, most of the proposed models present a symmetric structure. The only exceptions were the models from the Das laboratory (see Table 1). From the analysis of the Deformation Profile values (Fig. 1E), it is clear that the internal loops were the domains most difficult to predict (Fig. 1C,D) and that helix H2, probably because of its location between the loops, presents a particularly large interval of DP values. Several models present high values for the Clash Score, which could reflect the need for updated dictionaries of distances and angles or stronger constraints toward the dictionary values.

### Problem 2: Square

Thirteen predicted models were submitted with RMSDs ranging from 2.3 Å to 3.65 Å (mean RMSD of 2.9 Å) (see Table 2). The RMSDs of solutions to this problem are the lowest of all three problems, which is expected because one-half of each base pair was provided in the initial puzzle description. As expected, the helical regions are better predicted than loops, with mean DP values between 5 and 10 for all loops and <5 for three of the helices (Fig. 2C–E),





**FIGURE 3.** Problem 3—A riboswitch domain. (A) Secondary structure of the reference RNA molecule. (B) X-ray structure of the reference RNA molecule (P1 [red]; P2 [orange]; P3, P3a, P3b [yellow]; active site [green]) and the predicted model of the lowest RMSD Chen model 1 (blue). (C) Detail of the ligand binding pocket for the X-ray structure (green) and the lowest RMSD Chen model 1 (blue). (D) Deformation Profile values of the pairwise helical interdomain regions (P1, P2, P3, P3a, and P3b) for the three models with lowest RMSDs: Chen model 1 (dark red), Dokholyan model 2 (dark green), and Das model 5 (dark blue). (Radial red lines) The minimum, maximum, and mean DP values for each interdomain pair.

*RNA-Puzzles* will depend critically on engagement by the prediction community and the generosity of the experimental community. Most importantly, this work will, hopefully, convince additional structural biology groups to offer problems to the modeling community in the future.

This first contest had clear limitations, and several improvements have already been planned. (1) As established for CASP, in the future, modelers will be asked to predict the deviations of their own models from the unknown native structure, at the level of individual residues or atoms (in angstroms). These values could, for example, be

compactly encoded in the B-factor field of PDB atom records. The number of submissions should be limited and multiple submissions ranked by the authors. (2) In addition, it will be worthwhile to improve model scoring and ranking so as to produce an absolute ranking of all models, taking into account local and global model quality. (3) Because the RNA structure database continues to grow, template-based methods are becoming increasingly important, and, consequently, future RNA puzzles should also include structures of homologs of existing folds (for example, a riboswitch with an alternative ligand or mutation). (4) Finally, we plan to extend the contest to include structures of RNA-protein complexes.

The assessment of model accuracy requires reliable and meaningful metrics for comparisons between the models and the experimentally determined structures used as the accepted structure. In addition to the metrics currently used (generic to all macromolecules or specific for RNA), it may be worthwhile to include metrics that have been shown to perform very well at both global and local levels for assessing the very wide range of model qualities (from very inaccurate to very accurate) (Zemla 2003; Zhang and Skolnick 2004), as have been generally accepted in the protein structure prediction field and are used by assessors in the CASP challenges. We are hopeful that, with extensive community support, this round of *RNA-Puzzles* is the first of what will become a vigorous and ongoing discussion of the frontiers of RNA structure prediction and refinement.

## MATERIALS AND METHODS

The following provides a brief description of the methodology used by the modeling groups, together with comments.

### Bujnicki group

The Bujnicki group used a hybrid strategy previously developed for protein modeling in the course of the CASP experiment (Kosinski et al. 2003). Briefly, initial models were constructed by template-based modeling and fragment assembly, with constraints on secondary structure, using the comparative RNA modeling tool ModeRNA (Rother et al. 2011b). For RNA Puzzle Problem 2, the secondary

**TABLE 2.** Summary of the results for Puzzle 2

Problem 2															
Group <sup>a</sup>	Number <sup>b</sup>	RMSD <sup>c</sup>	Rank <sup>d</sup>	DI all <sup>e</sup>	Rank <sup>d</sup>	INF all <sup>f</sup>	Rank <sup>d</sup>	INF wc <sup>g</sup>	Rank <sup>d</sup>	INF nwc <sup>h</sup>	Rank <sup>d</sup>	INF stack <sup>i</sup>	Rank <sup>d</sup>	Clash Score <sup>j</sup>	Rank <sup>d</sup>
Bujnicki	2	2.3	1	2.83	1	0.81	8	0.92	9	0	13	0.79	7	14.54	2
Bujnicki	3	2.33	2	2.9	3	0.8	10	0.91	10	0	2	0.77	9	0.62	1
Das	1	2.5	3	2.9	2	0.86	2	0.96	5	0	8	0.85	2	19.8	5
Dokholyan	1	2.54	4	3.09	5	0.82	6	0.9	11	0	1	0.8	5	19.85	6
Bujnicki	1	2.65	5	2.99	4	0.89	1	0.96	4	0	3	0.86	1	15.47	3
Chen	1	2.83	6	3.74	9	0.76	13	0.9	12	0	9	0.69	13	18.66	4
Das	4	2.83	7	3.46	6	0.82	7	0.97	3	0	12	0.78	8	23.82	8
Major	1	2.98	8	3.82	10	0.78	12	0.95	7	0	10	0.71	12	134.26	12
Das	3	3.03	9	3.67	7	0.83	5	0.97	1	0	6	0.8	6	25.37	10
Das	2	3.05	10	3.69	8	0.83	4	0.97	2	0	7	0.81	3	23.51	7
Das	5	3.46	11	4.18	11	0.83	3	0.96	6	0	11	0.81	4	24.75	9
Flores	1	3.48	12	4.4	12	0.79	11	0.89	13	0	5	0.77	10	165.57	13
Santalucia	1	3.65	13	4.54	13	0.81	9	0.92	8	0	4	0.75	11	25.73	11
Mean		2.90		3.55		0.82		0.94		0.00		0.78			
Standard deviation		0.44		0.59		0.03		0.03		0.00		0.05			
													X-Ray Model	36.10	

Values in each row correspond to a predicted model.

<sup>a</sup>Name of the research group that submitted the model.

<sup>b</sup>Number of the model among all models from the same group.

<sup>c</sup>RMSD of the model compared with the accepted structure.

<sup>d</sup>Columns indicate the rank of the model with respect to the left-hand column metric.

<sup>e</sup>DI<sub>all</sub> is the Deformation Index taking into account all interactions (stacking, Watson-Crick, and non-Watson-Crick).

<sup>f</sup>INF<sub>all</sub> is the Interaction Network Fidelity taking into account all interactions.

<sup>g</sup>INF<sub>wc</sub> is the Interaction Network Fidelity taking into account only Watson-Crick interactions.

<sup>h</sup>INF<sub>nwc</sub> is the Interaction Network Fidelity taking into account only non-Watson-Crick interactions.

<sup>i</sup>INF<sub>stack</sub> is the Interaction Network Fidelity taking into account only stacking interactions.

<sup>j</sup>Clash Score as computed by the MolProbity suite (Davis et al. 2007).

structure was provided by the organizers, while for Problem 3, it was calculated as a consensus of more than 20 methods using the RNA metaserver (<http://genesilico.pl/rnametaserver/>). The initial models were expected to possess approximately correct Watson-Crick base-pairing and stacking interactions within individual structural elements, but their mutual orientations and tertiary contacts required optimization.

The initial models were subjected to global refinement using SimRNA, a de novo RNA folding method (Rother et al. 2012), which was inspired by the REFINER method for protein folding (Boniecki et al. 2003). SimRNA uses a coarse-grained representation, with only three centers of interaction per nucleotide residue. The backbone is represented by atoms P of the phosphate group and C4' of the ribose moiety, whereas the base is represented by just one nitrogen atom of the glycosidic bond (N9 for purines or N1 for pyrimidines). The remaining atoms are neglected. This simplified representation allows reproducing the main characteristics of the RNA molecule such as base-pairing and stacking, and the backbone in helix, while significantly lowering the computational cost for conformational transitions and energy calculations. As an “energy” function, SimRNA uses a statistical potential derived from frequency distributions of geometrical properties observed in experimentally determined RNA structures. Terms of the SimRNA energy function (for the virtual bond lengths, flat and torsion angles, pairwise interactions between the three atom types) were generated using reverse Boltzmann statistics. For searching the conformational space, SimRNA uses Monte Carlo dynamics

controlled by an asymmetric Metropolis method (Metropolis and Ulam 1949) that accepts or rejects new conformations depending on the energy change associated with the conformational change, with the probability of acceptance depending on the temperature of the system. Simulations can be run in the isothermal or energy minimization (simulated annealing) mode, or in conformation space search mode (replica exchange). While SimRNA allows for simulations that use only the sequence information, starting from an extended structure, it can accept user-defined starting structures and restraints that specify distances or allowed distance ranges for user-defined atom pairs. For *RNA-Puzzles*, the Bujnicki group used restraints on secondary structure that allowed the predicted base pairs to be maintained. Following a series of simulations, lowest-energy structures were selected for the final refinement.

The final models were built by first reconstructing the full-atom representation using RebuildRNA (P Lukasz, M Boniecki, and JM Bujnicki, unpubl.) and then optimizing atomic detail of selected residues with SCULPT (Surlis et al. 1994) and HyperChem 8.0 (Hypercube Inc.). For Problem 2, the known coordinates of four strands were used as provided by the organizers and “frozen” at the optimization stage.

The computer calculation time (on a single processor) was as follows: ModeRNA: <2 h; SimRNA and RebuildRNA: ~150 h; SCULPT <1 h; HyperChem: ~12 h.

In the case of the Bujnicki group, the ratio of human to computer time was relatively large (approximately equal), because

**TABLE 3.** Summary of the results for Puzzle 3

Problem 3																	
Group <sup>a</sup>	Number <sup>b</sup>	RMSD <sup>c</sup>	Rank <sup>d</sup>	DI all <sup>e</sup>	Rank <sup>d</sup>	INF all <sup>f</sup>	Rank <sup>d</sup>	INF wc <sup>g</sup>	Rank <sup>d</sup>	INF nwc <sup>h</sup>	Rank <sup>d</sup>	INF stack <sup>i</sup>	Rank <sup>d</sup>	Clash Score <sup>j</sup>	Rank <sup>d</sup>	P-value <sup>k</sup>	Rank <sup>d</sup>
Chen	1	7.24	1	9.84	1	0.74	2	0.86	5	0	6	0.73	1	1.1	3	2.01E-05	1
Dokholyan	2	11.46	2	16.1	2	0.71	6	0.82	9	0	9	0.71	6	41.21	10	3.90E-02	2
Das	5	11.97	3	16.42	3	0.73	5	0.9	1	0.36	5	0.71	3	1.1	4	6.92E-02	3
Bujnicki	1	12.19	4	17.49	5	0.7	7	0.82	10	0	10	0.7	7	14.72	8	8.71E-02	4
Das	2	12.2	5	16.6	4	0.74	3	0.86	6	0.4	2	0.73	2	0.74	2	8.83E-02	5
Major	2	13.7	6	23.33	10	0.59	11	0.67	11	0	8	0.61	10	93.52	12	3.03E-01	6
Bujnicki	2	14.06	7	22.51	7	0.62	10	0.83	8	0	7	0.59	11	5.15	7	3.75E-01	7
Das	1	15.48	8	20.9	6	0.74	1	0.87	4	0.57	1	0.71	5	0	1	6.81E-01	8
Dokholyan	1	15.92	9	23.28	9	0.68	9	0.9	2	0	12	0.66	9	39.37	9	7.629E-01	9
Das	3	16.95	10	23.17	8	0.73	4	0.89	3	0.4	3	0.71	4	1.47	5	9.02E-01	10
Das	4	18.3	11	26.55	11	0.69	8	0.85	7	0.38	4	0.67	8	2.21	6	9.79E-01	11
Major	1	22.99	12	45.27	12	0.51	12	0.39	12	0	11	0.59	12	75.11	11	1.00E+00	12
Mean		14.37		21.79		0.68		0.80		0.18		0.68					
Standard deviation		3.99		8.69		0.07		0.14		0.22		0.05					
													X-Ray Model	1.83			

Values in each row correspond to a predicted model.

<sup>a</sup>Name of the research group that submitted the model.

<sup>b</sup>Number of the model among all models from the same group.

<sup>c</sup>RMSD of the model compared with the accepted structure.

<sup>d</sup>Columns indicate the rank of the model with respect to the left-hand column metric.

<sup>e</sup>DI<sub>all</sub> is the Deformation Index taking into account all interactions (stacking, Watson-Crick, and non-Watson-Crick).

<sup>f</sup>INF<sub>all</sub> is the Interaction Network Fidelity taking into account all interactions.

<sup>g</sup>INF<sub>wc</sub> is the Interaction Network Fidelity taking into account only Watson-Crick interactions.

