NON-CANONICAL STRUCTURES AND FUNCTIONS OF THE HUMAN RIBOSOME: G-QUADRUPLEXES AND HEME APPROPRIATION

A Dissertation
Presented to
The Academic Faculty

by

Santiago Mestre Fos

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry & Biochemistry

Georgia Institute of Technology

August 2020

NON-CANONICAL STRUCTURES AND FUNCTIONS OF THE HUMAN RIBOSOME: G-QUADRUPLEXES AND HEME APPROPRIATION

Approved by:

Dr. Loren D. Williams, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Amit R. Reddi
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Francesca Storici
School of Biological Sciences
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biological Sciences
Georgia Institute of Technology

Dr. Adegboyega K. Oyelere
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: June 30th, 2020

This thesis is dedicated to my aunt Assumpta Fos Escrivà and her family (Joan Salvador, Anna, and Antoni) and my family (Manuel, Ana, and Carlos).

“You can’t start a fire without a spark.”

– Bruce F. J. Springsteen

ACKNOWLEDGMENTS

The work presented in this dissertation would not had been possible without the contributions of many scientists. Although I appear as the only author of this thesis, the help provided by numerous graduate students, postdoctoral fellows, research technicians and professors that I have met at Georgia Tech and elsewhere has been essential for its development and completion. For this reason, first I would like to acknowledge my advisor Dr. Loren Dean Williams and thank him for the all the guidance, mentorship, and knowledge provided during the past five years. I am very grateful for all the personal and professional invaluable advice he has given me. Graduate school has been an incredible and exciting journey mostly thanks to Dr. Williams. He has provided me with an excellent and positive environment where I have been able to grow as a scientist. His passion for science (and ribosomes) has deeply influenced me and I look forward to continuing using this passion for science in the future. I could not have asked for a better advisor than Dr. Williams.

I would also like to thank my committee members, Drs. Amit R. Reddi, Roger M. Wartell, Nicholas V. Hud, Francesca Storici, and Adegboyega Oyelere for their encouragement and personal and academic advice.

I would like to extend my gratitude to our collaborators Drs. Jonathan B. Chaires and Amit R. Reddi. Dr. Chaires has contributed very significantly to the development of my thesis, he brought encouragement when I needed it the most. I am honored to have collaborated and written a paper with him. Dr. Reddi has also been instrumental in the development of my thesis, especially the last chapter. Thanks to his expertise in the field

of cell biology we have been able to perform experiments and discover ribosomal functions that we could not have uncovered without his help. It has also been an honor to collaborate with Dr. Reddi and I look forward to seeing his new discoveries in the fields of heme and Sod1 biology.

Although a Ph.D. thesis is the result of the experiments performed individually, I have been extremely lucky to have been surrounded by many scientists that have provided me with an enormous amount of help and advice throughout these years. For this reason, I would like to thank all Williams lab members that I have met during my time in graduate school. Specifically, I would like to thank Ms. Jessica Bowman, Ms. Chieri Ito, and Dr. Lizzette Gómez-Ramos. Jessica and Chieri have always given me invaluable help and advice that has allowed me to grow both as a scientist and as a person. I am forever grateful for all the guidance they have given me during the ups and downs of graduate school. I was also extremely lucky to have Lizzette as my mentor. She was one of the hardest working graduate students I have ever met so having her mentoring me deeply influenced the way I have approached every day of work in lab. From her I learned that scientific curiosity always has to be supplemented with hard work. Thank you.

As students we are the result of the teachers that have influenced us the most. For this reason, I would like to thank Dr. Roser Nebot Castelló and Dr. John Haseltine. Dr. Nebot was my science teacher in high school. She is one of the most curiosity-driven scientists I know. She deeply influenced my early interests in biology; I would not had decided to pursue a career in science if it was not for her amazing lectures on geology and volcanoes. Dr. Haseltine was my organic chemistry professor in college. The time I spent in his

laboratory performing organic synthesis research convinced me to apply for graduate school. I would not be here without him either.

Despite being an exciting experience, graduate school also comes with lots of frustration due to negative and unexpected experimental results. I have found that having the help of your friends is a great way to overcome this frustration. The ones I made at Georgia Tech are one of the things I will remember the most about graduate school. For this reason, I would like to thank my great friends Dr. Abraham Jordan, Dr. Osiris Martínez-Guzman, Dr. Zachary Hood, and soon-to-be Drs. Eric Drew, Samuel Evans, and Stephen Zambrzycki for all the great times spent together. I will always remember our hiking trips and bonfires. I will miss them greatly.

During graduate school I have also met my girlfriend Claudia. Meeting her has been the best experience of graduate school. I am very grateful for all the support she has provided me throughout these years.

Lastly, I would like to thank my parents Manuel and Ana and my brother Carlos. Leaving Barcelona behind to start a new life in the US was difficult and it took a significant amount of effort from the four of us. I think one of the reasons why my parents decided to try the American adventure was to give me and my brother a better future and broaden our horizons. I hope this Ph.D., along with the great achievements that Carlos will accomplish in the near future, represents something similar to what they had in mind when they decided to start a new chapter in life across the pond. Carlos and I are forever grateful for their courageousness and for everything they have sacrificed in their lives for us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

°C	Degrees Celsius
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometer
A_{295}	Absorbance at 295 nm
ACN	Acetonitrile
ALA	5-aminolevulinic acid
AUC	Analytical Ultracentrifugation
BSA	Bovine Serum Albumin
CD	Circular Dichroism
cm	Centimeter
Cryo-EM	Cryogenic Electron Microscopy
C(t)	Threshold Cycle Value
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole

DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced green fluorescent protein
EMSA	Electromobility Shift Assay
ER	Endoplasmic Reticulum
ES	Expansion Segment
ES27	Expansion Segment 27
es3	Expansion Segment 3
es6	Expansion Segment 6
ES7	Expansion Segment 7
es12	Expansion Segment 12
E. coli	*Escherichia coli*
FPLC	Fast Protein Liquid Chromatography
G4	G-quadruplex

g	Gram
G-quadruplex	Guanine quadruplex
G-tract	Guanine tract
HD	Heme depletion
HEK293	Human Embryonic Kidney cells
H. sapiens	*Homo sapiens*
HS1	Heme sensor 1
KDa	Kilodalton
L	Liter
LSU	Large Subunit
M	Molar
MBN	Mung Bean Nuclease
mdeg	Millidegree
mL	Milliliter
mM	Millimolar
mRNA	Messenger RNA
MS	Mass Spectrometry

MSA	Multiple Sequence Alignment
NCBI	National Center for Biotechnology Information
nm	Nanometer
nM	Nanomolar
OD_{600}	Optical Density read at 600 nm
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDS	Pyridostatin
PTC	Peptidyl Transferase Center
rDNA	Ribosomal DNA
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
rProtein	Ribosomal Protein
rRNA	Ribosomal RNA
RT-qPCR	Quantitative Reverse Transcription PCR
S. cerevisiae	*Saccharomyces cerevisiae*

SA	Succinylacetone
SASA	Solvent-Accessible Surface Area
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEREB	Sparse and Efficient Representation of the Extant Biology
SILAC	Stable Isotope Labeling by Amino Acids in Cell Culture
SSU	Small Subunit
T. thermophilus	*Thermus thermophilus*
ThT	Thioflavin T
T_m	Melting Temperature
tRNA	Transfer RNA
UTR	Untranslated Region
UV	Ultraviolet
WT	Wild Type

SUMMARY

The current work describes the *in vitro* and *in vivo* G-quadruplex capabilities of human ribosomal RNAs (rRNAs) and investigates their involvement in heme bioavailability regulation. This is the first report on G-quadruplex formation on any rRNA, the most abundant RNA in cells of all living organisms on Earth (>80% of RNA is rRNA). Prior to this thesis work, RNA G-quadruplex research was mainly focused on messenger RNAs (mRNAs) partly due to next-generation sequencing (NGS) studies performed at the transcriptome-wide level according to which putative G-quadruplex regions are enriched in 5' and 3' untranslated regions (UTRs) of specific transcripts. In NGS studies it is a common practice to explicitly remove rRNAs due to their massive abundance. Doing so ensures efficient detection of other less abundant RNAs, such as mRNAs. However, it also eliminates the most abundant ones and it is one of the reasons why rRNAs have been put into oblivion within the G-quadruplex field. Another potential reason why rRNA G-quadruplexes had not been reported before our work was published is the fact that rRNA sizes differ very significantly across phylogeny, making the identification of rRNA G-quadruplex regions difficult. Eukaryotes (particularly metazoans) present massive rRNAs compared to bacteria. As it will be explained in the Introduction section of this thesis, this increase in size is mainly due to particular and enigmatic extensions in the form of double-stranded rRNA tentacles called expansion segments (ESs). Thus, unless rRNAs are studied across the entire tree of life, it is extremely challenging to identify specific sequence patterns such as G-quadruplex regions.

Professor Loren Williams and coworkers have been investigating the evolution of the ribosome at the molecular level by looking at ribosomal 3D structures from species across at the entire tree of life. By doing so, they have been able to investigate the molecular mechanisms by which eukaryotic rRNAs increase in size and, therefore, appreciate their changes in structure and sequence across phylogeny. When this thesis work started, in the Williams lab we realized that guanine-rich human rRNA expansion segments presented repetitive patterns of guanine runs near the termini of specific tentacles of the two largest expansion segments of the human ribosome (ES7 and ES27). These patterns suggested potential G-quadruplex capabilities, albeit rRNAs had never been reported to form these non-canonical secondary structures.

By applying a wide variety of biochemical assays such as circular dichroism, ultraviolet melting experiments, the use of fluorescence probes, electromobility shift assays with a G-quadruplex antibody, and competition experiments, we were able to detect the *in vitro* formation of highly stable G-quadruplexes in human rRNAs. We identified numerous G-quadruplex regions in human LSU rRNA (28S) and several other ones in human SSU rRNA (18S). Our results indicate that intact human 28S and 18S rRNAs extracted from human cells can form stable G-quadruplexes *in vitro*. We then tested their formation in cells. Our immunofluorescence and RNA pulldown experiments suggest these rRNAs can also form G-quadruplexes when in cells.