<sup>h</sup>INF<sub>nwc</sub> is the Interaction Network Fidelity taking into account only non-Watson-Crick interactions.

<sup>i</sup>INF<sub>stack</sub> is the Interaction Network Fidelity taking into account only stacking interactions.

<sup>j</sup>Clash Score as computed by the MolProbity suite (Davis et al. 2007).

<sup>k</sup>Significance of the predicted model, assuming that base-pairing was input as a structural constraint (Hajdin et al. 2010).

the RNA-Puzzles experiment was taken as an opportunity for training in the use of various modeling methods, in a spirit very similar to the collective work of that group during the CASP5 modeling season (Kosinski et al. 2003). Consequently, a large fraction of human time involved discussions and communication between the two parts of the team physically located in two different cities (Poznan and Warsaw). The human time devoted to interactions with software (preparation of input files, setting up simulations, analyses of output files, and manual refinement

using the graphical user interfaces of SCULPT and HyperChem) was ~30 h, with the majority of time devoted to Problems 2 and 3.

### Chen group

The Chen group used a multi-scale approach to predict the RNA 3D structure from the sequence (Cao and Chen 2011). For a given RNA sequence, they first predict the 2D structure from the free

**TABLE 4.** Pairwise interdomain Deformation Profile values for the helical domains P1, P2, P3, P3a, and P3b from Puzzle 3

	P1xP2	P1xP3	P1xP3a	P1xP3b	P2xP3	P2xP3a	P2xP3b	P3xP3a	P3xP3b	P3axP3b
3_bujnicki_1.dat	19.8	22.5	37.6	46.2	<b>7.0</b>	22.2	41.2	<b>11.2</b>	27.2	17.7
3_bujnicki_2.dat	27.3	37.3	49.4	68.2	20.1	32.3	47.4	<b>8.1</b>	21.5	16.0
3_chen_1.dat	<b>11.3</b>	<b>8.7</b>	26.0	28.3	<b>5.1</b>	18.6	26.8	<b>12.0</b>	<b>14.6</b>	<b>14.8</b>
3_das_1.dat	31.7	36.6	48.9	71.2	18.3	35.6	65.1	<b>11.9</b>	30.1	18.9
3_das_2.dat	30.7	32.9	34.1	34.0	24.9	32.9	27.0	<b>10.4</b>	<b>12.6</b>	<b>13.4</b>
3_das_3.dat	29.9	33.7	43.9	59.1	23.2	35.9	45.1	<b>13.3</b>	21.9	<b>13.5</b>
3_das_4.dat	33.1	36.5	54.0	71.3	<b>13.1</b>	22.9	37.9	<b>9.1</b>	<b>13.6</b>	<b>10.8</b>
3_das_5.dat	30.4	34.2	39.1	45.6	25.9	35.1	43.4	<b>8.9</b>	<b>13.5</b>	<b>11.8</b>
3_dokholyan_1.dat	34.9	21.5	32.9	59.4	<b>12.1</b>	28.3	32.9	<b>14.7</b>	26.8	25.9
3_dokholyan_2.dat	29.0	16.8	21.2	45.5	<b>13.8</b>	24.8	35.0	<b>10.1</b>	20.9	23.8
3_major_1.dat	23.0	40.3	44.4	46.4	27.5	43.3	56.5	27.9	53.4	48.2
3_major_2.dat	27.6	26.8	44.2	66.1	<b>9.6</b>	25.1	37.9	<b>10.4</b>	19.4	18.5

All DP values <15 are in boldface.

**TABLE 5.** The RMSD values for the ligand binding site of each predicted model from Puzzle 3, relative to the crystal structure

	Problem 3
3_das_5	2.842888
3_das_4	2.928573
3_bujnicki_2	3.042605
3_chen_1	3.703777
3_das_2	3.915769
3_dokholyan_2	4.138209
3_bujnicki_1	4.253633
3_major_2	4.447882
3_das_1	4.554707
3_das_3	4.681876
3_dokholyan_1	5.821877
3_major_1	17.289223

energy landscape using the Vfold model (Cao and Chen 2005, 2006a,b, 2009; Chen 2008). The Vfold model allows for the computation of the free energies for the different RNA secondary structures and pseudoknotted structures, from which the (low-free energy) folds can be predicted. Distinguished from other existing models, the Vfold model is based on a virtual bond (coarse-grained) structural model that enables direct evaluation of the entropy parameters for different RNA motifs, including pseudoknotted structures.

The Vfold-based approach to the evaluation of the entropy and the free energy may lead to more reliable 2D structure prediction. For the calculation of 2D structures, the base-stacking energies are adopted from the Turner energy rules (Serra et al. 1994). For the loops, the model enumerates all the possible intraloop mismatched base stacks and evaluates the free energy for each structure. Intraloop base stacks cause large entropic decrease. The parameters for such entropic decrease can be estimated by a theory such as the Vfold model. Next, a 3D coarse-grained scaffold is constructed, based on the predicted 2D structure. In the coarse-grained structure, three atoms (P, C<sub>4</sub>, N<sub>1</sub> or N<sub>9</sub>) are used to represent each nucleotide. To construct a 3D scaffold, the predicted helix stems are modeled by A-form helices. For the loops/junctions, 3D fragments from the known PDB database were used. Specifically, a structural template database was built by classifying the structures according to the different motifs such as hairpin loops and internal/bulge loops, three-way junctions, four-way junctions, pseudoknots, etc. Then the optimal structural templates for the predicted loops/junctions were searched from the structural template database. The structural templates may partly account for the tertiary contacts ignored in the 2D structure prediction. Third, the Chen group build the all-atom model from the coarse-grained scaffold by adding the bases to the virtual bond backbone. In the final step, they refine the all-atom structure by using AMBER energy minimization. Two thousand steps of minimization were run, applying 500.0 kcal/mol restraints to all the residues, followed by another 2000 steps of minimization without restraints. The final structure after minimization is the one submitted for evaluation in the structure prediction test.

In summary, the computation involves two steps: (a) the prediction of the 2D structure and the construction of the coarse-grained 3D structure and (b) AMBER energy minimization. The computer times (T<sub>a</sub>, T<sub>b</sub>) for the two steps are (<1 min,

53 min), (<1 min, 81 min), and (26 min, 143 min), for the predictions of the dimer, the square, and the riboswitch domain, respectively. The first step calculation was performed on a desktop PC with Intel Core 2 Duo CPU E8400 at 3.00 GHz, and the second step computation was performed on a Dell EM64T cluster (Intel Xeon 5150 at 2.66 GHz).

For predicting the dimer and the riboswitch structure (Problems 1 and 3), the Chen group only relied on the sequence information, and the 3D structures were generated by computer, with no human interference in the process. For the prediction of the square structure (Problem 2), they used the experimentally determined structure for one strand to refine the other strand. The loops and the secondary structure of the square were computer-predicted using the Vfold model (Cao and Chen 2011).

## Das group

The Das group used a newly developed ab initio method called *stepwise assembly* (SWA) for recursively constructing atomic-detail biomolecular structures in small building steps. Each step involved enumerating several million conformations for each monomer, and all step-by-step build-up paths were covered in polynomial computational time. The method is implemented in Rosetta and uses the physically realistic, Rosetta all-atom energy function (Das and Baker 2008; Das et al. 2010). The Das group has recently benchmarked SWA on small RNA loop-modeling problems (Sripakdeevong et al. 2011). They also applied de novo fragment assembly with full-atom refinement (FARFAR, also implemented in Rosetta), but did not submit those solutions because they either agreed with the SWA models (Problem 1; parts of Problems 2 and 3) or did not give converged solutions (other parts of Problems 2 and 3) (Das et al. 2010).

Due to the deterministic and enumerative nature of SWA, the computational expense is high relative to stochastic and knowledge-based methods. The computational expense ranged from 20,000 (Problem 1) to 50,000 CPU-hours (Problems 2 and 3). Also, because the code was developed “on-the-fly,” there was no time to fully optimize the run-time, which is being done now.

The SWA modeling runs were fully automated. Manual input was used near the beginning to set up the runs, and near the end to ensure that models presented a diversity of base-pairing patterns—both of these steps could be easily automated, but the Stepwise Assembly method was still under development during the course of this community-wide, blind RNA prediction experiment.

The lessons learned from the three models are the following.

### Problem 1

SWA models 1 and 3 (out of five submitted) performed reasonably well on the 46-nt homodimer, especially at the 9-nt L1 region (see Fig. 1C). Both models correctly predicted the non-canonical *cis* WC/WC C9–C37 base pair and the extrahelical bulge at U39. This accuracy was aided by a strategy that gave entropic bonuses to bulged nucleotides that make no other interactions; the bulges are “virtualized” within Rosetta (Sripakdeevong et al. 2011). In this L1 region, both models gave 1.0 Å all-heavy-atom RMSD to the crystallographic model, excluding the U39 extrahelical bulge. In contrast, none of the five SWA models achieved atomic accuracy in the sequence-identical 9-nt L2 loop (see Fig. 1D; >3.0 Å RMSD). Model 3 did correctly predict C14 to be an

extrahelical bulge and C15 and C32 to be base-paired. However, the exact geometry of the predicted C15–C32 base pair and an additional U16–G31 base pair were incorrect. In the crystallographic model, the base of U16 bulged out and its phosphate formed hydrogen-bonding interactions with the base of G31; in the Das group implementation at the time, they “virtualized” the phosphates of any bulged nucleotides along with their bases. Partial virtualization of bulged bases and more rigorous modeling of conformational entropy are under investigation.

### Problem 2

The SWA models performed well in the regions of the 100-nt “self-assembling RNA square” within putatively regular secondary structure. The Das group did not assume these to be ideal A-form helices but modeled them from scratch. These regions are composed mainly of Watson-Crick base pairs but also included a non-canonical *cis* WC/WC base pair at corner E/F (see Fig. 2C) that SWA model 1 correctly predicted. In contrast, the SWA models did not reach atomic accuracy for any of the 5-nt loops at each of the four corners of the square RNA. This was partly expected because the “corners” of the nanosquare originated from a 5-nt bulge in HCV IRES domain IIa (PDB number: 2PN4) (Zhao et al. 2008), which happened to be part of our comprehensive SWA benchmark (Sripakdeevong et al. 2011). There, it was possible to sample the crystallographic loop conformation but not to select it as the lowest-energy structure; the loop forms direct hydrogen bonds to metal ions, and these interactions are not yet modeled in Rosetta.

With this result in mind, after the nanosquare crystal structure was released, it was compared with the full ensemble of models generated by SWA. Loops in corners C/D and G/H were engaged in significant crystal contacts; but loops A/B and E/F should have been amenable to high-accuracy modeling. Indeed, for both of these loops, SWA sampled the crystallographic conformation of these loop regions with a  $<1.0$  Å RMSD, but these models had significantly worse Rosetta energy than the submitted ones. Again, these corners (and indeed all four corners) involved the binding of either divalent metal ions or cobalt hexamine (III). The lesson learned (or verified) from this puzzle is that approximations in the Rosetta all-atom energy function, especially with regard to metal ions, still remain too inaccurate to permit atomic-resolution RNA modeling on a consistent basis. This puzzle has inspired us to develop approaches to include metal ions during the *de novo* buildup of models.

### Problem 3

The Das group’s recent research has focused on the prediction of high-resolution motifs as stepping stones to modeling larger RNAs. This glycine riboswitch puzzle was thus currently out of range—its core three-way junction and glycine binding site form an intricate noncanonical pairing network involving more than a dozen residues. Furthermore, interactions across a dimer interface appear crucial for stabilizing the riboswitch conformation, but this information was not available. The group’s models were based on generating low-energy Rosetta SWA models for individual loops, two-way junctions, and three-way junctions, and then connecting them with ideal helices. Surprisingly, this basic approach, ignorant of higher-order interactions, gave the best base-pair recoveries (INF all, INF wc, INF nwc; see Models 1 and 2 in Table 3) among submitted models. Other submitted models (Models 4 and 5) gave the best RMSDs for the glycine-binding

site. However, these were very far from atomic accuracy (2.8 Å and 2.9 Å). Most critically, the global structure of the RNA was not recapitulated (RMSD and DI) (Table 3). The helices formed the correct tuning-fork-like rearrangement but were twisted relative to the crystallographic model (Tables 4, 5). Globally correct solutions require global optimization, and this puzzle has motivated the group to develop iterative hybrid high-resolution/low-resolution approaches to RNA modeling, analogous to the rebuild-and-refine method used in Rosetta template-based modeling (Qian et al. 2007). As a final note, in the article describing this puzzle’s crystal structure, a striking structural similarity of the glycine riboswitch core to a previously solved SAM-I riboswitch (Montange and Batey 2006) was noted. If such similarities could be inferred from sequence or multiple sequence alignments (analogous to fold recognition methods in protein modeling), we expect that substantially more accurate models could be built. We are therefore hopeful about further development of RNA structural bioinformatics approaches such as Rmdetect (Cruz and Westhof 2011) and FR3D (Sarver et al. 2008).

## Dokholyan group

The Dokholyan group adopted a multi-scale, molecular dynamics approach (Ding and Dokholyan 2012). Briefly, coarse-grained discrete molecular dynamics (DMD) simulations are used to sample the vast conformational space of RNA molecules. The representative structures are selected from the coarse-grained simulations based on energies and/or additional filters such as the radius of gyration or other experimentally known parameters. RNA nucleotides are represented in coarse-grained simulations by three pseudo-atoms corresponding to the base, sugar, and phosphate groups (Ding et al. 2008b). The neighboring beads along the sequence are constrained to reflect RNA chain connectivity and local geometry, including covalent bonds, bond angles, and dihedral angles. The parameters for bonded interactions are derived from high-resolution RNA structures. Nonbonded interactions include base-pairing, base-stacking, short-range phosphate–phosphate repulsion, and hydrophobic interactions. These interaction parameters are derived from the sequence-dependent free energy parameters of the individual nearest-neighbor hydrogen bond model (INN-HB) (Mathews et al. 1999). Given an initial coarse-grained RNA model, the corresponding all-atom model is reconstructed and further optimized with all-atom DMD simulations (data not shown). The all-atom DMD RNA modeling approach is an extension of all-atom DMD protein modeling (Ding et al. 2008b). In DMD simulations, the structural information of a given RNA, such as base pairs and distances between specific nucleotides, can be incorporated as constraints to guide RNA folding (Gherghe et al. 2009; Lavender et al. 2010).

The CPU time for DMD simulations depends on RNA length. For the coarse-grained simulations, previous benchmarks suggested a near-linear dependence on RNA length (Ding et al. 2008a). For example, for an RNA of  $\sim 80$  nt (such as Puzzle 3), the total computational time for the coarse-grained DMD simulation is  $\sim 12$  h. The procedure to identify representative structures using the clustering algorithm usually requires  $<1$  h. The CPU time of the all-atom DMD simulation also depends on RNA length  $n$ , with the computational complexity of order  $\sim n \ln(n)$ . For the 84-nt RNA (Problem 3), the CPU time was  $\sim 18$  h; and the CPU time for 100-nt RNA (Problem 2) was  $\sim 24$  h.

In the current three RNA puzzles, base pairs derived either from previous knowledge (input from the experimentalists in Puzzle 2; RNA secondary structure prediction combined with biochemical validation in Puzzle 3) or from biochemical intuition (Puzzle 1) were included. Once the structural information is gathered and prepared for the refinement simulations, the computational effort is then fully automated.

The Dokholyan group's DMD approach has been designed for fold refinement of relatively large RNAs with complex 3D architectures. It was especially successful with Puzzle 3, where the models recapitulated the global fold well. In an independent structure-prediction exercise, the Dokholyan group also recently predicted a structure for the pseudoknot domain of the hepatitis C virus internal ribosome entry site (Lavender et al. 2010). The structure of a closely related RNA construct was subsequently determined by crystallography (Berry et al. 2011). Their prediction for the HCV pseudoknot domain shows good agreement with the global fold of the experimental structure (RMSD  $\approx 11$  Å and  $P$ -value  $5 \times 10^{-3}$ ), although some local interactions were missed.

For the simpler RNAs in Puzzles 1 and 2, the learned lesson is that inclusion of as much experimentally validated structural information as possible improves predictions, but it is important to avoid over-constraining the simulation. Instead, the DMD simulations should be allowed to sample the favorable conformational space, where constraints are unclear. For example, in their solution to Puzzle 1, the Dokholyan group overestimated the internal base-pairing in the middle of the monomer sequence based on misinterpreting the statement in the puzzle that "The strands hybridize with blunt ends (C–G closing base pairs)." As a result, their prediction for Puzzle 1 had the highest RMSD among the predictions. In a post priori simulation, in which only the G–C pairs at the ends were constrained to form base pairs, the predicted model structure had a much smaller RMSD (4.3 Å) and would have ranked among the top third of models.

### Flores group

For 3D structure prediction, the Flores group used RNABuilder (named MMB in a subsequent release), an internal coordinate-mechanics code that allows the user to specify the flexibility, forces, constraints, and full or partial structural coordinates to model the structure and/or dynamics of an RNA molecule (Flores and Altman 2010; Flores et al. 2011). Working in internal coordinates has the advantage that regions of known structure in a model can be rigidified, thus eliminating the cost associated with solving the equations of motion for internal rearrangements of that region (Flores and Altman 2011). Steric exclusion can be accounted for economically using collision-detecting spheres that are applied to a subset of atoms in user-specified residues. Any canonical or noncanonical base-pairing interaction catalogued in Leontis et al. (2002), plus stacking and a "Superimpose" threading force can be enforced between any and all pairs of residues specified by the user. These features have been used for RNA threading (Flores et al. 2010) and for generating an all-atoms trajectory of ribosomal hybridization using structural and biochemical information (Flores and Altman 2011).

The processing time on a single core of a 3.0 GHz Intel processor was  $\sim 94$  min. Notice that this run was not optimized for speed, and also that a newer version of RNABuilder (named MMB) is at least twice as fast due to improvements in the

underlying Simbody internal coordinate dynamics engine (Sherman et al. 2011).

RNABuilder is intended to be easy to use, and this goal is supported by the use of a single, free-format command file that is prepared using a relatively intuitive syntax comprising terms recognizable by any biologist. However, the package is also designed to provide the user control over the flexibility, forces, and parameters of the model, in order to be useful for a wide variety of applications; hence, it is not automated. The human time required for preparing a run is thus dependent on the experience of the user and the complexity of the task. RNABuilder is designed to enable fast runtimes; most tasks undertaken require minutes to hours, depending on the task. A trained user can also reduce the degrees of freedom and structure the problem to allow larger integration time steps for greater efficiency. Also, most users in practice will do multiple calculations before coming to a biological conclusion. In this group's experience, the human/computer time ratio is typically much greater than unity.

### Major group

This CASP-like contest was an opportunity for the Major group to test their fully automated MC-Fold and MC-Sym pipeline (Parisien and Major 2008). Two students in the laboratory, Véronique Lisi (PhD student in molecular biology) and Marc-Frédéric Blanchet (PhD student in computer science), who did not participate in the development of the pipeline were selected to participate. Lisi solved the Homodimer and Square problems, and Blanchet solved the Riboswitch Domain problem. Except for the Riboswitch Domain, no human intervention or numerical refinement was applied to the structures that were submitted to the contest, or in other words, the structures that were submitted were taken directly from the output of the MC-Sym program. This explains their high Clash Scores.

#### Homodimer

Lisi concatenated two copies of the given sequence into one that was submitted to MC-Fold using the default parameters, i.e., not considering pseudoknots; best 20 structures; and, explored the best 15% suboptimal structures. The structure predicted with the highest probability (i.e., minimum free energy) was then submitted to MC-Sym. The first 3D structure generated by MC-Sym was submitted to the contest (atomic clashes 1.5 Å all-atoms but hydrogens; backtrack probabilist, width limit 25%, height limit 33%; backbone method estimate, threshold 2.0 Å; maximum of 1000 models, CPU time limit 180 min, seed 3210, RMSD min 3.0 Å side chain only).

#### Square

Lisi directly used the 2D structure provided as the input to MC-Sym. The first structure generated by MC-Sym was submitted to the contest (using the same parameters as for the Homodimer).

#### Riboswitch domain

Maria Abella, a MSc student in bioinformatics, suspected that this sequence was a riboswitch. She checked its matching DNA sequence in GenBank and determined that it was from *Vibrio*

*cholerae*. Then, using BLAST, she found that the sequence was ~100 bp away from the sodium/glycine symporter GlyP gene in *Bacillus subtilis*, which was previously reported to be controlled by a riboswitch (Mandal et al. 2004). The sequence of the *B. subtilis* riboswitch is the same except for one nucleotide as that for Problem 3. Its 2D structure supported by chemical probing data was also published by Mandal et al. (2004).

Suspecting the riboswitch, Blanchet decided to submit two structures. He first predicted 2D structures using MC-Fold (not considering pseudoknots; best 1000 structures; explore the best 15% suboptimal structures). Generating the set of 1000 suboptimal structures took <2 min on an Intel i7 @ 2.67 GHz. He noted that the 2D structure published by Mandal et al. (2004) was absent in the set. He grouped the 1000 structures according to their topologies. Among the topologies, he kept a cloverleaf (four-way junction; the most frequent with 796 structures among the 1000) and a Y-shape (three-way junction), which corresponded to the same topology of the structure published by Mandal et al. (2004). However, the base-pairing pattern of the predicted structure differed much from that published. He then generated decoys of 3D structures using MC-Sym for the best-scoring 2D structure in each chosen topology (atomic clashes 1.5 Å all-atoms but hydrogens; backtrack probabilist, width limit 25%, height limit 33%; backbone method ccm, pucker = C3'-endo, threshold 2.0 Å; maximum of 9999 models, CPU time limit 12 h, seed 3210, RMSD min 1.0 Å side chain only). He edited the MC-Sym input scripts to explore independently and more of the conformational space of each stem in the same allowed time. The structures of the individual stems were merged in complete structures for an additional 12 h. For each topology (5685 Y-shape and 9999 cloverleaf structures), Blanchet selected the centroid model of the 20 best scoring models (according to the P-Score described at the MC-Sym command page; the centroid structure had a P-Score of -61.44). A "Relieve" minimization (see the MC-Sym command page) was applied to both selected models. This operation corrects the major steric conflicts in the backbone but does not refine the overall structure. This is reflected by the high Clash Score in the submitted models.

#### Lesson for the Major group

Obviously, in Problem 3, the minimum free energy structure predicted by MC-Fold differs from that of the crystal. Worst, it is not even predicted among the 1000 suboptimal structures. Just for P3, the 2D structure corresponding to the crystal is evaluated at -39.7 kcal/mol (minimum free energy structure for P3 = -47.9 kcal/mol) and ranks near 50th only. It is not possible to see at this time how this structure could be selected by the program unless more information than the sequence is provided. It would be interesting to see the precision of 3D structures that would have been generated by MC-Sym, given that the crystal 2D structure could have been selected. Thus, a decoy of 3D structures using an input script was generated from the correct 2D dot-bracket (from Fig. 3A without the G29:C83:A11 triple), and after applying the "Relieve" minimization, the structures ranged between 7 and 21 Å of RMSD with the crystal structure (data not shown) (best RMSD = 6.8 Å; P-Score = -25.38). However, selecting this best RMSD structure using our P-Scores is not possible, because the best P-Score structure, -53.84, has an RMSD of 10.3 Å with the crystal structure.

## SantaLucia group

The SantaLucia group submitted models for Problems 1 and 2. Both models were generated using the de novo modeling module within the RNA123 software suite (Sijen et al. 2012). Below is a brief description of the methodology.

### Problem 1

The first step was to predict the secondary structure of the submitted sequence. To accomplish this, the sequence 5'-CCGCC GCGCAUGCCUGUGGCGGUUCGCCGCCGCCAUGCCUG UGGCGG-3' was submitted to a secondary structure-folding algorithm. The UUCG hairpin was added in order to make a continuous chain of RNA, because RNA123 folds a single chain of RNA. The UUCG hairpin was later manually removed once the tertiary structure was predicted. The secondary structure was predicted using a thermodynamics-based dynamic programming algorithm within RNA123 that produced 10 optimal and suboptimal secondary structure folds. A tertiary structure model was computed only for the optimal secondary structures by decomposing the secondary structure into constituent motifs such as internal loops, helices, and hairpins. The 3D structure was then assembled using a motif library. The motif library was generated from RNA structures previously deposited in the PDB (Protein Data Bank). The selection and assembly of the motifs are automated within the RNA123 via an algorithm called BUILDER (Sijen et al. 2012), which uses an energy function to score and assemble the 3D model. The manual effort in performing the prediction for Problem 1 was minor, notably in removing the UUCG hairpin after the models were generated.

### Comments

After the results for the prediction of Problem 1 were released, the group discovered that they had submitted a model generated from the wrong sequence. Specifically, residue 15 was a C instead of a U, and residue 18 was a U instead of a C. This meant that the dimer ended up with four incorrect residues and thereby compromised the quality of their prediction. It is important to note that this error was later fixed and produced a model that scores better against the crystal structure.