Finally, we studied some of their potential physiological roles. To test this, we considered that these regions are located on the surface of the human ribosome. Therefore, we reasoned that one of their roles could be to recruit specific proteins to the ribosome. By using mass spectrometry, we identified the cytosolic proteins that bind to one of the G-

quadruplex regions on the 28S rRNA, identifying several RNA helicases (such as DEAD-box helicases DDX3X or DDX17) , other RNA remodeling proteins such as CNBP, and proteins involved in phase separation systems like Fus.

In addition, since G-quadruplexes are known to be capable of aggregating via inter-molecular association, we tested the potential involvement of ribosomal G-quadruplexes in polysome association. Our dot blots with a G-quadruplex antibody suggest rRNA G-quadruplexes form more easily in polysomes than in monomer ribosomes.

Lastly, based on the high affinity of heme for G-quadruplexes *in vitro* ($K_D \sim 10$ nM), our work demonstrating that rRNA forms extensive G-tracts *in vitro* (Chapters 1 and 2), the extreme stabilities of rRNA G4s *in vitro,* and the extraordinary abundance of rRNA *in vivo* we hypothesized that rRNA G-quadruplexes function *in vivo* to regulate intracellular heme bioavailability. This work (described in detail in Chapter 4) provides strong evidence that G-quadruplexes in human 28S and 18S rRNAs appropriate heme and regulate the bioavailable levels of labile heme in the cytosol, making the ribosome a hub for heme metabolism.

Altogether, completion of this work opens new sets of questions on ribosomal and G-quadruplex functionalities. Since rRNA is the most abundant RNA in cells, our results suggest that the so-called "RNA G-quadruplexome" (the entire body of cellular RNAs capable of forming G-quadruplexes) is ribosome-centered. Indeed, our immunofluorescence experiments suggests ~83% of the extra-nuclear RNA G-quadruplexes are associated with ribosomes. In addition, our work suggests rRNA G-quadruplexes are sites of exchangeable heme, appropriating it to the ribosome. We propose

that heme-rRNA G-quadruplex interactions may be important for protein hemylation reactions and buffering cytosolic heme.

CHAPTER 1

INTRODUCTION

1.1.The Ribosome

Ribosomes are complex macromolecular machines responsible for all protein synthesis in cells. Ribosomes are ribonucleoproteins, which means they are made both of RNA (ribosomal RNA, or rRNA) and proteins (ribosomal proteins, or rProteins). They were first identified as RNA-rich particles in close proximity to the endoplasmic reticulum (1). In 1958, Francis Crick proposed that ribosomes, which he named “microsomal particles”, could be involved in gene expression and suggested for the first time that proteins could be made by an RNA machine (2). Four decades later, Harry Noller reported that ribosomal activity was unusually resistant to proteases (enzymes that digest proteins), indicating that rRNA was the main ribosomal component involved in peptide bond formation (3).

Every single cell of any living organism on Earth presents ribosomes. Each ribosome is composed of two subunits that, in combination with transfer RNA (tRNA) and messenger RNA (mRNA), assemble when protein synthesis is required. The large subunit (LSU) is responsible for peptide bond formation whereas the small subunit (SSU) contains the decoding center, where tRNA and mRNA bind. Each subunit is composed of rRNAs and rProteins and they are commonly identified by their sedimentation coefficient (in Svedbergs, S). In bacteria, the LSU (50S) is composed of two rRNAs (23S and 5S) and 31 rProteins, whereas the small subunit (30S) is made of one rRNA (16S) and 21 rProteins (4). Eukaryotic ribosomes are larger in size and present higher structural complexity. For

example, the LSU (60S) of the *Homo sapiens* ribosome is composed of three rRNAs (28S, 5.8S, and 5S) and over 50 rProteins whereas the SSU (40S) is made of one rRNA (18S) and 33 rProteins (4).

1.1.1 rRNA size across phylogeny

Although all rRNAs across phylogeny have conserved functions (LSU rRNAs: peptide bond formation, SSU rRNAs: decoding mRNA) their difference in size varies very significantly across the tree of life (Figure 1.1) (5). Bacterial rRNAs, for example, are small (23S: ~2,900 nucleotides or nts, 16S: ~1,600 nts). Yeast (a eukaryote) presents considerably larger rRNAs (25S: ~3,400 nts, 16S: ~1,800 nts) and the human ribosome presents the largest rRNAs known in nature (28S: ~5,000 nts, 18S: ~1,870 nts). This difference in size is very apparent when the secondary structures of these rRNAs are compared to each other (Figure 1.2) (5).

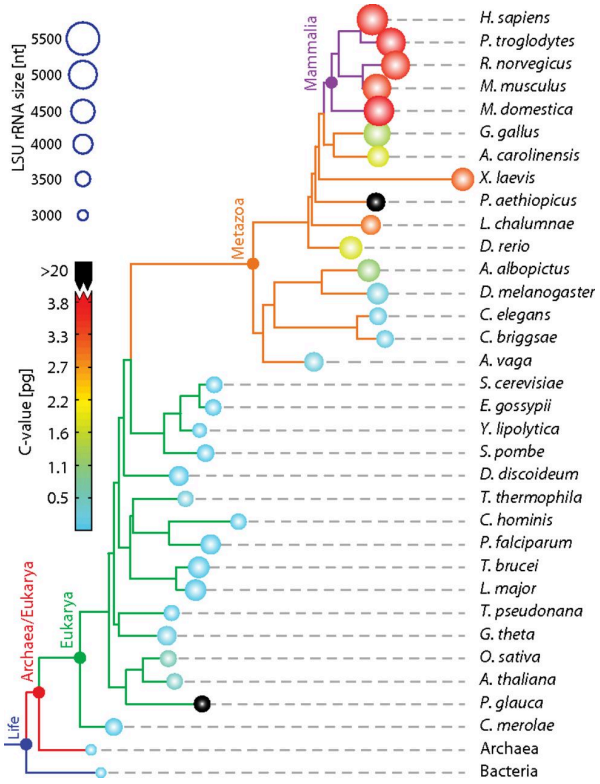


Figure 1.1. Phylogenetic tree indicating the sizes of LSU rRNAs and of genomes. Reprinted with permission (5).

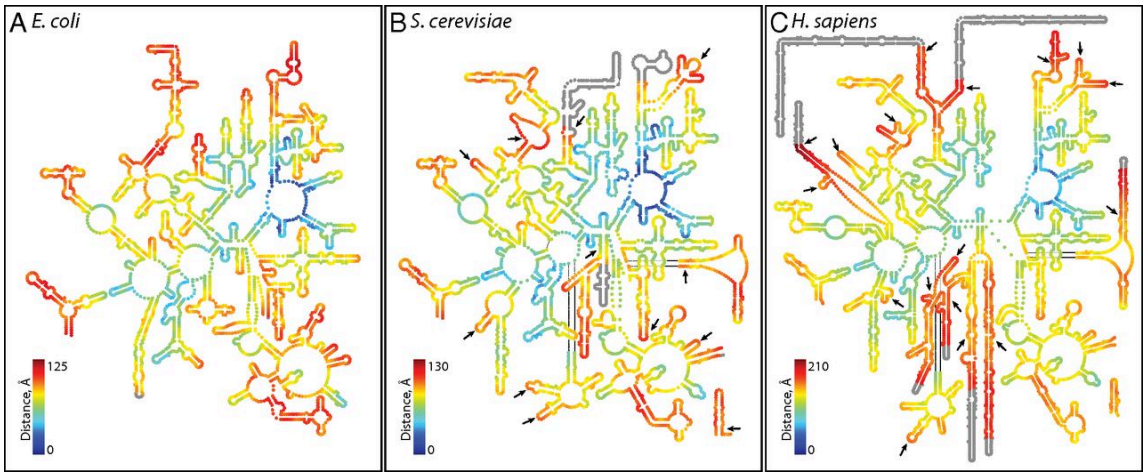


Figure 1.2. LSU rRNA secondary structures. (A) *E. coli*, (B) *S. cerevisiae*, (C) *H. sapiens*. Reprinted with permission (5).

Data in Figure 1.2. shows that the increase in size occurs through the elongation of rRNA in the form of tentacle-like double-stranded ramifications that protrude from the ribosomal surface. These regions are called rRNA expansion segments.

1.1.2 rRNA expansion segments: structure and known functions

Eukaryotic rRNA expansion segments (ESs) were discovered when comparative studies on rRNA secondary structures between bacteria and eukaryotes were performed (6-9). In general, ESs do not perturb the ribosomal core. Eukaryotic ribosomes present numerous ESs and each one of them is identified and numbered. Across phylogeny, expansion segments 7 (ES7) and 27 (ES27) experience the most dramatic elongation processes. The serial increase of ES7, for example, was monitored by comparing the ES7 3D structures across the tree of life (Figure 1.3) (5).

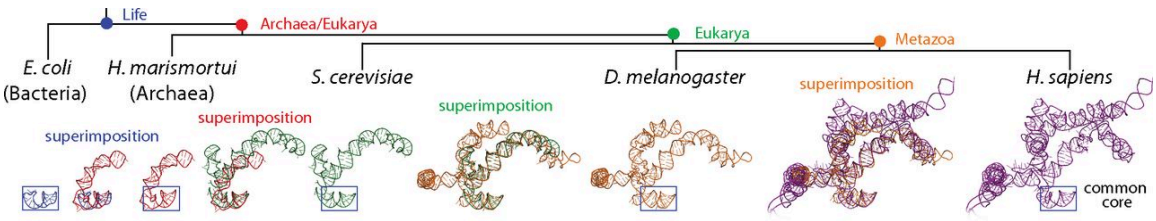


Figure 1.3. Serial accretion of rRNA helix 25/ES7 across phylogeny. Reprinted with permission (5).