### Problem 2

This problem was solved by a combined effort of both manual and automated steps. The given secondary structure was decomposed into four identical "L"-shaped secondary structures with dangling ends on the 5' end (CCGG) and 3' end (GGCC). The idea was to generate four identical tertiary structures and then base-pair the dangling ends so that a perfect square (Fig. 2A) can be assembled from the four "L"-shaped structures. Using the de novo modeling module in RNA123, a 3D model with the lowest energy was computed. Fortunately, this model happened to have an "L"-shaped tertiary structure. Four copies of this structure were then created and superimposed onto the provided coordinates of the inner strand from Problem 3, forming an initial coarse model consistent with the "square" secondary structure. This coarse model had distorted base pairs between the 5' and 3' dangling ends of each of the preceding fragments, and therefore we subjected

the entire model to the energy optimization algorithm in RNA123. This algorithm, named DSTA (Discrete Sampling of Torsion Angles), uses a multi-dimensional search and a novel method for modeling the local potential energy surface and finding an analytical minimum (Sijen et al. 2012).

#### *An estimate of the time required to produce the models*

The computer calculation time (on a laptop with Intel Core 2 Duo CPU P8600 @ 2.4 GHz processor) was as follows:

Problem 1. It took ~25 min to predict a single tertiary structure using the de novo prediction platform in RNA123.

Problem 2. It took ~20 min to predict and generate four “L”-shaped tertiary structures using the de novo prediction platform in RNA123. Manual assembly and running of the DSTA optimization algorithm took ~1 h.

#### *An estimate of the proportion human effort/machine effort*

Problem 1. Twenty-five minutes of machine effort, negligible human effort.

Problem 2. Twenty minutes of machine effort, 1 h of human effort.

### PDB file normalization

Both files for the accepted experimental structures and predicted model files, submitted in PDB format, were normalized to comply with a common standard. Only the first model present in the file was considered. All records except for the ATOM and TER records were ignored. Only the four nucleotides A, C, U, and G were considered. Modified nucleotides were treated as unmodified bases, and extra atoms were discarded (e.g., a 5-bromouracil is treated as a normal uracil and the extra bromine atom is discarded). The only atoms kept are those for the bases (C2, C4, C6, C8, N1, N2, N3, N4, N6, N7, N9, O2, O4, and O6) and for the sugar-phosphate backbone (C1', C2', C3', C4', C5', O2', O3', O4', O5', OP1, OP2, and P).

### Stereochemical evaluation

The stereochemical evaluation was performed using the MolProbity (Davis et al. 2007) tool. In a first step, hydrogen atoms were added to the model using the “reduce-build” command line utility, and the Clash Score value was computed using the “online-analysis -nocbeta -norota -norama” command.

### RMSD computation

The RMSD is computed using the “Superimposer” class from the “Bio.PDB” package (Hamelryck 2003). The “Superimposer” class translates and rotates the comparing model to minimize its RMSD in respect to the reference model. It uses a singular value decomposition algorithm as described in Golub and Van Loan (1989).

### Deformation Index and Deformation Profile computations

The base–base interactions (BBI) of both solution and predicted models are extracted using the MC-Annotate (Gendron et al. 2001) tool. The Interaction Network Fidelity (INF) value is computed as:

$$INF = \sqrt{\left(\frac{TP}{TP+FP}\right) \times \left(\frac{TP}{TP+FN}\right)},$$

where TP is the number of correctly predicted BBI, FP is the number of predicted BBI with no correspondence in the solution model, and FN is the number of BBI in the solution model not present in the predicted model. The Deformation Index is then computed as:

$$DI = \frac{RMSD}{INF}.$$

Several partial INF (and respective DI) can be computed if one considers only the Watson-Crick (WC) base pairs ( $INF_{WC}$ ), the non-Watson-Crick (NWC) base pairs ( $INF_{NWC}$ ), both WC and NWC base pairs ( $INF_{BPS}$ ), or the stacking interactions ( $INF_{STACK}$ ).

The Deformation Profile is a distance matrix computed as the average RMSD between the individual bases of the predicted and the reference models while superimposing each nucleotide of the predicted model over the corresponding nucleotide of the reference model one at a time. It is computed using the “dp.py” command from the “SIMINDEX” package (Parisien et al. 2009).

### P-value computation

The P-value is computed as described (Hajdin et al. 2010) using:

$$P - value = \frac{1 + \operatorname{erf}\left(\frac{(RMSD - \langle RMSD \rangle) / 1.8}{\sqrt{2}}\right)}{2}, \text{ with}$$

$$\langle RMSD \rangle = a \times N^{0.41} - b.$$

the constants  $a$  and  $b$  depend on whether the secondary structure base-pairing information is provided ( $a = 5.1$  and  $b = 15.8$ ) or not ( $a = 6.4$  and  $b = 12.7$ ). This metric is only valid for RNAs with true higher-order 3D folds and thus only applies to Problem 3 (with base-pairing as an assumed constraint).

### Graphics

Interactive molecular module images in the *RNA-Puzzles* website are produced with Jmol (<http://www.jmol.org>) and the secondary structures with VARNA (Darty et al. 2009).

### ACKNOWLEDGMENTS

J.A.C. is supported by the PhD Program in Computational Biology of the Instituto Gulbenkian de Ciência, Portugal (sponsored by Fundação Calouste Gulbenkian, Siemens SA, and Fundação para a Ciência e Tecnologia; SFRH/BD/33528/2008). T.H. is supported by the National Institutes of Health, grants AI72012 and CA132753. The work of the Bujnicki group was supported by the

Polish Ministry of Science (HISZPANIA/152/2006 grant to J.M.B. and PBZ/MNiSW/07/2006 grant to M.B.), by the EU (6FP grant “EURASNET” LSHG-CT-2005-518238 and structural funds POIG.02.03.00-00-003/09), by the Faculty of Biology, Adam Mickiewicz University (PBWB-03/2009 grant to M.R.), and by the German Academic Exchange Service (grant D/09/42768 to K.R.). The work of the Chen group was supported by NIH grant GM063732 and NSF grants MCB0920067 and MCB0920411 to S.-J.C. The Weeks and Dokholyan groups were supported by the U.S. National Institutes of Health (grant GM064803). M.F.B. holds a scholarship from the Fonds Québécois de Recherche: Nature et Technologies. V.L. holds a scholarship from the Canadian Institutes of Health Research (CIHR). The Major group is supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant), the CIHR (grant MOP93679), and the U.S. National Institutes of Health (grant GM088813). The RNA123 software used to predict the structures and all related activities have been funded by the National Institutes of Health Grants R01-GM073179 (P.I. John SantaLucia), U01-AI061192 (P.I. Philip Cunningham), Grants R44 GM85889 (P.I. Norman E. Watkins, Jr., and Fredrick Sijenyi), and R44 GM095251 (P.I. John SantaLucia, Jr. and Fredrick Sijenyi).

Received October 22, 2011; accepted December 20, 2011.

## REFERENCES

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242.
- Berry KE, Waghay S, Mortimer SA, Bai Y, Doudna JA. 2011. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. *Structure* **19**: 1456–1466.
- Boniecki M, Rotkiewicz P, Skolnick J, Kolinski A. 2003. Protein fragment reconstruction using various modeling techniques. *J Comput Aided Mol Des* **17**: 725–738.
- Cao S, Chen S-J. 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* **11**: 1884–1897.
- Cao S, Chen S-J. 2006a. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* **34**: 2634–2652.
- Cao S, Chen S-J. 2006b. Free energy landscapes of RNA/RNA complexes: With applications to snRNA complexes in spliceosomes. *J Mol Biol* **357**: 292–312.
- Cao S, Chen S-J. 2009. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA* **15**: 696–706.
- Cao S, Chen S-J. 2011. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* **115**: 4216–4226.
- Chen S-J. 2008. RNA folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* **37**: 197–214.
- Cruz JA, Westhof E. 2011. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat Methods* **8**: 513–519.
- Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**: 1974–1975.
- Das R, Baker D. 2008. Macromolecular modeling with rosetta. *Annu Rev Biochem* **77**: 363–382.
- Das R, Karanicolas J, Baker D. 2010. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* **7**: 291–294.
- Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB III, Snoeyink J, Richardson JS, et al. 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**: W375–W383. doi: 10.1093/nar/gkm216.
- Dibrov S, McLean J, Hermann T. 2011a. Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA. *Acta Crystallogr D Biol Crystallogr* **67**: 97–104.
- Dibrov SM, McLean J, Parsons J, Hermann T. 2011b. Self-assembling RNA square. *Proc Natl Acad Sci* **108**: 6405–6408.
- Ding F, Dokholyan N. 2012. Multiscale modeling of RNA structure and dynamics. In *RNA 3D structure analysis and prediction* (ed. NB Leontis, E Westhof). Springer-Verlag, Berlin.
- Ding F, Tsao D, Nie H, Dokholyan N. 2008a. Ab initio folding of proteins using all-atom discrete molecular dynamics. *Structure* **16**: 1010–1018.
- Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. 2008b. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* **14**: 1164–1173.
- Ditzler MA, Otyepka M, Sponer J, Walter NG. 2010. Molecular dynamics and quantum mechanics of RNA: Conformational and chemical change we can believe in. *Acc Chem Res* **43**: 40–47.
- Flores SC, Altman RB. 2010. Turning limited experimental information into 3D models of RNA. *RNA* **16**: 1769–1778.
- Flores SC, Altman R. 2011. Structural insights into pre-translocation ribosome motions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 205–11. <http://www.ncbi.nlm.nih.gov/pubmed/21121048>.
- Flores SC, Wan Y, Russell R, Altman RB. 2010. Predicting RNA structure by multiple template homology modeling. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 216–27. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2872935&tool=pmcentrez&rendertype=abstract>.
- Flores SC, Sherman MA, Bruns CM, Eastman P, Altman RB. 2011. Fast flexible modeling of RNA structure using internal coordinates. *IEEE/ACM Trans Comput Biol Bioinformatics* **8**: 1247–1257.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, et al. 2009. Rfam: Updates to the RNA families database. *Nucleic Acids Res* **37**: D136–D140. doi: 10.1093/nar/gkn766.
- Gendron P, Lemieux S, Major F. 2001. Quantitative analysis of nucleic acid three-dimensional structures. *J Mol Biol* **308**: 919–936.
- Gherghe CM, Leonard CW, Ding F, Dokholyan NV, Weeks KM. 2009. Native-like RNA tertiary structures using a sequence-encoded cleavage agent and refinement by discrete molecular dynamics. *J Am Chem Soc* **131**: 2541–2546.
- Golub G, Van Loan C. 1989. *Matrix computations*. Johns Hopkins University Press, Baltimore.
- Hajdin CE, Ding F, Dokholyan NV, Weeks KM. 2010. On the significance of an RNA tertiary structure prediction. *RNA* **16**: 1340–1349.
- Hamelryck T. 2003. PDB file parser and structure class implemented in Python. *Bioinformatics* **19**: 2308–2310.
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* **125**: 167–188.
- Huang L, Serganov A, Patel DJ. 2010. Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Mol Cell* **40**: 774–786.
- Jonikas MA, Radmer RJ, Altman RB. 2009. Knowledge-based instantiation of full atomic detail into coarse-grain RNA 3D structural models. *Bioinformatics* **25**: 3259–3266.
- Jossinet F, Ludwig TE, Westhof E. 2010. Assemble: An interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**: 2057–2059.
- Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. 2003. A “Frankenstein’s monster” approach to comparative modeling: Merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. *Proteins* **53**: 369–379.

- Lavender CA, Ding F, Dokholyan NV, Weeks KM. 2010. Robust and generic RNA modeling using inferred constraints: A structure for the hepatitis C virus IRES pseudoknot domain. *Biochemistry* **49**: 4931–4933.
- Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30**: 3497–3531.
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**: 275–279.
- Martinez HM, Maizel JV, Shapiro B. 2008. RNA2D3D: A program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* **25**: 669–683.
- Marti-Renom MA, Madhusudhan MS, Fiser A, Rost B, Sali A. 2002. Reliability of assessment of protein structure prediction methods. *Structure* **10**: 435–440.
- Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940.
- Matthews B. 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* **405**: 442–451.
- Metropolis N, Ulam S. 1949. The Monte Carlo method. *J Am Stat Assoc* **44**: 335–341.
- Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216**: 585–610.
- Montange RK, Batey RT. 2006. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**: 1172–1175.
- Moult J. 2006. Rigorous performance evaluation in protein structure modelling and implications for computational biology. *Philos Trans R Soc Lond B Biol Sci* **361**: 453–458.
- Murray LJW, Arendall WB, Richardson DC, Richardson JS. 2003. RNA backbone is rotameric. *Proc Natl Acad Sci* **100**: 13904–13909.
- Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55.
- Parisien M, Cruz JA, Westhof E, Major F. 2009. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* **15**: 1875–1885.
- Qian B, Raman S, Das R, Bradley P, McCoy AJ, Read RJ, Baker D. 2007. High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**: 259–264.
- Reuter JS, Mathews DH. 2010. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**: 129. doi: 10.1186/1471-2105-11-129.
- Rother K, Rother M, Boniecki M, Puton T, Bujnicki JM. 2011a. RNA and protein 3D structure modeling: Similarities and differences. *J Mol Model* **17**: 2325–2336.
- Rother M, Rother K, Puton T, Bujnicki JM. 2011b. ModeRNA: A tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* **39**: 4007–4022.
- Rother K, Rother M, Boniecki M, Puton T, Tomala K, Lukasz P, Bujnicki JM. 2012. Template-based and template-free modeling of RNA 3D structure: Inspirations from protein structure modeling. In *RNA 3D structure analysis and prediction* (ed. NB Leontis, E Westhof). Springer-Verlag, Berlin.
- Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. 2008. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* **56**: 215–252.
- Serra MJ, Axenson TJ, Turner DH. 1994. A model for the stabilities of RNA hairpins based on a study of the sequence dependence of stability for hairpins of six nucleotides. *Biochemistry* **33**: 14289–14296.
- Sherman Ma, Seth A, Delp SL. 2011. Simbody: Multibody dynamics for biomedical research. *Procedia IUTAM* **2**: 241–261.
- Sijenyi F, Saro P, Ouyang Z, Damm-Ganamet K, Wood M, Jiang J, SantaLucia J Jr. 2012. The RNA folding problems: Different levels of RNA structure prediction. In *RNA 3D structure analysis and prediction* (ed. NB Leontis, E Westhof). Springer-Verlag, Berlin.
- Sripakdeevong P, Kladwang W, Das R. 2011. An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc Natl Acad Sci* **108**: 20573–20578.
- Surles MC, Richardson JS, Richardson DC, Brooks FP. 1994. Sculpting proteins interactively: Continual energy minimization embedded in a graphical modeling system. *Protein Sci* **3**: 198–210.
- Tinoco I Jr, Bustamante C. 1999. How RNA folds. *J Mol Biol* **293**: 271–281.
- Tinoco I Jr, Borer PN, Dengler B, Levin MD, Uhlenbeck OC, Crothers DM, Bralla J. 1973. Improved estimation of secondary structure in ribonucleic acids. *Nat New Biol* **246**: 40–41.
- Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, et al. 1980. Secondary structure model for bacterial 16S ribosomal RNA: Phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* **8**: 2275–2293.
- Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. 1999. Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* **285**: 1711–1733.
- Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. 2003. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* **31**: 3450–3460.
- Zemla A. 2003. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* **31**: 3370–3374.
- Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**: 702–710.
- Zhao Q, Han Q, Kissinger CR, Hermann T, Thompson P. 2008. Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* **64**: 436–443.
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

## **II. Chemical probing of RNA in living cells**

M. Wildauer, G. Zemora, A. Liebeg, V. Heisig, C. Waldsich,

Methods Mol Biol, 1086: pp 159-76, 2014.

## Chemical Probing of RNA in Living Cells

Michael Wildauer, Georgeta Zemora, Andreas Liebeg,  
Verena Heisig, and Christina Waldsich

### Abstract

RNAs need to adopt a specific architecture to exert their task in cells. While significant progress has been made in describing RNA folding landscapes *in vitro*, understanding intracellular RNA structure formation is still in its infancy. This is in part due to the complex nature of the cellular environment but also to the limited availability of suitable methodologies. To assess the intracellular structure of large RNAs, we recently applied a chemical probing technique and a metal-induced cleavage assay *in vivo*. These methods are based on the fact that small molecules, like dimethyl sulfate (DMS), or metal ions, such as  $Pb^{2+}$ , penetrate and spread throughout the cell very fast. Hence, these chemicals are able to modify accessible RNA residues or to induce cleavage of the RNA strand in the vicinity of a metal ion in living cells. Mapping of these incidents allows inferring information on the intracellular conformation, metal ion binding sites or ligand-induced structural changes of the respective RNA molecule. Importantly, *in vivo* chemical probing can be easily adapted to study RNAs in different cell types.

**Key words** *In vivo* chemical probing, DMS, RNA structure, RNA folding, Metal-induced cleavage, Metal ions, Lead cleavage

---

## 1 Introduction

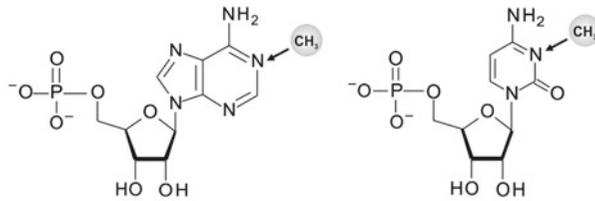
Exploring RNA folding paradigms *in vitro* has been a major objective in RNA biology [1–10]. RNA in general folds in a hierarchical manner: first monovalent ions are required to shield the negatively charged RNA backbone, allowing counterion-mediated condensation and assembly of secondary structure. By binding to distinct metal ion pockets within RNA,  $Mg^{2+}$  ions allow further compaction of the molecule and the formation of long-range, tertiary interactions. Although *in vitro* studies are inevitable for characterizing RNA structure and folding pathways, the *in vitro* refolding conditions are very different from the intracellular environment [11, 12]. Especially temperature and ion concentrations, two very important factors in RNA folding, can vary considerably.

Furthermore, the directionality and velocity of transcription and translation as well as the presence of *trans*-acting factors, such as proteins, can influence RNA folding in vivo [11, 12]. However, there are only few techniques available for determining the structure of RNAs in the cell. As a result little is known about RNA folding in vivo and the contribution of the cellular environment to RNA structure formation.

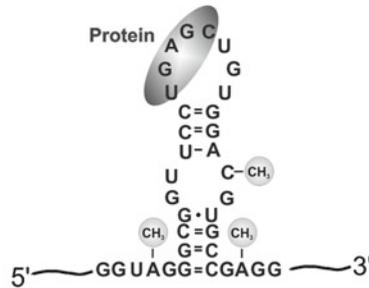
Chemical probing is a powerful method, which has been used most extensively to map RNA structure in vivo [13–25]. It is a simple, well-adaptable, and inexpensive approach. Dimethyl sulfate (DMS), the most widely used chemical probing reagent, rapidly penetrates the cell and all of its compartments, modifying N1 of adenines and N3 of cytosines. DMS only methylates proton accepting ring nitrogens, if these atoms are not engaged in hydrogen bonding (e.g., Watson–Crick base pairing or sheared AA base pairs) and if they are solvent accessible, while a reduced solvent exposure or binding of a protein results in protection from DMS modification (Fig. 1). Occasionally, certain uridines and guanines are reactive to DMS at their N3 or N1 position, respectively, if these are stabilized in an enol-tautomer due to a specific local environment [26, 27]. The sites of modification are mapped by reverse transcription, as the bulky methyl group at the Watson–Crick face of adenines and cytosines leads to termination of the primer extension. Thus, reverse transcription of the modified RNA pool results in a variety of cDNAs of different length, which can be resolved on a standard denaturing polyacrylamide gel. Information on the RNA structure is obtained by analyzing the DMS modification pattern in the context of secondary structure maps derived from phylogenetic or bioinformatic studies [13–25]. The experimental framework for the  $\text{Pb}^{2+}$ -induced cleavage assay is not fundamentally different to the DMS approach with one notable exception: in contrast to DMS, which is a nucleobase-specific probe,  $\text{Pb}^{2+}$  is able to displace  $\text{Mg}^{2+}$  ions in their binding pockets and induces strand scission of the RNA backbone in its vicinity [28–31]. These cleavage events can be mapped by reverse transcription as well.

Employing in vivo DMS chemical probing we recently provided the first structure-based insights into DEAD-box protein-facilitated RNA folding in living cells [18]. In yeast mitochondria efficient splicing of all group I and group II introns is dependent on the DEAD-box helicase Mss116p [32]. To monitor Mss116p-induced conformational changes within the ai5 $\gamma$  group II intron in yeast, we mapped the ai5 $\gamma$  structure in different genetic backgrounds (Fig. 2). While the intron adopts the native conformation in the wild-type yeast strain, ai5 $\gamma$  appears to be largely unfolded in the *mss116*-knockout strain, as most of the secondary structure elements, but none of the tertiary interactions are

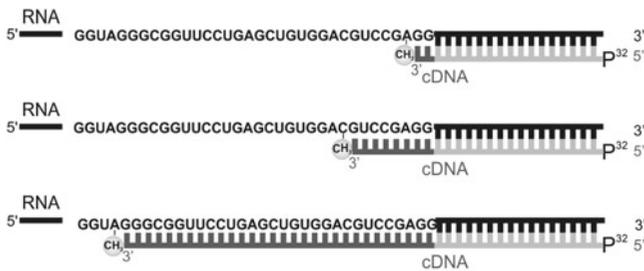
### a Methylation by Dimethyl sulfate



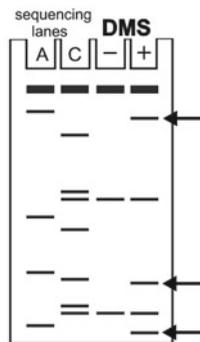
### b Modification of accessible bases



### c Reverse Transcription



### d PAGE analysis of cDNA products



**Fig. 1** Schematic representation of *in vivo* chemical probing using DMS. **(a)** The N1 atom of adenines and the N3 atom of cytosines are methylated by DMS, **(b)** if these atoms are not involved in H-bonding or protected by a protein. **(c)** The modified residues can be detected by reverse transcription using 5'-end-labeled, gene-specific DNA primers (*light gray*), as the extension by the reverse transcriptase terminates when the enzyme encounters the bulky methyl group. This reaction generates a family of radio-labeled cDNAs of different length (*dark gray*). **(d)** The pool of cDNA is separated by a standard denaturing PAGE. A and C denote the sequencing **Fig. 1** (continued) lanes which are used to determine the position of the modification. The -lane is the RT stop control, showing natural stops occurring during reverse transcription, for which RNA extracted from cells that were not treated with the DMS was used. In the +lane RNA has been reverse transcribed that was extracted from cells treated with DMS. Comparing these two lanes reveals the sites of DMS modifications (indicated by an *arrow*)



formed [18]. In brief, most of the Mss116p-induced structural changes are observed within domain D1 (Fig. 2); thus Mss116p appears to facilitate the formation of this largest domain, which is the scaffold for docking of other intron domains. Based on the chemical probing data we proposed that Mss116p assists the ordered assembly of the ai5 $\gamma$  intron in vivo and is critical for folding of the RNA at an early step along the pathway [18]. In light of the fact that metal homeostasis plays a crucial role in yeast mitochondrial intron splicing [33] and that Mss116p lowers the Mg<sup>2+</sup> requirement for intron folding in vitro [34, 35], we were also interested in assessing the impact of Mss116p on the formation of metal ion binding pockets within the ai5 $\gamma$  intron in vivo. Using the Pb<sup>2+</sup>-induced cleavage assay we determined the metal ion binding sites within the ai5 $\gamma$  intron in vivo (Fig. 3) and observed so far that these correlate nicely with those identified previously in vitro [36], suggesting that Mss116p may not significantly influence the formation of metal ion binding sites in vivo (*Wildauer and Waldsich, unpublished*). Here we provide a detailed description of both the in vivo DMS chemical probing technique and the lead-induced cleavage assay for mapping RNA structure in eukaryotic cells.

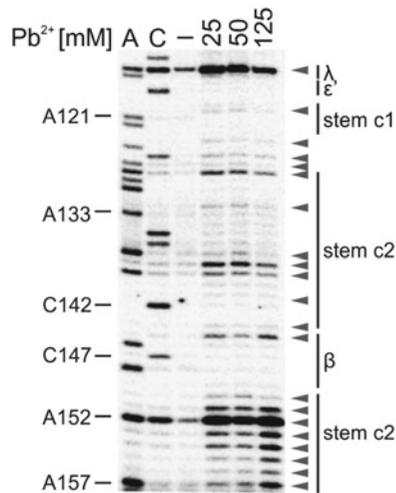
---

## 2 Materials

Prepare all solutions with RNase-free ultrapure water (deionized water with a sensitivity of  $\geq 18$  M $\Omega$  cm at 25 °C) unless indicated otherwise. Sterilize all solutions by autoclaving or filtering (membrane pore size 0.2  $\mu$ m) and store them at room temperature if not indicated otherwise. Use RNase-free glassware and plasticware for the experiments.