ES7 is formed by rRNA accretion from a 22-nucleotide stem loop (helix 25) in the last universal common ancestor (LUCA), which is approximated by *E. coli*. ES7 grows to a double-stranded 210-nucleotide ramification in yeast. In metazoans, ES7 elongation is more extreme and starts to present long surface-exposed tentacles, becoming a 342-nucleotide long structure in *D. melanogaster* and 876 nucleotides in humans. *H. sapiens* presents the largest ES7 known in nature.

The extreme cellular concentrations of rRNA (>80% of RNA is rRNA) (10) and the fact that eukaryotes present repetitive rDNA loci on multiple chromosomes (11) make the genetic modification of rDNA very challenging and hence the study of the potential physiological roles of ESs is incredibly difficult. *S. cerevisiae*, on the other hand, presents a single rDNA locus in chromosome XII (12). This has allowed researchers to entirely delete this region and supplement cells with an rDNA plasmid (13). Large deletions of yeast ESs result in defects in ribosome biogenesis (14). By deleting distal regions of yeast ES27, Maria Barna and co-workers found ES27 is involved in translation fidelity (13). Specifically, ribosomes with deletions on ES27 presented amino acid misincorporation as well as readthrough and frameshifting errors. As an alternative way of studying the cellular functions of ESs, Loren Williams and co-workers performed RNA pull-downs with *in-vitro*-synthesized yeast ES7 and identified the cytosolic proteins that interact with it using

mass spectrometry (15). Taken together, these studies have provided new insights on the mysterious roles of ESs of the yeast ribosome. However, yeast ribosomes present significantly shorter ESs when compared to those of humans (e.g. ES7: 210 nts in yeast, 876 nts in humans). The fact that human ESs, being significantly longer than those in yeast, suggests they could have additional cellular functions. However, as aforementioned, the repetitive nature of rDNA loci and their location on multiple chromosomes makes the study of human ESs even more difficult than those in yeast and it is the reason why their functions remain mostly unknown and enigmatic. Interestingly, most human ESs are rich in guanine and cytosine residues (>80% GC).

1.2. G-quadruplexes in biology

Guanine-rich nucleic acid sequences can form non-canonical secondary structures known as guanine quadruplexes (or G-quadruplexes) (16). Four guanine residues interact with each other through Hoogsteen base-pairing, forming a G-quartet. When two or more G-quartets π-stack onto each other, a G-quadruplex is formed (Figure 1.4). The central cavity of a G-quartet sequesters monovalent cations, with potassium being the one that favors G-quadruplex formation the most.

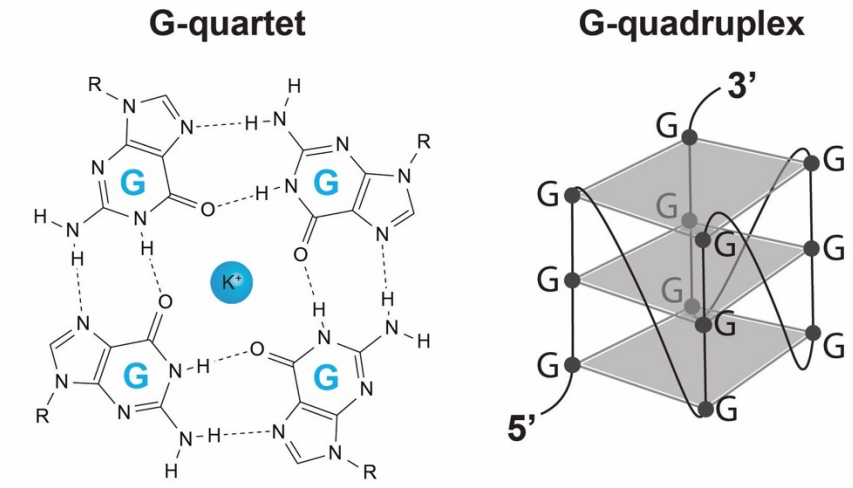


Figure 1.4. Representation of a G-quartet and a G-quadruplex. Adapted from (17) with permission.

Putative G-quadruplex sequences fall within the motif $G_xN_{1-7}G_xN_{1-7}G_xN_{1-7}G_x$, where x is usually at least 3 and N means any nucleotide. Such motifs allow for the formation of "canonical" G-quadruplexes (made of at least 3 G-quartets). However, recent studies have shown that "non-canonical" G-quadruplexes, with G-tracts shorter than 3 guanine residues and connecting loops longer than 7 nucleotides, also exist (18-21). Although both DNA and RNA can form G-quadruplexes, those in RNA have been reported to be more stable than those in DNA, presumably through additional hydrogen bonds due to the presence of the 2' hydroxyl group in RNA (22, 23).

G-quadruplex research was predominantly focused on *in vitro* experimentation until a G-quadruplex specific antibody was developed and used to demonstrate that both DNA and RNA G-quadruplexes can form in fixed human cells (24, 25). Since then, several methods have been developed to study the formation of G-quadruplexes in cells, such as RNA pulldowns using a biotin tagged G-quadruplex ligand (26, 27), fluorescent probes (28, 29), or using ^{19}F nuclear magnetic resonance (NMR) (30). However, while more studies reporting new potential roles of G-quadruplexes in physiology were published, a thought provoking article challenged the entire G-quadruplex field (31). By using dimethyl sulfate (DMS) footprinting and next-generation sequencing, Guo and Bartel provided a model according to which all G-quadruplexes are unfolded in eukaryotic cells. Their data suggested G-rich sequences were capable of forming stable G-quadruplexes *in vitro* but when in cells their chemical probing techniques suggested these same sequences were unfolded. Most interestingly, G-rich sequences folded into G-quadruplexes when transformed into bacterial cells, which led them to propose a model according to which

eukaryotic cells have developed a robust and precise machinery involved in unfolding G-quadruplexes, most likely with the direct participation of RNA helicases and other RNA remodeling factors. Bacteria would lack this machinery and hence G-quadruplex sequences are able to fold in bacterial cells, according to their experiments and model. Although some concerns regarding their experimental approach have been raised (32), this study has been very instrumental and has pushed the G-quadruplex field forward very significantly. It is possible that G-quadruplexes form only transiently, and their formation might be highly regulated by helicases and other G-quadruplex unwinding factors.

Other key methods used to identify putative DNA and RNA G-quadruplex sequences at the genome- and transcriptome-wide levels, respectively, involve next-generation sequencing (33, 34). RNA G-quadruplex next-generation sequencing (rG4-seq) showed G-quadruplexes are enriched in 5' and 3'-untranslated regions (UTRs) (34), in agreement with previous results that demonstrated that G-quadruplexes formed in specific 5'-UTRs and inhibited translation of those transcripts (35-40). Due to the significant finding of G-quadruplex-mediated regulation of protein translation, most RNA G-quadruplex work has been centered on mRNAs. However, G-quadruplexes have also been reported to form in other RNAs such as microRNAs (41-43), human telomerase RNA (44, 45), and viral genomes (46, 47). Other known functions of G-quadruplexes involve promoting phase separation (48, 49) and binding heme, an interaction that can promote oxygen transfer reactions (50-52).

1.3. Heme in biology

Hemes are iron-coordinated porphyrins essential for life in every living organism that metabolizes oxygen. In mammals, most of the cellular iron is heme iron due to its role in oxygen transport in hemoglobin. In humans, about 95% of the functional iron is bound to heme (53). There are several types of hemes with different shapes and roles (54). The most abundant one is heme b, which acts as a cofactor in most globin proteins, such as myoglobin and hemoglobin. Hemes *a* and *c* are synthesized from heme *b* through specific side chain modifications. The former acts as a cofactor in mitochondrial cytochrome *c* oxidase whereas the latter is known to form covalent bonds with the hemoproteins it interacts with, such as cytochrome *c* (54).

The total pool of cellular heme is the sum of inert and labile heme (LH). Inert heme is more abundant, but it is not available for exchange in heme-dependent processes since it is bound with high-affinity to heme-binding proteins, such as cytochromes and globins. LH is less abundant, but it interacts with hemoproteins with less affinity and hence it is exchangeable and trafficked in cells through unknown buffering factors (55). The roles of heme as a cofactor (e.g. oxygen transport) and as signaling molecule (e.g. regulating transcription factors, kinases, etc. (56, 57)) make it an essential molecule. However, these properties also make heme cytotoxic: it can catalyze the formation of reactive oxygen species (ROS) (58) and bind to proteins and alter their structure and function (59). Therefore, heme bioavailability and trafficking are processes that have to be highly regulated. The identification of these factors is mostly focused on proteins, such as GAPDH (60), GTPases (61), and enzymes involved in heme biosynthesis (62). However, other biomolecules could be involved in the appropriation and trafficking of cellular heme, such

as nucleic acids. As aforementioned, the heme-binding capabilities of DNA and RNA G-quadruplexes have been studied for decades (52). Furthermore, these interactions have been shown to enhance oxygen transfer reactions (50-52), indicating these interactions could have physiological relevance. A recent study reported that treatment of human cells with a G-quadruplex stabilizer resulted in upregulation of proteins involved in the degradation of free heme, such as heme oxygenase 1 (HMOX1), suggesting that one of the potential physiological roles of G-quadruplexes is to sequester free cytotoxic heme (63).

1.4. Scope of Thesis

In this dissertation, I will describe our results on i) the identification of G-quadruplex sequences in human LSU and SSU ribosomal RNAs, ii) their *in vitro* and *in vivo* formation, and iii) the role of rRNA G-quadruplexes in regulation of heme bioavailability in human cells.

In Chapter 2, I will describe our published work in which we used a "canonical" G-quadruplex motif to identify the G-quadruplex regions in human rRNAs, most of which are located on surface-exposed regions of the LSU rRNA. By performing a wide range of biochemical assays, our results suggest that the regions derived from the LSU rRNA form highly stable G-quadruplexes *in vitro*. In addition, we also identified the cytosolic proteins that bind to one of these regions. By performing phylogenic analyses, we discovered that the G-quadruplex regions we found in human rRNA are conserved across the phylum Chordata. Finally, our *in vitro* data also suggest human rRNA G-quadruplexes form more readily in polysomes than in monomer ribosomes, indicating a potential involvement in

polysome linkage via inter-molecular G-quadruplexes. This was the first report on G-quadruplex formation on any ribosomal RNA.

In Chapter 3, I will present the published work that continued our first article. In this second study, we applied a more relaxed G-quadruplex motif to identify numerous additional G-quadruplex regions in human 28S rRNA (from 4 to 17 regions) and 3 G-quadruplex regions in human 18S rRNA. Our data suggest regions derived from the human 18S rRNA as well as intact 28S and 18S rRNAs extracted from human cells can form stable G-quadruplexes *in vitro*. Our phylogenic studies indicate that the SSU G-quadruplex regions are highly conserved in mammals. This was the first report on G-quadruplex formation on any small subunit rRNA.

In Chapter 4, I will describe our work on 1) rRNA G-quadruplex formation *in vivo* and 2) the involvement of these non-canonical ribosomal structures in regulation of heme bioavailability in human cells. By taking a chemical biology approach that integrates results from immunofluorescence, G-quadruplex ligands, heme affinity reagents, and a genetically encoded fluorescent heme sensor, we observed that human ribosomes can form G-quadruplexes *in vivo* that regulate heme bioavailability. Our immunofluorescence experiments indicated that the vast majority of extra-nuclear G-quadruplexes are associated with rRNA (~83%). Moreover, titrating human cells with a G-quadruplex ligand we observed a dose-dependent alteration in the ability of ribosomes to bind heme and in the bioavailable levels of cytosolic labile as measured by a genetically encoded fluorescent heme sensor. Overall, results in Chapter 4 suggest ribosomes are central hubs of heme metabolism.

Lastly, in Chapter 5 I will present the conclusions of this work and some future directions.

In short, the results presented in this thesis indicate human ribosomal RNAs can form G-quadruplexes *in vitro* and *in vivo* and suggest that one of their physiological functions is to regulate heme bioavailability in cells.

CHAPTER 2

G-QUADRUPLEXES IN HUMAN RIBOSOMAL RNA

This chapter is adapted from the previously published work: Mestre-Fos, S., Penev, P. I., Suttapitugsakul, S., Ito, C., Petrov, A. S., Wartell, R. M., Wu, R., Williams, L. D. "G-Quadruplexes in Human Ribosomal RNA", *J. Mol. Biol.* 431, pp 1940-1955 (2019). The author of this document contributed to this work by conceiving, designing and performing all experiments except the mass spectrometry and phylogenetic analyses, making all figures, making all tables, making all supplementary materials except the mass spectrometry analysis, and co-writing the manuscript.

2.1. Abstract

rRNA is the single most abundant polymer in most cells. Mammalian rRNAs are nearly twice as large as those of prokaryotes. Differences in rRNA size are due to expansion segments, which contain extended tentacles in metazoans. Here we show that the terminus of an rRNA tentacle of Homo sapiens contains 10 tandem G- tracts that form highly stable G-quadruplexes in vitro. We characterized rRNA of the H. sapiens large ribosomal subunit by computation, circular dichroism, UV melting, fluorescent probes, nuclease accessibility, electrophoretic mobility shifts, and blotting. We investigated Expansion Segment 7 (ES7), oligomers derived from ES7, intact 28S rRNA, 80S ribosomes, and polysomes. We used mass spectrometry to identify proteins that bind to rRNA G-quadruplexes in cell lysates.

These proteins include helicases (DDX3, CNBP, DDX21, DDX17) and heterogeneous nuclear ribonucleoproteins. Finally, by multiple sequence alignments, we observe that G-quadruplex-forming sequences are a general feature of LSU rRNA of Chordata but not, as far as we can tell, of other species. Chordata ribosomes present polymorphic tentacles with the potential to switch between inter- and intramolecular G-quadruplexes. To our knowledge, G-quadruplexes have not been reported previously in ribosomes.

2.2. Introduction

Cytosolic ribosomes are the most abundant assemblies in any cell, containing over 80% of cellular RNA (10). Ribosomes are built on a 'common core' (64) of rRNA with universal structure and function in all extant species. Common core rRNA is approximated by prokaryotic rRNA; around 90% of prokaryotic rRNA is contained in the common core.

The LSU of eukaryotic ribosomes contains additional rRNA in a secondary shell that surrounds the common core. The rRNA of the eukaryotic shell is composed of expansion segments (ESs) that attach to common core rRNA at a handful of specific sites (6, 65-67). ESs are the most variable rRNA structures over phylogeny. In *S. cerevisiae*, ESs are important in ribosome biogenesis (14) and chaperone association (68, 69). Beckman and coworkers recently described a 3D structure in which *S. cerevisiae* ESs associate with and localize N-terminal acetylases (70).

ESs of chordates contain "tentacles" (Figure 2.1) (71). These tentacles reach a zenith in primates and birds, extending for hundreds of Ångstroms from the ribosomal surface. Here we observe G-quadruplexes in rRNA tentacles of human ribosomes. Tandem G-tracts are found in the tentacles of expansion segments 7 and 27 of *H. sapiens* ribosomes ($ES7_{HS}$

and $ES27_{HS}$). $ES7_{HS}$ contains ten tandem G-tracts in *tentacle a* and four in *tentacle b*. $ES27_{HS}$ contains six tandem G-tracts in *tentacle a*, two sets of three in *tentacle b*, and four within a base helix (Table 2.1).

G-quadruplexes are favored by tandem G-tracts separated by short non-specific sequences. To investigate the possibility that G-tracts in rRNA tentacles form G-quadruplexes, we used computation, circular dichroism, fluorescent probes, thermal melting, nuclease accessibility, electrophoretic mobility shift assays (EMSA), dot blotting, Western blotting, and pull-down assays combined with stable isotope labeling of amino acids in cell culture (SILAC) and Mass Spectrometry. To investigate phylogenetic distribution of ribosomal G-quadruplexes we conducted Multiple Sequence Alignments (MSAs) and database analysis.

The results indicate that G-quadruplexes form in oligomers composed of sequences derived from $ES7_{HS}$. In addition, G-quadruplexes form in intact $ES7_{HS}$ and in purified human 28S rRNA. We present data supporting formation of G-quadruplexes in 80S ribosomes and in polysomes. SILAC experiments show that known G-quadruplex-binding proteins associate with the G-tracts of $ES7_{HS}$. MSAs indicate that G-quadruplex-forming sequences are found in ES7s of all chordates.

Table 2.1. G-quadruplex-forming regions within ES7 and ES27 rRNA

ES7		**Nucleotides**
tentacle a		
Homo sapiens	**GGGGGCGGG**CUCCGGC**GGG**UGC**GGGGGU****GGG**C**GGG**C**GGGG**CC**GGGGGU****GGGG**UCG GC**GGGGG**	587-648
Pan troglodytes	**GGGGGCGGG**CUCCGGC**GGG**UGC**GGGGGU****GGG**C**GGG**C**GGGG**CC**GGGGGU****GGGG**UCG GC**GGGGG**	583-644
Mus musculus	**GGG**C**GGGG**CC**GGGGGU****GGGG**UCGGC**GGGGG**	627-656
Gallus gallus	**GGGG**C**GGGG**C**GGG**CCCA**GGGGGGGG**C**GGG**C**GGG**CC**GGGG**	557-594
tentacle b		
Homo sapiens	**GGG**A**GGG**CGCGC**GGG**UC**GGGG**	829-849
Pan troglodytes	**GGG**A**GGG**CGCGC**GGG**UC**GGGG**	816-836
tentacle d		
Mus musculus	**GGG**C**GGG**CGU**GGGGGU****GGGGG**CC**GGG**	907-932
Gallus gallus	**GGGG**CGC**GGGGGC****GGGGGGGG**UC**GGG**	933-958
ES27		
tentacle a		
Homo sapiens	**GGGGG**AGCGCCGCGU**GGGGGC**GGCGGC**GGGGGG**AGAA**GGG**UC**GGGG**CGGCA**GGGG**	3095-3149
tentacle b		
Homo sapiens[a]	**GGGGGC****GGGG**AGCGGUC**GGG**CGGCGGCGGUCGGC**GGG**CGGC**GGGGC****GGGG**	3373-3422
Helix 63		
Homo sapiens	**GGG**CU**GGG**UCGGUC**GGG**CU**GGGG**	2896-2918

a) This sequence falls outside the $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$ motif.