---

**Fig. 2** (continued) represent an increase in accessibility, while open ones highlight bases with reduced accessibility). Strong changes in accessibility are displayed in dark gray (>2-fold), while smaller changes are shown in *light gray* (1.5 to 2-fold). These values were derived from normalized gel plots. In the –lane natural stops of the reverse transcriptase are seen. In the +lane the in vivo DMS pattern is shown. Notably, comparing these two lanes reveals the DMS-induced stops of the reverse transcriptase and thus accessible residues (N1-A, N3-C). Importantly, the – and +lanes are shown for both the wt and *mss116*-knockout strain. Comparing *lanes 4* and *6* reveals the altered DMS modification pattern and thus conformational changes within the ai5 $\gamma$  intron due to the absence of Mss116p. **(b)** Differential summary map: residues, the modification intensity of which changes in the absence of Mss116p ( $\Delta$ *mss116* strain), are highlighted. This map is based on normalized plots, which had been derived from the modification gels. The *closed squares* indicate an increase in accessibility, while *open squares* represent bases with reduced accessibility in the absence of Mss116p. The A and C residues whose modification remains unaltered (i.e., equally modified or protected in both strains) are not highlighted. This figure has been adapted from [18], with permission



**Fig. 3** Mapping metal ion binding sites within the ai5 $\gamma$  intron in yeast mitochondria. Pb<sup>2+</sup>-induced cleavage sites (indicated with a *gray arrow*) were mapped via primer extension. A and C denote sequencing lanes; –lane is an RT stop control to detect natural stops of the extension (RNA, which was extracted from untreated cells, was reverse transcribed). Lanes labeled 25–125 indicate the concentration of Pb(OAc)<sub>2</sub> [mM] used to modify the intracellular RNA

### 2.1 Culturing Human Cells

1. Growth medium DMEM–FBS–PS for HEK 293 cells: Dulbecco's modified Eagle's medium (DMEM) with high glucose and supplemented with 10 % (v/v) fetal bovine serum (FBS) and 1 % (v/v) penicillin/streptomycin. Store at 4 °C.
2. Washing buffer for adherent HEK 293 cells: 1× Dulbecco's phosphate-buffered saline (1× DPBS) without Ca and Mg.
3. Detachment reagent for adherent HEK 293 cells: 1× Trypsin–EDTA. Store at 4 °C.

### 2.2 Growing Yeast Cells

1. YP growth medium: dissolve 1 % (w/v) yeast extract, 2 % (w/v) peptone from meat, pancreatic digest, in ddH<sub>2</sub>O and autoclave. Add carbon source directly before use (*see Note 1*).
2. 10 % (w/v) glucose.
3. 10 % (w/v) raffinose (*see Note 2*).
4. TM buffer: 10 mM Tris–HCl, pH 7.5, 1 mM MgCl<sub>2</sub>.

### 2.3 DMS Chemical Probing

1. 10.7 M dimethyl sulfate (DMS; *see Note 3*). Store at 4 °C.
2. 14.3 M  $\beta$ -mercaptoethanol. Store at 4 °C.
3. Isoamyl alcohol.

### 2.4 Pb<sup>2+</sup>-Induced Cleavage Assay

1. 0.5 M EDTA, pH 8.0.
2. 50 mM lead(II) acetate.

**2.5 RNA Isolation from Mammalian Cells**

1. Isol-RNA lysis reagent or comparable product. Store at 4 °C.
2. Chloroform.
3. Isopropanol.
4. 75 % (v/v) ethanol.

**2.6 RNA Isolation from Yeast**

1. AE buffer: 50 mM NaOAc, pH 4.5, 10 mM EDTA, pH 8.0.
2. 10 % (w/v) SDS.
3. Water-saturated phenol, pH 4.5. Store at 4 °C.
4. PCI (phenol:chloroform:isoamyl alcohol, 25:24:1). Store at 4 °C.
5. CI (chloroform:isoamyl alcohol, 24:1).
6. Ethanol/0.3 M NaOAc, pH 5.0.
7. RNase-free DNase I (2 U/μL) and supplied 10× DNase I reaction buffer. Store at -20 °C.
8. RNase inhibitor (40 U/μL). Store at -20 °C.
9. 0.5 M EDTA, pH 8.0.

**2.7 5' End-Labeling of DNA Primers**

1. 10 μM stocks of gene-specific DNA primers. Store at -20 °C.
2. [ $\gamma$ -<sup>32</sup>P]-ATP (10 μCi/μL, 6,000 Ci/mmol). Store at 4 °C.
3. T4 polynucleotide kinase (PNK, 10 U/μL). Store at -20 °C.
4. 10× PNK buffer (supplied with enzyme). Store at -20 °C.
5. 0.5 M EDTA, pH 8.0.
6. Glycogen (10 mg/mL). Store at -20 °C.
7. Ethanol/0.3 M NaOAc, pH 5.0.

**2.8 Primer Extension****2.8.1 Reverse Transcription with Transcriptor Reverse Transcriptase**

1. 4.5× hybridization buffer: 225 mM K-HEPES, pH 7.0, 450 mM KCl. Store at -20 °C.
2. Transcriptor reverse transcriptase (20 U/μL; Roche), or comparable product. Store at -20 °C.
3. 5× reaction buffer supplied with Transcriptor reverse transcriptase (Roche). Store at -20 °C.
4. 10 mM dNTP mix. Store at -20 °C.
5. 10 mM ddTTP. Store at -20 °C.
6. 10 mM ddGTP. Store at -20 °C.
7. 0.1 M DTT (dithiothreitol). Store at -20 °C.
8. RNase inhibitor (40 U/μL). Store at -20 °C.

**2.8.2 Reverse Transcription with AMV Reverse Transcriptase**

1. 4.5× hybridization buffer: 225 mM K-HEPES, pH 7.0, 450 mM KCl. Store at -20 °C.
2. 10× extension buffer: 1.3 M Tris-HCl, pH 8.0, 0.1 M MgCl<sub>2</sub>, 0.1 M DTT. Store at -20 °C.
3. 2.5 mM dNTP mix. Store at -20 °C.

4. 1 mM ddTTP. Store at  $-20^{\circ}\text{C}$ .
5. 1 mM ddGTP. Store at  $-20^{\circ}\text{C}$ .
6. AMV reverse transcriptase (10 U/ $\mu\text{L}$ ). Store at  $-20^{\circ}\text{C}$ .

**2.8.3 Solutions for Both  
Reverse Transcription  
Protocols**

1. 1 M NaOH.
2. 1 M HCl.
3. 0.5 M EDTA, pH 8.0.
4. Glycogen (10 mg/mL). Store at  $-20^{\circ}\text{C}$ .
5. Ethanol/0.3 M NaOAc, pH 5.0.
6. 10 $\times$  TBE buffer: 0.89 M Tris-base, 0.89 M boric acid, 20 mM EDTA.
7. Loading buffer: 7 M urea, 25 % (w/v) sucrose, 0.025 % (w/v) bromophenol blue, 0.025 % (w/v) xylene cyanol in 1 $\times$  TBE.

**2.9 Denaturing  
Polyacrylamide Gel  
Electrophoresis (PAGE)**

1. Vertical gel electrophoresis apparatus (adjustable).
2. Glass plate sets: 42 cm (L) $\times$ 20 cm (W).
3. Metal plate: 26.5 cm (L) $\times$ 20 cm (W).
4. Combs and spacers, 0.4 mm thickness.
5. High-voltage power supply.
6. Whatman 3MM paper.
7. Saran wrap.
8. 10 $\times$  TBE buffer: 0.89 M Tris-base, 0.89 M boric acid, 20 mM EDTA.
9. *N,N,N',N'*-tetramethylethylenediamine (TEMED). Store at  $4^{\circ}\text{C}$ .
10. 10 % (w/v) ammonium persulfate (APS) (*see Note 4*). Store at  $4^{\circ}\text{C}$ .
11. 8 % denaturing acrylamide solution: 7 M urea, 8 % (v/v) acrylamide/bisacrylamide (19:1) in 1 $\times$  TBE. Store at  $4^{\circ}\text{C}$ .
12. Phosphorimager exposure cassette and screen (GE Healthcare) or comparable product.
13. Phosphorimager (e.g., STORM 820, GE Healthcare) or comparable equipment.

---

## 3 Methods

### 3.1 Modification of RNA *In Vivo*

#### 3.1.1 DMS Modification of RNA in HEK Cells

1. Grow HEK cells to confluency in a 100 mm  $\varnothing$  dish (approx.  $8.8 \times 10^6$  cells) in DMEM-FBS-PS medium at  $37^{\circ}\text{C}$  with 5 %  $\text{CO}_2$ .
2. Discard the medium and wash the cells with 5 mL 1 $\times$  DPBS.
3. Add 1 mL 1 $\times$  Trypsin-EDTA and incubate at  $37^{\circ}\text{C}$  with 5 %  $\text{CO}_2$  for 2–3 min.

4. Add 9 mL DMEM–FBS–PS medium and transfer the cell suspension to 15 mL centrifugation tubes.
5. Centrifuge the cells at  $1,200 \times g$  for 5 min ( $4\text{ }^{\circ}\text{C}$ ).
6. Remove the supernatant and add 1 mL pre-warmed ( $37\text{ }^{\circ}\text{C}$ ) DMEM–FBS–PS medium.
7. Add DMS to a final concentration of 50 mM (vortex briefly). Incubate the cells at  $37\text{ }^{\circ}\text{C}$  for 2 min (*see* **Notes 3** and **5**).
8. Stop the DMS reaction by adding 50  $\mu\text{L}$  14.3 M  $\beta$ -mercaptoethanol and 50  $\mu\text{L}$  isoamyl alcohol. Vortex strongly and centrifuge at  $6,000 \times g$  for 2 min (*see* **Notes 6** and **7**).
9. Remove the supernatant carefully and resuspend the cells in 1 mL cold  $1 \times$  DPBS and add another 50  $\mu\text{L}$   $\beta$ -mercaptoethanol.
10. Vortex briefly and centrifuge at  $1,200 \times g$  for 5 min.
11. Discard the supernatant and proceed with RNA isolation (Subheading **3.2.1**).

### 3.1.2 DMS Modification of RNA in Yeast

1. Grow an overnight culture of the desired yeast strain at  $30\text{ }^{\circ}\text{C}$  in growth medium (e.g., YPD).
2. Inoculate 100 mL growth medium (e.g., YP + 2 % (w/v) raffinose, *see* **Notes 1** and **2**) with 5 mL of the overnight culture. Grow the culture at  $30\text{ }^{\circ}\text{C}$  to an  $\text{OD}_{600}$  of 1.0.
3. Harvest the cells ( $2 \times 30\text{ mL}$ ) by centrifugation at  $4,500 \times g$  for 5 min.
4. Discard the supernatant and resuspend the cell pellet in 1 mL pre-warmed ( $30\text{ }^{\circ}\text{C}$ ) growth medium (e.g., YPD).
5. Centrifuge the samples at  $6,000 \times g$  for 2 min. Carefully discard the supernatant.
6. Add DMS to a final concentration of 50 mM to one of the samples and vortex (*see* **Notes 3** and **5**).
7. Immediately incubate both samples at  $30\text{ }^{\circ}\text{C}$  for 2 min.
8. Stop the DMS reaction by adding  $\beta$ -mercaptoethanol to a final concentration of 0.7 M and 50  $\mu\text{L}$  isoamyl alcohol. Vortex strongly and centrifuge both samples at  $6,000 \times g$  for 2 min (*see* **Notes 6** and **7**).
9. Discard the supernatant and resuspend the pellet in 1 mL growth medium (e.g., YPD).
10. Add another 50  $\mu\text{L}$   $\beta$ -mercaptoethanol to the DMS-treated sample, vortex, and repeat the centrifugation step ( $6,000 \times g$ , 2 min).
11. Discard the supernatant and freeze the pellet at  $-80\text{ }^{\circ}\text{C}$  for at least 20 min but not longer than overnight.
12. Proceed with RNA isolation (Subheading **3.2.2**).

### 3.1.3 *Pb<sup>2+</sup>-Induced Cleavage Assay of RNA in Yeast*

1. Perform **steps 1–3** as described in Subheading **3.1.2**.
2. Discard the supernatant and resuspend the cell pellet in 0.5 mL pre-warmed (30 °C) growth medium (e.g., YPD).
3. Add 500 µL of 50 mM Pb(OAc)<sub>2</sub> (prepare freshly) to the sample, yielding a final volume of 1 mL and a final concentration of 25 mM Pb(OAc)<sub>2</sub> (*see* **Notes 5** and **8**). To the second sample, add 500 µL ddH<sub>2</sub>O as control.
4. Incubate the samples at 30 °C for 10 min with moderate shaking to prevent cells from pelleting at the bottom.
5. Stop the reaction by adding 125 µL of 0.5 M EDTA, pH 8.0 (2.5× molar excess relative to the [Pb<sup>2+</sup>]) (*see* **Note 6**).
6. Pellet the cells by centrifugation at 4,500×*g* for 5 min and carefully remove the supernatant.
7. Freeze the cells at –80 °C for at least 20 min to overnight.
8. Proceed with RNA isolation (Subheading **3.2.2**).