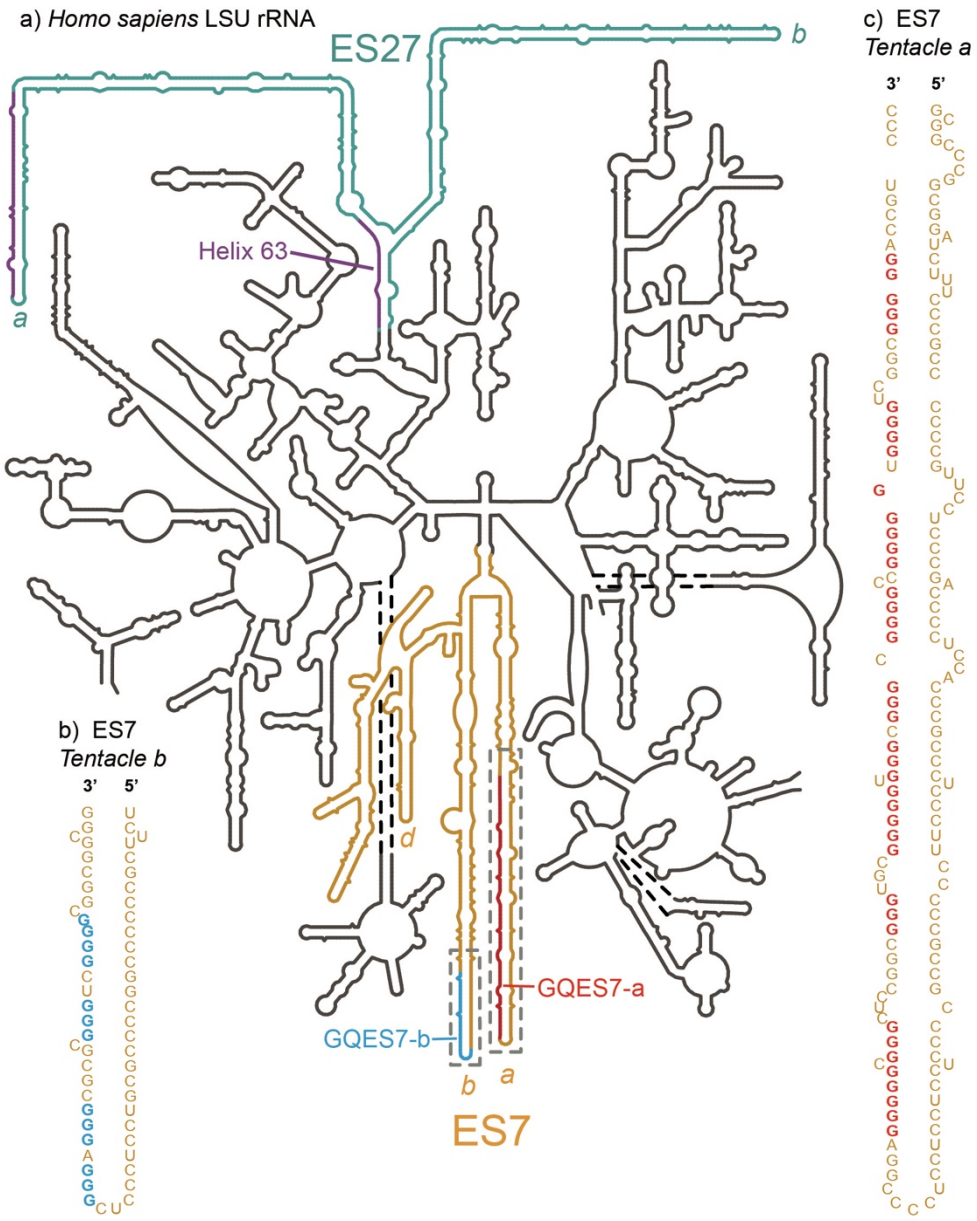


Figure 2.1. Model of the secondary structure of the LSU rRNA of *Homo sapiens*. G-quadruplex-forming regions (defined by $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$) are highlighted. a) Expansion segment $ES7_{HS}$ is orange. *Tentacles a, b* and *d* of $ES7_{HS}$ are indicated. G-quadruplex-forming regions of $ES7_{HS}$ are GQES7-a (red, in *tentacle a*) and GQES7-b (cyan, in *tentacle b*) and are boxed by dashed lines. Expansion segment $ES27_{HS}$ is green with purple G-tracts. Helix 63, at the base of $ES27_{HS}$, contains a G-quadruplex motif (purple). *Tentacles a* and *b* of $ES27_{HS}$ are indicated. b) An expanded view of GQES7-b indicates the nucleotide sequence. c) An expanded view of GQES7-a indicates the nucleotide sequence.

2.3. Results

2.3.1. ES7 and ES27 of the human LSU contain G-quadruplex-forming sequences.

The propensity of an RNA to form G-quadruplexes can be estimated from sequence - by lengths of guanine tracts and the lengths and compositions of loops regions. The program QGRS Mapper (72) provides "G-scores", which quantitate this propensity. We have identified ES7 and ES27 as the primary regions in the human LSU with sequences that appear to be capable of forming G-quadruplexes. The computational results suggest that G-quadruplexes can form near the termini of the longest rRNA tentacles of these two ESs. In tentacles *a* and *b* of $ES7_{HS}$, two regions, here named GQES7-a and GQES7-b (Figure 2.1), meet the G-quadruplex consensus ($G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$). The G-scores of GQES7-a and GQES7-b are in the range of well-established RNA G-quadruplexes.

The sequence 5' **GGGG**CC**GGGGG**U**GGGG**UCGGC**GGGG** 3' (nts 623-647, from within GQES7-a, Figure 2.1, Table 2.1) gives a G-score of 60. The sequence 5' **GGG**UGC**GGG**GGU**GGG**C**GGG** 3' (nts 603-621, also within GQES7-a) gives a G-score of 40. The sequence 5' **GGG**A**GGG**CGCGC**GGG**UC**GGGG** 3' (nts 829-849 within GQES7-b, Figure 2.1, Table 2.1) gives a G-score of 38. Differences in the number of tandem G-tracts (10 in GQES7-a and 4 in GQES7-b), the lengths of the G-tracts, and G-scores suggest more stable and more extensive G-quadruplex formation in GQES7-a than in GQES7-b. The G-tracts of GQES7-a are longer than those of GQES7-b and there are more of them. A greater propensity of GQES7-a over GQES7-b for G-quadruplex formation is seen in all experiments below.

As a positive control for both computation and experiment, we used the G-quadruplex from the 5'-UTR of the mRNA of the ADAM10 metalloprotease (73). This stable and well-characterized RNA G-quadruplex gives a G-score of 42. As negative controls, we used two mutant RNA oligomers (*mt*ES7-a and *mt*ES7-b) that are analogous to GQES7-a and GQES7-b in composition and length, with disrupted G-tracts (Table A.1). Neither gives a G-score. In several experiments we used yeast-tRNA^{phe} as an additional negative control.

We focused our experiments primarily on $ES7_{HS}$. However, the end of *tentacle a* of $ES27_{HS}$ contains the sequence 5' **GGGG**AGAA**GGG**UC**GGGG**CGGCA**GGG** 3' (nts 3124-3148, *tentacle a*), which gives a G-score of 40 (Figure 2.1, Table 2.1). $ES27_{HS}$ also contains a G-quadruplex-forming region within Helix 63, near the junction of *tentacles a and b*. Based on the high G-scores and our experimental observation of G-quadruplexes within $ES7_{HS}$, we expect G-quadruplexes to form in $ES27_{HS}$.

2.3.2. Circular dichroism.

CD spectra of GQES7-a and GQES7-b indicate G-quadruplex formation, with the expected dependence on type of counterion. CD is used widely to study RNA and DNA G-quadruplexes (74-78). CD spectra of GQES7-a and GQES7-b (Figure 2.2b) show the characteristic peak at 260 nm and trough at 240 nm. It is known that G-quadruplex formation is promoted by K^+ and is inhibited by Li^+ or Na^+ (79). The intensities of the 260 nm peaks of both GQES7-a and GQES7-b are attenuated when the monovalent cation is switched from K^+ to Li^+. GQES7-a gives a more intense CD signal than GQES7-b under all conditions.

2.3.3. *ThT fluorescence.*

ThT is known to yield intense fluorescence at 487 nm upon association with G-quadruplexes (80, 81). ThT fluorescence results here suggest formation of G-quadruplexes in GQES7-a, GQES7-b and intact $ES7_{HS}$ (Figure 2.2c). Intact $ES7_{HS}$ and GQES7-a give more intense ThT fluorescence signals than the positive control (ADAM10). The GQES7-b signal is less than that of ADAM10 but is significantly greater than the controls. ThT fluorescence of GQES7-a and GQES7-b is attenuated when the monovalent counterion is switched from K^+ to Li^+. (Figure 2.2d). Consistent with results of QGRS Mapper and CD spectroscopy, the ThT-induced fluorescence signal for GQES7-b is less than that of GQES7-a under all conditions.

The formation of G-quadruplexes by GQES7-a, GQES7-b and intact $ES7_{HS}$ is supported by competition assays with pyridostatin (PDS) (Figures 2.2e and 2.2f). PDS is a G-quadruplex stabilizer and a ThT competitor with a greater affinity than ThT for G-quadruplexes (82). As expected if G-quadruplexes form in these rRNAs, PDS displaces ThT. In this series of experiments *mt*ES7-a, *mt*ES7-b and tRNA were used as negative controls, giving signals near background.

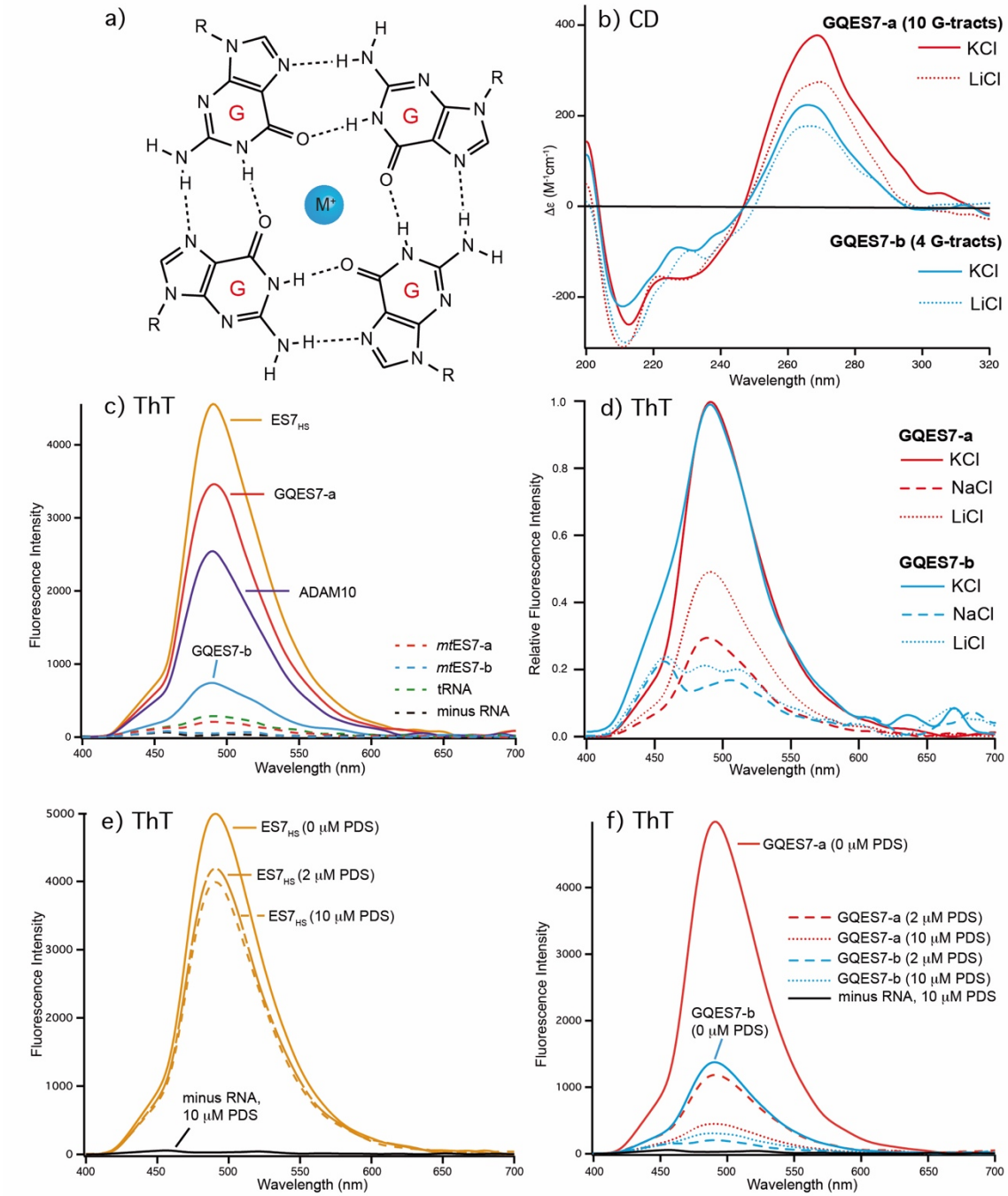
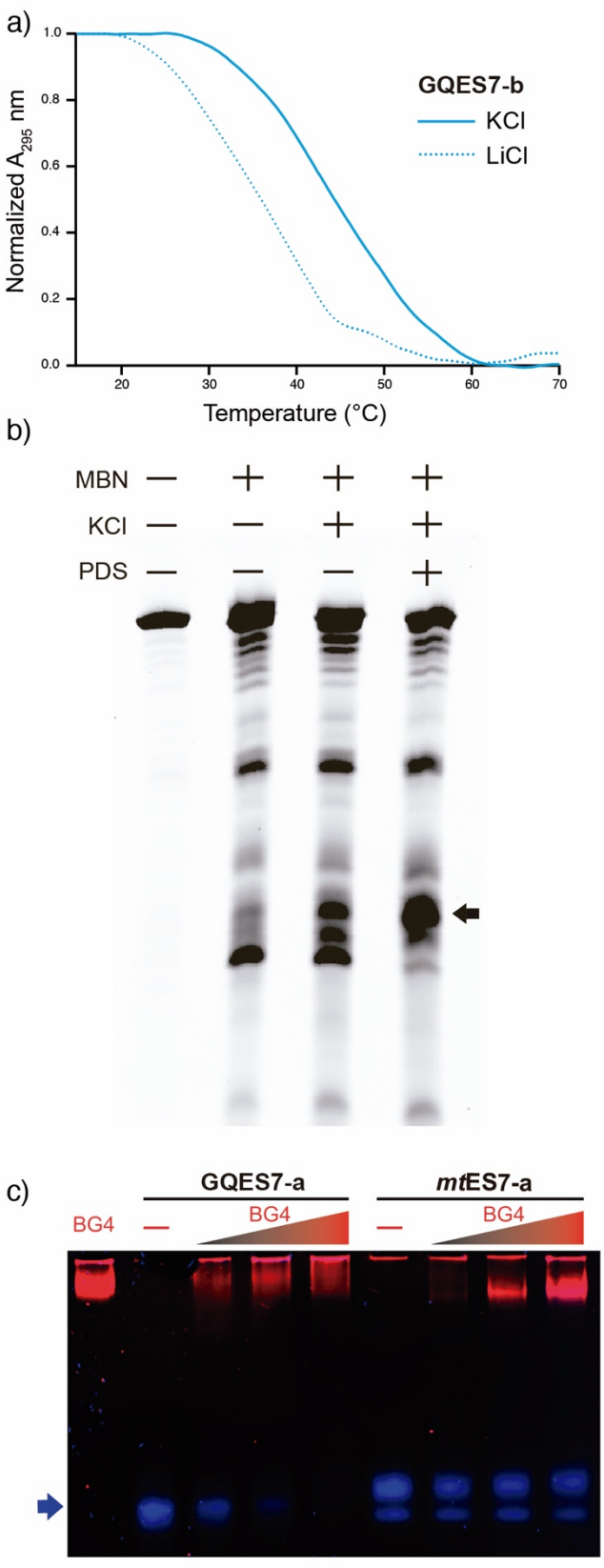


Figure 2.2. Formation of G-quadruplexes by rRNA fragments GQES7-a, GQES7-b, and $ES7_{HS}$. a) G-quadruplexes preferentially coordinate K^+. b) CD spectra of GQES7-a and GQES7-b in the presence of either K^+ or Li^+. c) Fluorescence emission at 487 nm of the G-quadruplex probe ThT in the presence of $ES7_{HS}$, GQES7-a, GQES7-b, or positive control ADAM10. Negative controls (dashed) are tRNA, *mt*ES7-a, *mt*ES7-b and minus RNA. d) ThT fluorescence emission of GQES7-a and GQES7-b in the presence of various monovalent cations. Intensities of GQES7-a and GQES7-b are normalized in the presence of K^+ to highlight cation-induced differences. e) PDS competes with ThT in association with $ES7_{HS}$. f) PDS competes with ThT in association with GQES7-a and GQES7-b.



2.3.4. *UV thermal melting.*

The melting of G-quadruplexes is distinguishable from melting of other RNA secondary structures. Melting of G-quadruplexes, but not other RNA structures, is accompanied by *hypochromicity* at 295 nm (83). Melting of G-quadruplexes, but not other RNA structures, shows an acute dependence on type of monovalent counterion. The Tm's of G-quadruplex melting are expected to be greater in K^+ than in Li^+.

Figure 2.3. a) UV thermal melting profile of GQES7-b at 295 nm. Before melting RNA was annealed in the presence of either 100 mM KCl or 100 mM LiCl. b) $ES7_{HS}$ cleavage by mung bean nuclease. ES7 was annealed with or without KCl and with or without PDS. The black arrow indicates cleaved rRNA. c) EMSA of the BG4 antibody with GQES7-a and its non-G-quadruplex-forming mutant *mt*ES7-a, visualized on a native gel. GQES7-a and *mt*ES7-a RNAs were loaded at a constant strand concentration with increasing concentrations of BG4 antibody. The RNA (arrow) is blue and the protein is red.

The UV melting profile of GQES7-b demonstrates the characteristic hypochromic shift at 295 mm and the expected salt dependence (Figure 2.3a). The T_m of the melting transition is increased by around 9 °C when the counterion is switched from Li^+ ($T_m \cong 36$ °C) to K^+ (~45 °C). We were unable to observe a melting transition under any conditions for GQES7-a. Even under low concentrations of Li^+ the Tm of GQES7-a appears to be greater than experimentally accessible temperatures. The observed differences between melting behaviors of GQES7-a and GQES7-b are consistent with the greater G-quadruplex propensity of GQES7-a than of GQES7-b observed by GQRS Mapper, CD spectroscopy and ThT fluorescence.

2.3.5. Mung bean nuclease cleavage.

In the conventional secondary model of *H. sapiens* LSU rRNA (Figure 2.1), G-quadruplex-forming sequences are represented as doubled-stranded and are paired with C-rich strands. If these tentacles form G-quadruplexes, the C-rich strands would be dissociated, presumably as single strands. As an additional test for G-quadruplexes in intact $ES7_{HS}$ rRNA, we examined cleavage by mung bean nuclease (MBN, Figure 2.3b). MBN preferentially cleaves single-stranded RNA or DNA and would cleave $ES7_{HS}$ rRNA more rapidly if G-quadruplexes form than if they do not. MBN cleavage has been used previously to test for G-quadruplexes in DNA (84). The results here show that MBN cleaves $ES7_{HS}$ most rapidly under G-quadruplex-stabilizing conditions (Figure 2.3b). Addition of K^+ increases the extent of cleavage (at constant time). Addition of PDS to K^+ further increases the extent of cleavage. The simplest interpretation of the MBN results is that $ES7_{HS}$ exists as a mixture of duplex and G-quadruplex forms and that the equilibrium is shifted by the type of counterion and by G-quadruplex stabilizers. In a negative control,

extent of MBN hydrolysis of tRNA did not increase upon addition of K^+ and/or PDS (Figure A.2).

2.3.6. Antibody binding.

BG4 is an antibody developed by Balasubramanian and coworkers (24, 25) that binds to a variety of G-quadruplex types but not to other nucleic acids such as RNA hairpins, single-stranded or double-stranded DNA. Here, to test for G-quadruplex formation in GQES7-a, an EMSA was performed with BG4 (Figure 2.3c). BG4 was also used for dot blotting experiments with GQES7-a, GQES7-b, and intact $ES7_{HS}$ (Figure 2.4). We observe binding of BG4 to GQES7-a, GQES7-b, and intact $ES7_{HS}$ (Figure 2.4a-b). Consistent with the results above, BG4 binds more tightly to GQES7-a than to GQES7-b.

The experiments presented above are consistent with *in vitro* formation of G-quadruplexes by $ES7_{HS}$ and by oligomers derived from $ES7_{HS}$. Below we investigate whether intact *H. sapiens* 28S rRNA can form G-quadruplexes when protein-free or when assembled in ribosomes. The 28S rRNA was extracted from HEK293T cells and dot blotting was performed with BG4 (Figure 2.4d). The results suggest that 28S rRNA forms G-quadruplexes.