## 3.2 *Isolating Total RNA from Eukaryotic Cells*

### 3.2.1 *RNA Isolation from HEK Cells*

1. Add 6 mL Isol-RNA lysis reagent to the cell pellet and resuspend it rapidly.
2. Keep the tube at room temperature for 5 min.
3. Add 1.2 mL chloroform and vortex vigorously for 15 s.
4. Repeat **step 2**.
5. Centrifuge at 12,000×*g* for 15 min (4 °C).
6. Transfer the aqueous phase to a new tube.
7. Add 3 mL isopropanol and vortex 15 s. Incubate the tube at room temperature for 10 min. Centrifuge at 12,000×*g* for 10 min (4 °C).
8. Carefully discard the supernatant.
9. Add 1 mL 75 % (v/v) ethanol.
10. Centrifuge at 7,500×*g* for 5 min (4 °C).
11. Remove the supernatant completely and air-dry the RNA pellet for 5 min.
12. Dissolve the RNA in 20 µL ddH<sub>2</sub>O (*see* **Note 9**). Store the RNA at –20 °C.

### 3.2.2 *RNA Isolation from Yeast*

1. Resuspend the frozen cells (*see* **Note 10**) in 600 µL AE buffer and 100 µL 10 % (w/v) SDS.
2. Split the cell suspension into two 1.5 mL tubes containing 700 µL water-saturated phenol, pH 4.5, preheated to 65 °C.
3. Vortex strongly and freeze the tubes in liquid nitrogen (ensure that the content is frozen completely). Thaw the samples at 65 °C, followed by vortexing for 30 s.
4. Repeat the freeze–thaw cycle twice.

5. Shake the samples at 65 °C for 4 min.
6. Centrifuge the samples at 18,000 × *g* for 5 min.
7. Transfer the aqueous phase to a new tube and add 700 μL PCI.
8. Vortex strongly for 30 s and centrifuge the samples at 18,000 × *g* for 5 min.
9. Transfer the aqueous phase to a new tube containing 700 μL CI.
10. Mix the sample thoroughly by vortexing for 30 s and centrifuge again at 18,000 × *g* for 5 min.
11. Transfer the aqueous phase to a new 2.0 mL tube. Precipitate the sample with 2.5× volumes of ethanol/0.3 M NaOAc, pH 5.0, and keep it at -20 °C for at least 60 min.
12. Centrifuge the samples at 18,000 × *g* for 30 min (4 °C) and discard the supernatant. Dry the pellet for 5 min at room temperature and then resuspend it in 100 μL ddH<sub>2</sub>O (*see Note 9*).
13. Remove residual DNA by adding the following: 17.5 μL RNase-free DNase I (2 U/μL), 0.5 μL RNase inhibitor (40 U/μL), and 12 μL 10× DNase I reaction buffer. Incubate the samples on ice for 60 min.
14. Adjust the volume to 200 μL by adding 70 μL ddH<sub>2</sub>O and perform a phenol extraction. Add 200 μL PCI, vortex vigorously and centrifuge at 18,000 × *g* for 5 min. Transfer the aqueous phase to a new tube and add 200 μL CI. After vortexing repeat the centrifugation step and transfer the aqueous phase to a new tube.
15. Precipitate the samples by adding 2 μL 0.5 M EDTA, pH 8.0, and 2.5× volumes of ethanol/0.3 M NaOAc, pH 5.0. Incubate the sample at -20 °C for at least 60 min.
16. Centrifuge samples at 4 °C, 18,000 × *g* for 30 min and discard the supernatant. Dry the pellet for 5 min at room temperature.
17. Resuspend the pellet in 10 μL ddH<sub>2</sub>O (*see Note 9*).

### **3.3 5' End-Labeling of Gene-Specific DNA Primers**

1. Mix the reagents in the following order: 10 μL ddH<sub>2</sub>O, 2 μL 10× T4 PNK buffer, 1 μL 10 μM DNA primer, 6 μL [γ-<sup>32</sup>P]-ATP (10 μCi/μL, 6,000 Ci/mmol), and 1 μL T4 PNK (10 U/μL) (*see Note 11*).
2. Incubate the sample at 37 °C for 30–40 min.
3. Add 1 μL 0.5 M EDTA, pH 8.0, to the sample.
4. Incubate the sample at 95 °C for 1 min.
5. Immediately, place the sample on ice for 2 min.
6. Add 2 μL glycogen (10 mg/mL) and 2.5× volumes of ethanol/0.3 M NaOAc, pH 5.0. Keep the sample at -20 °C for at least 60 min.

7. Centrifuge the sample at 4 °C and 18,000×*g* for 30 min. Carefully remove the supernatant. Dry pellet for 5 min at room temperature and dissolve it in 40 μL ddH<sub>2</sub>O.

### **3.4 Mapping the Sites of Modification or Strand Scission by Reverse Transcription**

#### **3.4.1 Reverse Transcription Using Transcriptor Reverse Transcriptase**

1. Set up the annealing reaction by mixing 2.5 μL RNA (20–40 μg), 1 μL 0.25 μM <sup>32</sup>P-labeled primer (*see Note 12*), and 1 μL 4.5× hybridization buffer.
2. Incubate the sample at 95 °C for 1 min.
3. Immediately place the sample on ice for 2 min.
4. Add 15.5 μL extension mix consisting of 4 μL 5× reaction buffer, 2 μL 10 mM dNTPs, 1 μL 0.1 M DTT, 0.5 μL RNase inhibitor (40 U/μL), 0.5 μL Transcriptor reverse transcriptase (20 U/μL), and 5.5 μL ddH<sub>2</sub>O. Incubate the samples at 50 °C for 60 min. To generate A and C sequencing lanes add 2 μL of a 10 mM ddTTP or ddGTP solution to the sample in addition to the extension mix, respectively (*see Notes 13–15*).
5. Degrade the RNA by adding 3 μL 1 M NaOH and incubate at 50 °C for 45 min.
6. Neutralize the pH with 3 μL 1 M HCl.
7. Precipitate the cDNAs by adding 2 μL glycogen (10 mg/mL), 2 μL 0.5 M EDTA, pH 8.0 and 2.5× volumes of ethanol/0.3 M NaOAc, pH 5.0. Incubate the samples at –20 °C for 1 h to overnight.
8. Centrifuge the sample at 4 °C and 18,000×*g* for 30 min. Discard the supernatant and resuspend the pellet in 8 μL loading buffer.

#### **3.4.2 Reverse Transcription Using AMV Reverse Transcriptase**

1. Set up the annealing reaction as described in **steps 1–3** in Subheading **3.4.1**.
2. Add 2.2 μL extension mix consisting of 1.08 μL ddH<sub>2</sub>O, 0.67 μL 10× extension buffer, 0.3 μL 2.5 mM dNTP mix, 0.15 μL AMV reverse transcriptase (10 U/μL) and incubate the sample at 42 °C for 1 h. To generate A and C sequencing lanes add 0.75 μL of 1 mM ddTTP or 1 mM ddGTP solution to the extension mix, respectively (*see Notes 13–15*).
3. Degrade the RNA by adding 1.5 μL 1 M NaOH and incubate the sample at 42 °C for 60 min.
4. Neutralize the pH with 1.5 μL 1 M HCl.
5. Precipitate the cDNAs by adding 2 μL glycogen (10 mg/mL), 1 μL 0.5 M EDTA, pH 8.0 and 2.5× volumes of ethanol/0.3 M NaOAc, pH 5.0. Incubate the sample at –20 °C for 1 h to overnight.
6. Centrifuge the sample at 4 °C and 18,000×*g* for 30 min. Discard the supernatant, dry the pellet at room temperature for 5 min and resuspend the pellet in 8 μL loading buffer.

### 3.5 Resolving the cDNA Pool on a Denaturing PAGE

1. Assemble the glass plates.
2. Mix 50 mL 8 % denaturing acrylamide solution with 500  $\mu$ L 10 % (w/v) APS (*see Note 4*) and 50  $\mu$ L TEMED.
3. Pour the gel solution immediately between the two glass plates. Avoid introducing air bubbles. Insert a suitable comb and let the gel polymerize for at least 60 min.
4. Place the gel in the electrophoresis apparatus and pre-run the gel for 30–60 min at 40 W (*see Note 16*).
5. Denature the samples at 95 °C for 1 min.
6. Rinse the wells thoroughly with 1 $\times$  TBE buffer.
7. Load 4  $\mu$ L of the sample and run the gel at 40 W for approximately 2 h or until bromophenol blue dye has reached the bottom of the gel.
8. Disassemble the gel by removing the top glass plate and transfer it to Whatman 3MM paper. Cover the gel with saran wrap.
9. Dry the gel in a vacuum dryer at 80 °C for 90 min.
10. Expose the gel to a phosphorimager screen overnight and scan the screen using a STORM 820 or comparable equipment.
11. Analyze the gel using ImageQuant or comparable software (*see Notes 17–19*).

---

## 4 Notes

1. Growth medium containing a carbon source gets easily contaminated. Therefore sugar is added directly prior use.
2. Using raffinose instead of glucose as a carbon source leads to proliferation of yeast mitochondria and in turn to an increased yield of mitochondrial RNA.
3. Care should be taken when working with DMS. DMS is a corrosive, toxic, and potentially carcinogenic substance that is readily absorbed through skin. Therefore DMS should be handled in the hood wearing gloves that do not permit penetration of organic chemicals.
4. The 10 % (w/v) APS solution can be kept at 4 °C for up to 2 weeks. Alternatively, store aliquots of the solution at –20 °C for up to 6 months.
5. The DMS concentration used in the modification reaction has to be optimized for each target RNA prior to experiments. To ensure single-hit conditions (one or less DMS modification event per RNA molecule). The amount of full-length cDNA of samples treated with 25–200 mM DMS should be compared to the RT stop control in order to decide which DMS concentration correlates best with single-hit conditions for the

target RNA. Single hit conditions have to be achieved for the  $\text{Pb}^{2+}$ -induced cleavage assay as well. Again, this is done by performing a concentration and/or time series; useful parameters for such series range from 10 mM to 150 mM  $\text{Pb}(\text{OAc})_2$  as final concentration and incubation times from 5 to 120 min. Like for DMS probing, the amount of full-length cDNA should be comparable between the RT stop control and the modified sample.

6. The DMS reaction has to be stopped efficiently before proceeding with total RNA extraction, to avoid extended modification during the RNA preparation. If this is not the case, the majority of the RNA bases will be modified by DMS during the denaturing conditions of the RNA preparation. To confirm efficient quenching, a stop control has to be carried out in the following manner: the indicated amount of  $\beta$ -mercaptoethanol has to be added prior to DMS and incubated for 2 min at 30 or 37 °C (in case of yeast or HEK cells, respectively). If no DMS pattern is observed for the stop control, this implies that the  $\beta$ -mercaptoethanol has successfully quenched the DMS reactivity and as a consequence it cannot methylate the RNA during subsequent RNA preparation. A stop control has to be prepared for  $\text{Pb}^{2+}$ -induced cleavage assay as well by adding the 2.5× molar excess of 0.5 M EDTA, pH 8.0, prior to  $\text{Pb}(\text{OAc})_2$  to the sample and subsequent incubation at 30 °C for 2 min. If no cleavage pattern is observed, this indicates that the amount of EDTA is sufficient to chelate all  $\text{Pb}^{2+}$ -ions present in the sample.
7. The cells treated with DMS are sometimes difficult to pellet. To avoid losing parts of the pellet, it is suggested to increase the centrifugation force or to increase the time of centrifugation.
8.  $\text{Pb}(\text{OAc})_2$  does precipitate in normal growth medium (e.g., YPD). The amount of precipitate correlates with the concentration of the stock solution added to the medium. We therefore propose to dilute the sample with the  $\text{Pb}(\text{OAc})_2$  stock only 1:2 (e.g., mix 500  $\mu\text{L}$  cells suspension with 500  $\mu\text{L}$  50 mM  $\text{Pb}(\text{OAc})_2$  to achieve a final concentration of 25 mM) to reduce precipitation as much as possible. Furthermore,  $\text{Pb}(\text{OAc})_2$  has a strong preference to form precipitants under the following conditions: (1) the presence of chloride ions in the buffer and (2) as a function of the pH value of the buffers used. The lower the pH, the more precipitate is formed.
9. RNA is very sensitive to degradation. Therefore it is advisable to follow some general considerations when working with RNA: avoid as much as possible RNase contamination by always wearing gloves and by using consumables of “RNase-free” grade. Store the RNA stock at -20 °C and avoid repeated freeze-thaw cycles (aliquot, if necessary). In addition, when handling RNA,

keep it on ice if not stated otherwise. Prepare solutions with RNase-free water or DEPC-treated water. Also, the chemicals have to be of high quality and should be used for RNA only. If affordable, a separate set of pipettes should be used for working with RNA. In case of a severe RNase contamination, it is helpful to clean the equipment and bench thoroughly with “RNaseZap” (Ambion) or a comparable product.