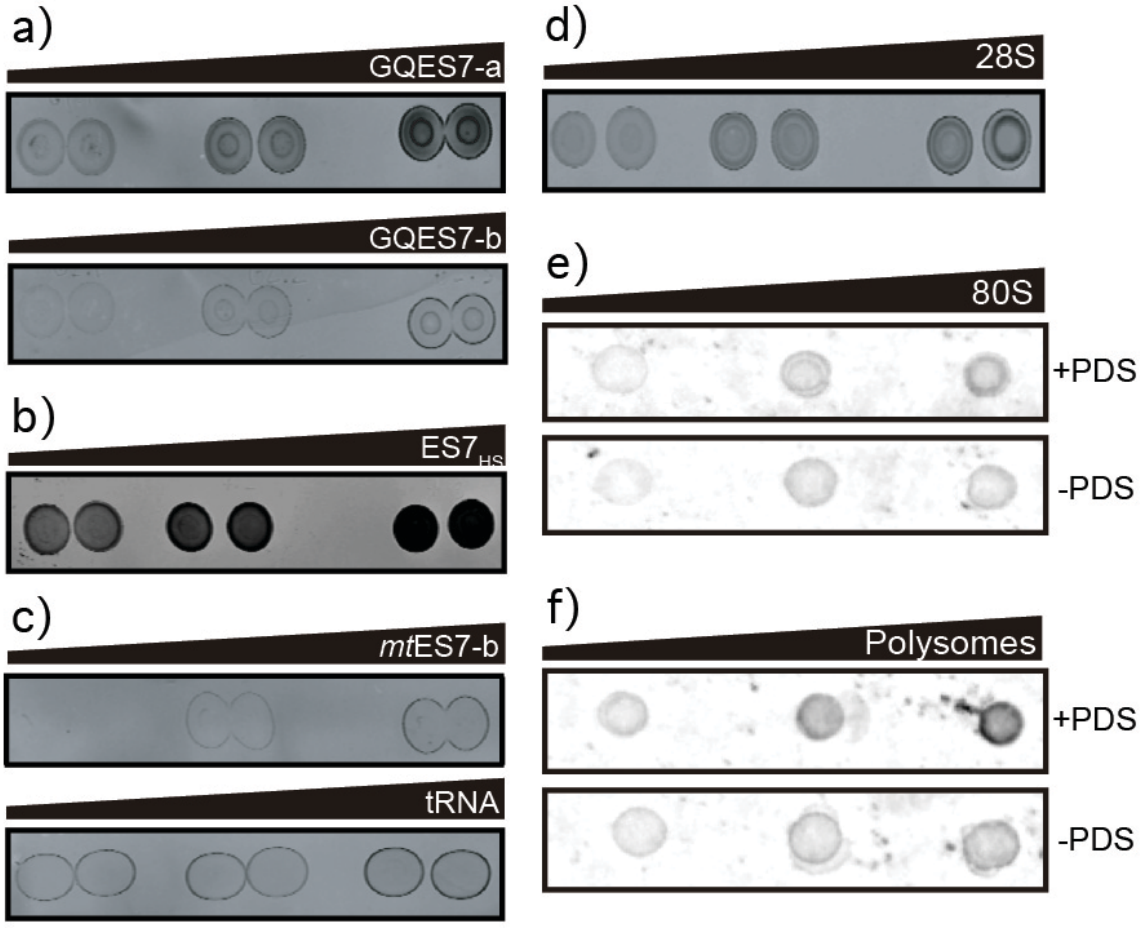


Figure 2.4. Dot blots performed with the BG4 antibody on a) GQES7-a and GQES7-b, b) intact $ES7_{HS}$, c) the negative controls *mt*ES7-b and tRNA, d) the 28S rRNA extracted from HEK293T cells and on e) human 80S ribosomes and f) polysomes purified from HEK293 cells. All samples were incubated in the presence of 50 mM KCl and ribosomes and polysomes were further analyzed with or without 10 μM PDS, which stabilizes G-quadruplexes. Samples were loaded onto the membrane in increasing amounts from left to right.

To determine if 28S rRNA from *H. sapiens* forms G-quadruplexes when assembled in intact ribosomes, dot blotting was also performed with purified 80S human ribosomes and with polysomes (Figure 2.4e and 2.4f). The results show that the BG4 antibody binds preferentially to intact human ribosomes and polysomes in a concentration-dependent manner. PDS enhances binding of the antibody, as expected for G-quadruplex formation.

The observation of more extensive binding of the antibody to polysomes than to monomer ribosomes suggests formation of intermolecular G-quadruplexes in polysomes (i.e., that G-quadruplexes link tentacles of adjacent ribosomes).

2.3.7. G-quadruplex sequences in ribosomes throughout Chordata.

Focusing specifically on translation, we have developed the SEREB Database (64), which contains fully curated and cross-validated sequences of rRNAs from all major phyla, yet samples the tree of life in a sparse, efficient and unbiased manner. Here we extended the SEREB database, increasing the number of chordate species from 10 to 17, for a fine-grained analysis of ES7.

2.3.7.1.*G-quadruplex-forming sequences in chordate ES7s*.

Our MSA confirms that the lengths of rRNA tentacles of eukaryotes are variable, reaching maxima in species such as *G. gallus* and *H. sapiens* (Figure 2.5). Aligned sequences of relevant segments of ES7s of various eukaryotes demonstrate G-quadruplex-forming sequences in chordates, indicated by the motif $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$ (number of G tracts (n) >3). The motif is observed near the termini of ES7 tentacles in all warm-blooded chordates, although the exact locations and specific sequences are variable. The maximum number of tandem G-tracts in *tentacle a* of ES7 is ten in human and chimpanzee and eight in rat and chicken. Fish, reptiles and amphibians appear to lack the G-tract motif in ES7 tentacles.

2.3.7.2.*G-quadruplex-forming sequences in chordate ESs other than ES7.*

The extended SEREB Sequence Database suggests that G-quadruplex-forming sequences are universal to chordates (Table A.2). Several chordate species present G-quadruplex-forming sequences in tentacles other than ES7 (Table A.2). These G-quadruplex-forming sequences are not shown in Figure 2.5. In addition, repeated G-tracts outside of the motif can form G-quadruplexes (85, 86). Therefore, additional G-quadruplexes cannot be excluded in these ribosomes. It is possible that G-tracts with $n<4$ form intermolecular G-quadruplexes with other tentacles or with other ribosomes as in polysomes.

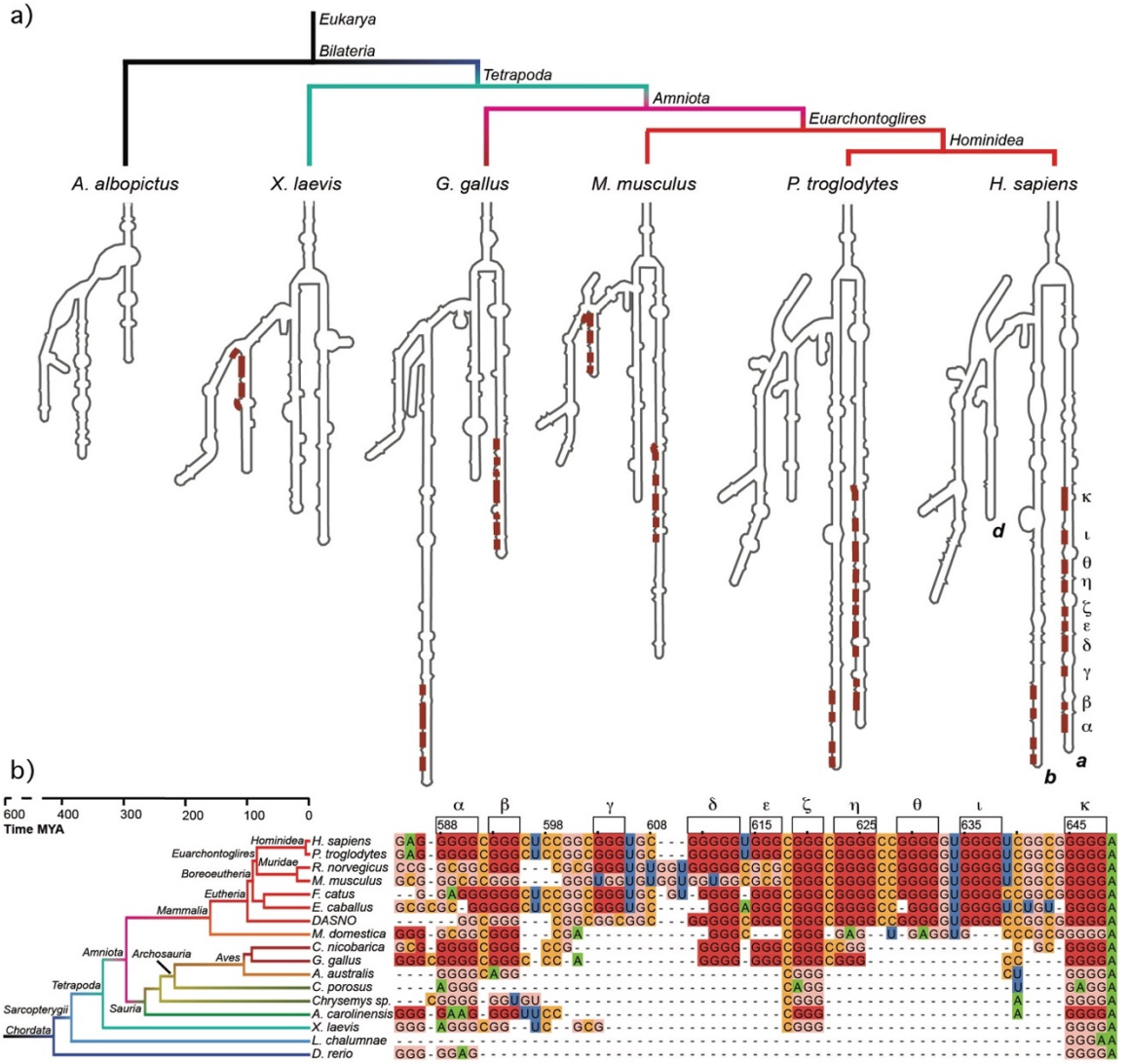


Figure 2.5. G-tracts are observed in ES7 tentacles. a) Conventional secondary structural models of ES7 from various eukaryotes. G-tracts within the $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$ motif are highlighted in red. b) Sequence alignment of ES7 *tentacle a* showing G-quadruplex-forming sequences are common in chordates. Individual G-tracts in both panels are labeled with Greek symbols. Nucleotides are colored by type. G's within G-tracts are dark red. Other G's are pink. All nucleotides are numbered in accordance with *H. sapiens* 28S rRNA. Sizes of eukaryotic ES7 secondary structures are not to scale. Complete species nomenclature is provided in Table A.3.