10. When working with yeast mitochondrial RNA, the best results in terms of quality and yield of mtRNA were obtained using the RiboPure Yeast Kit from Ambion. Using the method described in Subheading 3.2.2 we obtained RNA of good quality and high yield compared to all other protocols tested to extract total RNA from yeast.
11. We recommend gel-purification of the DNA primers before end-labeling. The purification is necessary to remove any organic remnants left from oligo synthesis. These could impair the T4 polynucleotide kinase during labeling and may result in a low signal-to-noise ratio. We purify DNA primers (18–30 nucleotides) on a 20 % denaturing polyacrylamide gel run with 25 W. Purification is time-consuming and should therefore be done ahead of time, at least 1 day prior to the labeling reaction.
12. Working with total RNA can be problematic due to mispriming events during reverse transcription. It is therefore advisable to prepare sequencing lanes from in vitro transcribed RNA and compare them with those obtained from in vivo isolated RNA, thereby confirming that the primer used does only bind to the RNA of interest. Otherwise the obtained DMS pattern might be misleading and/or inconclusive. Simple comparison of the band pattern in the sequencing lanes and the gene sequence is often insufficient.
13. The cDNA synthesis can be accomplished by different reverse transcriptases and is influenced by several parameters, like RNA sequence, highly structured RNA elements, low amount of target RNA, reaction temperature and ionic concentrations, among others. AMV reverse transcriptase is the most commonly used enzyme. For highly structured or GC-rich RNA, a very sensitive and thermostable enzyme like Transcriptor reverse transcriptase should be used instead of AMV reverse transcriptase. Protocols for both enzymes are provided.
14. Include a stop control during reverse transcription. This is absolutely necessary for distinguishing between so-called natural stops of the reverse transcriptase enzyme and those induced by DMS modification or Pb<sup>2+</sup>-cleavage, respectively. Secondary structures that form after denaturing of the RNA as well as G/C sequence stretches can block the reverse transcriptase enzyme, thereby terminating the extension. Also, RNAs tend to have “hot spots” for breakage due to a strain on

the backbone. Along this line, partial degradation by RNases results in RT stops as well. By reverse transcribing RNA from cells that were not treated with DMS or Pb<sup>2+</sup>, these natural stops of different origin can be easily detected.

15. To generate sequencing lanes of good quality, the ddNTP concentration has to be adjusted for the respective RNA. If the concentration is inadequate, sequencing lanes tend to either give a very faint signal in the upper part of the gel or some nucleotides remain undetected (even though the flanking residues are well detectable).
16. Pre-run the gel for at least 30 min before loading the samples to reach optimal temperature (45–55 °C). A temperature indicator attached to the glass plates is recommended to ensure that the optimal temperature is applied. Higher temperature might damage the glass plates and cause bands to smear and therefore reduces the resolution of the gel. On the other hand uneven heat distribution during gel electrophoresis causes the “smile effect.” Attaching a metal plate to the glass ensures a better heat distribution and thus improves the gel quality.
17. If the cDNA sample has a high salt content, this results in narrowing of the lanes during the gel electrophoresis. To avoid this problem, the cDNA pellet should be washed with 70 % (v/v) ethanol before resuspending it in loading buffer. The resolution of the gel is also drastically reduced, if the cDNA sample is precipitated for longer than overnight. Alternatively, it is possible to overcome this problem by increasing the percentage of the gel and the running time, as the salt front will run out of the gel.
18. A low signal-to-noise ratio usually points to the need for optimization of different steps. First, check the <sup>32</sup>P-labeling efficiency by comparing the amount of radioactivity of the supernatant versus the pelleted DNA oligo in the labeling step. Notably, the T4 PNK enzyme and the PNK buffer are both very sensitive to changes in temperature. Aliquoting the buffer helps avoiding frequent freeze–thaw cycles. Furthermore, it is advisable to use the radioactive stock no later than at its calibration date (the <sup>32</sup>P-isotope has a short half-life of 14.3 days). If no improvement is observed, several other parameters can be adjusted: the choice of reverse transcriptase used (AMV vs. Transcriptor reverse transcriptase); the amount of RNA and primer used for reverse transcription reaction; also the primer binding site and its melting temperature should be reevaluated along with the annealing and extension temperature, respectively; dNTP concentration and/or the preparation date of the stock.
19. Quantification is highly dependent on loading equal amounts of sample on each lane. In case of uneven loading, the second half of the sample can be used for a rerun with adjusted sample amount.

## Acknowledgement

This work was funded by the Austrian Science Foundation (FWF; grants Y401 and P21017 to C.W.).

## References

1. Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137
2. Herschlag D (1995) RNA chaperones and the RNA folding problem. *J Biol Chem* 270:20871–20874
3. Pyle AM, Fedorova O, Waldsich C (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem Sci* 32:138–145
4. Schroeder R, Barta A, Semrad K (2004) Strategies for RNA folding and assembly. *Nat Rev Mol Cell Biol* 5:908–919
5. Sosnick TR, Pan T (2003) RNA folding: models and perspectives. *Curr Opin Struct Biol* 13:309–316
6. Treiber DK, Williamson JR (1999) Exposing the kinetic traps in RNA folding. *Curr Opin Struct Biol* 9:339–345
7. Treiber DK, Williamson JR (2001) Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* 11:309–314
8. Woodson SA (2000) Compact but disordered states of RNA. *Nat Struct Biol* 7:349–352
9. Woodson SA (2005) Structure and assembly of group I introns. *Curr Opin Struct Biol* 15:324–330
10. Woodson SA (2010) Compact intermediates in RNA folding. *Annu Rev Biophys* 39:61–77
11. Schroeder R, Grossberger R, Pichler A et al (2002) RNA Folding *in vivo*. *Curr Opin Struct Biol* 12:296–300
12. Zemora G, Waldsich C (2010) RNA folding in living cells. *RNA Biol* 7:634–641
13. Ares M Jr, Igel AH (1990) Lethal and temperature-sensitive mutations and their suppressors identify an essential structural element in U2 small nuclear RNA. *Genes Dev* 4:2132–2145
14. Balzer M, Wagner R (1998) Mutations in the leader region of ribosomal RNA operons cause structurally defective 30 S ribosomes as revealed by *in vivo* structural probing. *J Mol Biol* 276:547–557
15. Climie SC, Friesen JD (1988) *In vivo* and *in vitro* structural analysis of the rplJ mRNA leader of *Escherichia coli*. Protection by bound L10-L7/L12. *J Biol Chem* 263:15166–15175
16. Doktycz MJ, Larimer FW, Pastrnak M et al (1998) Comparative analyses of the secondary structures of synthetic and intracellular yeast MFA2 mRNAs. *Proc Natl Acad Sci USA* 95:14614–14621
17. Harris ME, Pace NR (1995) Analysis of the tertiary structure of bacterial RNase P RNA. *Mol Biol Rep* 22:115–123
18. Liebeg A, Mayer O, Waldsich C (2010) DEAD-box protein facilitated RNA folding *in vivo*. *RNA Biol* 7:803–811
19. Mereau A, Fournier R, Gregoire A et al (1997) An *in vivo* and *in vitro* structure-function analysis of the *Saccharomyces cerevisiae* U3A snoRNP: protein-RNA contacts and base-pair interaction with the pre-ribosomal RNA. *J Mol Biol* 273:552–571
20. Moazed D, Robertson JM, Noller HF (1988) Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. *Nature* 334:362–364
21. Senecoff JF, Meagher RB (1992) *In vivo* analysis of plant RNA structure: soybean 18S ribosomal and ribulose-1,5-bisphosphate carboxylase small subunit RNAs. *Plant Mol Biol* 18:219–234
22. Waldsich C, Grossberger R, Schroeder R (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the *td* group I intron *in vivo*. *Genes Dev* 16:2300–2312
23. Waldsich C, Masquida B, Westhof E et al (2002) Monitoring intermediate folding states of the *td* group I intron *in vivo*. *EMBO J* 21:5281–5291
24. Wells SE, Hughes JM, Igel AH et al (2000) Use of dimethyl sulfate to probe RNA structure *in vivo*. *Methods Enzymol* 318:479–493
25. Zaug AJ, Cech TR (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
26. Brunel C, Romby P (2000) Probing RNA structure and RNA-ligand complexes with chemical probes. *Methods Enzymol* 318:3–21
27. Ehresmann C, Baudin F, Mougel M et al (1987) Probing the structure of RNAs in solution. *Nucleic Acids Res* 15:9109–9128
28. Erat MC, Sigel RK (2011) Methods to detect and characterize metal ion binding sites in RNA. *Met Ions Life Sci* 9:37–100
29. Lindell M, Brannvall M, Wagner EG et al (2005) Lead(II) cleavage analysis of RNase P RNA *in vivo*. *RNA* 11:1348–1354

30. Lindell M, Romby P, Wagner GH (2002) Lead (II) as a probe for investigating RNA structure *in vivo*. *RNA* 8:534–541
31. Sigel RK, Pyle AM (2003) Lanthanide ions as probes for metal ions in the structure and catalytic mechanism of ribozymes. *Met Ions Biol Syst* 40:477–512
32. Huang HR, Rowe CE, Mohr S et al (2005) The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc Natl Acad Sci USA* 102:163–168
33. Gregan J, Kolisek M, Schweyen RJ (2001) Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing *in vivo*. *Genes Dev* 15:2229–2237
34. Halls C, Mohr S, Del Campo M et al (2007) Involvement of DEAD-box proteins in group I and group II intron splicing. Biochemical characterization of Mss116p, ATP hydrolysis-dependent and -independent mechanisms, and general RNA chaperone activity. *J Mol Biol* 365:835–855
35. Solem A, Zingler N, Pyle AM (2006) A DEAD protein that activates intron self-splicing without unwinding RNA. *Mol Cell* 24:611–617
36. Sigel R, Vaidya A, Pyle A (2000) Metal ion binding sites in a group II intron core. *Nat Struct Biol* 7:1111–1116

## VI. Acknowledgements

*“We choose to go to the moon [...] and do the other things, not because they are easy, but because they are hard”*

John F. Kennedy

Doing a PhD felt sometimes like going to the moon. Throughout this journey a tremendous number of challenges must be accomplished in order to be successful and finally obtain one’s PhD. Luckily, I had and still have a number of great people I could always rely on and who helped me to never lose interest in science. These people I owe a debt of gratitude. In particular, I would like to thank my supervisor Dr. Mag. Christina Waldsich for giving me the opportunity to work on an extremely interesting topic. I am grateful for the countless scientific discussions and the excellent guidance I was provided with. I really enjoyed the opportunity to work independently on my project, and nevertheless knowing that all my questions, doubts and suggestions are always welcome.

I also would like to thank the members of my PhD committee: Prof. Renée Schroeder and Prof. Eric Westhof, who took the time to listen to my ideas and gave valuable scientific input for my project.

Without doubt, I have to thank my family, especially my parents, for showing interest in the research I was conducting as well as providing motivation in times of failure and set-backs. This kind of support made it much easier for me.

Lastly, I would like to thank my wife Victoria for her ability to listen to long monologues about my research, her constant support and simply for being part of my life. I couldn’t have done it without you.

The final thank you goes to all of my colleagues and friends, especially the members of the Waldsich lab. I really enjoyed our discussion about science, the meaning of life and board games. You have created a pleasant working atmosphere and I enjoyed every day of it.



07/2007 – 08/2006	blotting, allgemeine Laborarbeit
2000 bis 2005	Diverse Ferialjobs, unter anderem Schichtarbeit bei VOEST Alpine

---

## Zusätzliche Qualifikationen

Sprachen	Deutsch (Muttersprache), Englisch (fließend in Wort und Schrift), Französisch (Maturaniveau)
Soziale Kompetenzen	Mitarbeiterführung, erworben als Tutor und Betreuer von Bachelor und Masterstudenten. Teamfähigkeit, Präsentationstechnik
Fachliche Zusätze	Grundlagen im Patentrecht und Projektmanagement, erworben durch Fortbildungsseminare
Publikationen	Cruz et. al., 2012. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. RNA 2012. 18: 610-62  Wildauer et al., 2014, Chemical probing of RNA in living cells, Methods Mol Biol. 014;1086:159-76
Konferenzen	Seventeenth annual meeting of the RNAsociety, Ann Arbor, 2012  Telomeres and Telomerase 2013, Cold Spring Harbor, 2013  EMBO conference on telomeres, telomerase and disease, Bussels, 2014
EDV	Souveräner Umgang mit MS Office und Corel Draw X4
Tätigkeiten	Mitglied im Organisationskomitee für den PhD & PostDoc Retreat 2012 und 2013  Moderator European Youth Parlimant 2011 zu Wissenschaft und Gesellschaft
Hobbies	Laufsport, Brettspiele, Reisen und Wandern