2.3.7.3.Absence of G-quadruplex sequences in non-chordate rRNAs.

To determine the phylogenic distribution of G-quadruplexes in LSU rRNA, we inspected highly curated sequences of 20 non-chordate eukaryotes from the SEREB database (64). Thus far we can find no evidence of G-quadruplex-forming sequences in ribosomes of non-chordate eukaryotes.

2.3.8. RNA remodeling proteins bind to rRNA G-quadruplex sequences.

The localization of G-quadruplex-forming sequences to ribosomal tentacles suggests the possibility of interaction with non-ribosomal proteins. To identify the proteins that bind to rRNA G-quadruplexes, we performed pull-down experiments and used stable isotope labeling with amino acids in cell culture (SILAC) for protein quantification. We focused on GQES7-a, the longest and most stable G-quadruplex-forming region in human rRNA (Figure 2.6). GQES7-a rRNA was linked on the 3' end to biotin (GQES7-a-Biotin) and associated proteins in human cell lysates were pulled down and analyzed by mass spectrometry. The biotinylation of GQES7-a does not disrupt the G-quadruplexes (Figure A.3). Known G-quadruplex-binding proteins were pulled down by this assay including CNBP, YBOX1, hnRP F, hnRP H, DDX21, DDX17, DDX3X (87-93). A significant number of helicases were identified (DDX3X, CNBP, DDX21, DDX17), all of which have been reported to unfold G-quadruplexes. In addition, a significant number of heterogeneous nuclear ribonucleoproteins (hnRNPs) were shown in this experiment to bind to GQES7-a, including hnRNP G-T/RMXL2, hnRNP M, hnRNP G/RBMX, hnRNP H2, hnRNP H, hnRNP F, hnRNP H3, and FUS. hnRNPs are a family of RNA-binding proteins with functions including pre-mRNA processing and transport of mRNAs to ribosomes (94).

Several of these proteins have been previously identified as ribosome-binding proteins (95).

To support results of the pull-down experiments, Western blotting was performed with four of the proteins obtained in the pull-down experiments (Figure 2.6d). We assayed a DEAD-box RNA helicase (DDX3X), a heterogeneous nuclear ribonucleoprotein (hnRNP H), the RNA-binding protein FUS, and a pre-mRNA polyadenylation stimulator (FIP1). hnRNP H and DDX3X have been previously identified as G-quadruplex-binding proteins. All four proteins bind to GQES7-a in the Western blot, suggesting we have tapped an uncharacterized pool of G-quadruplex-binding proteins.

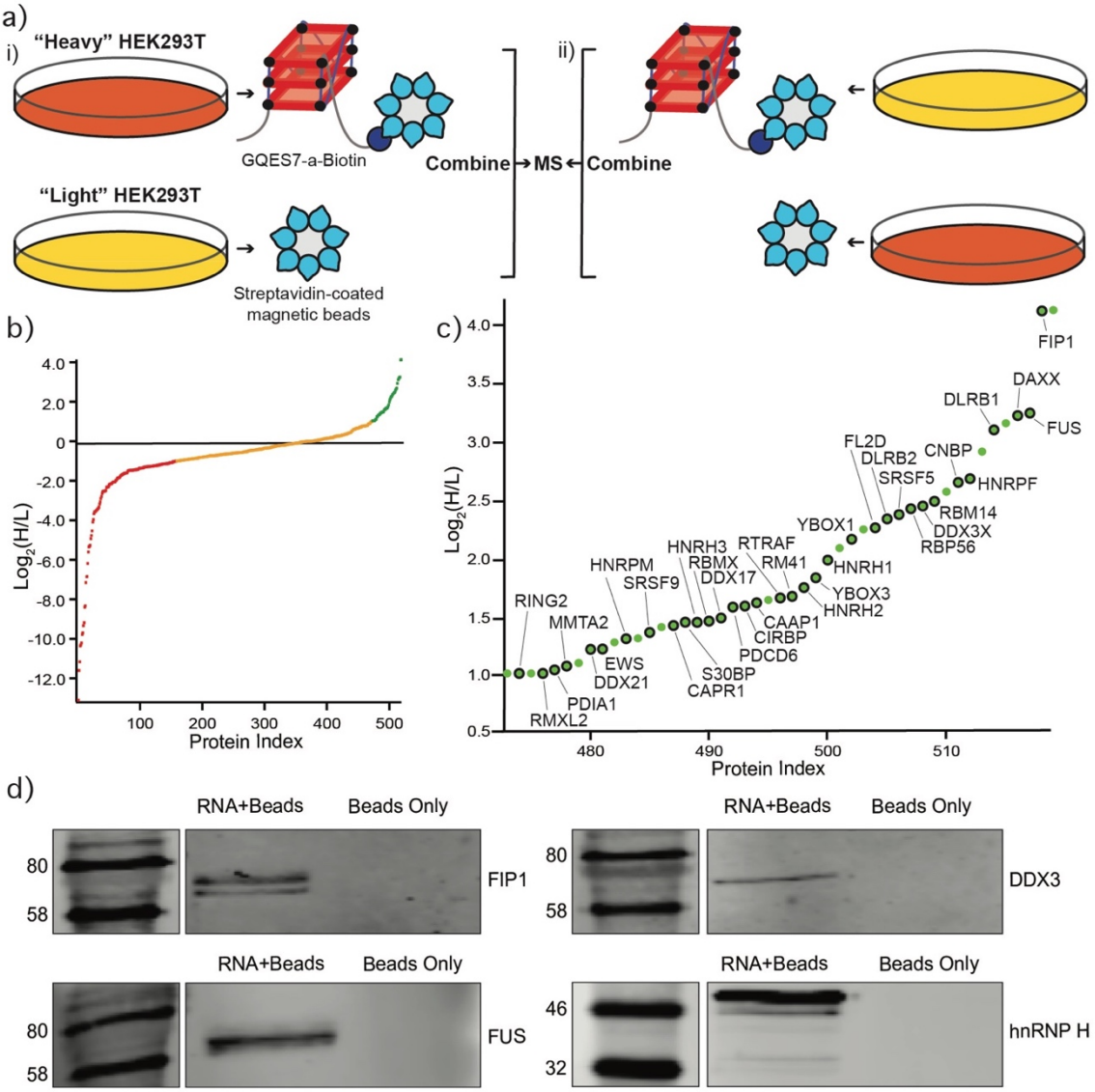


Figure 2.6. Identification of GQES7-a-binding proteins. a) Scheme of the SILAC experiment. "RNA+beads" samples were combined in HEK293T grown in heavy media (panel i). The "Beads Only" control sample was combined in HEK293T grown in light media. To verify the proteins identified by this method, the experiment was performed using reverse labeling (panel ii). b) Scatter plot representing fold enrichment of the proteins binding to GQES7-a in "Heavy" HEK293T. Color representation indicates specific proteins that bound more tightly to GQES7-a than to the beads (green), to the beads than to GQES7-a (red) or bound to the beads and GQES7-a to a similar extent (orange). c) A close-up view of the green region of the scatter plot represented in panel b. Dots with a black contour are used to indicate proteins that appeared in the green region of the two replicate experiments described in panel a. d) Western blotting analyses of the eluted proteins from the RNA pull-down of HEK293T. All four blotted proteins (FIP1, FUS, DDX3 and hnRNP H) eluted from the GQES7-a sample (RNA+Beads) but not from the control (Beads Only), confirming the SILAC results.

2.4. Discussion

The results presented here suggest that G-quadruplexes are far more profuse than previously conceived. rRNA is the most abundant macromolecule in most cells. Human LSU rRNA and rRNA of other chordates contain sequences with strong propensity to form G-quadruplexes (Figure 2.1). We have identified ten tandem G-tracts in *tentacle a* of human ES7 rRNA and four in *tentacle b*. These tandem G-tracts form stable G-quadruplexes under a variety of conditions *in vitro*. Computation, ThT fluorescence, CD spectroscopy, UV melting, EMSAs, nuclease digestion and blotting with a G-quadruplex antibody provide a consistent picture of the propensities of various regions of 28S rRNA to form G-quadruplexes.

G-quadruplex-forming sequences have been shown previously to cluster within regulatory mRNA regions such as 5′ and 3′ untranslated regions (96) and within the first intron (97). The extent to which such G-quadruplexes form *in vivo* remains uncertain. Bartel and coworkers suggest that mRNA G-quadruplexes are globally unfolded by unwinding factors in eukaryotic cells (31). By contrast, Wong and Monchaud support a model in which G-quadruplexes continuously form and unfold *in vivo* (27). However, rRNAs were explicitly excluded from both of these investigations; the extent of G-quadruplex formation *in vivo* remains an open question.

In our view, the following merit a ribosome centric reinvestigation of G-quadruplex structure and function *in vivo:*

i. inherent flexibility and polymorphism of rRNA tentacles,

ii. their ability to extend hundreds of Ångströms from the ribosomal surface,

iii. the large number of tandem G-tracts on some rRNA tentacles,

iv. high stability of tentacle G-quadruplexes *in vitro*, and

v. extreme concentrations of rRNA on the rough ER and in polysomes.

Our results suggest that nature's most complex organisms have evolved long rRNA tentacles with unexpected structural polymorphism, including the ability to form G-quadruplexes (Figure 2.7).

G-quadruplex-forming rRNA sequences appear to be a universal feature of tentacles of chordate ribosomes. We have inferred the locations of tandem rRNA G-tracts in various species by Multiple Sequence Alignments. The specific sequences and exact locations of the G-quadruplexes on tentacles are variable across phylogeny. We searched the SEREB database and thus far could find no evidence of G-quadruplex-forming sequences outside of the Chordata phylum. The SEREB database is specifically designed for analysis of rRNA, and includes species from all major phyla, and samples the tree of life in a sparse, efficient and accurate manner (64). It contains complete and highly curated rRNA sequences.

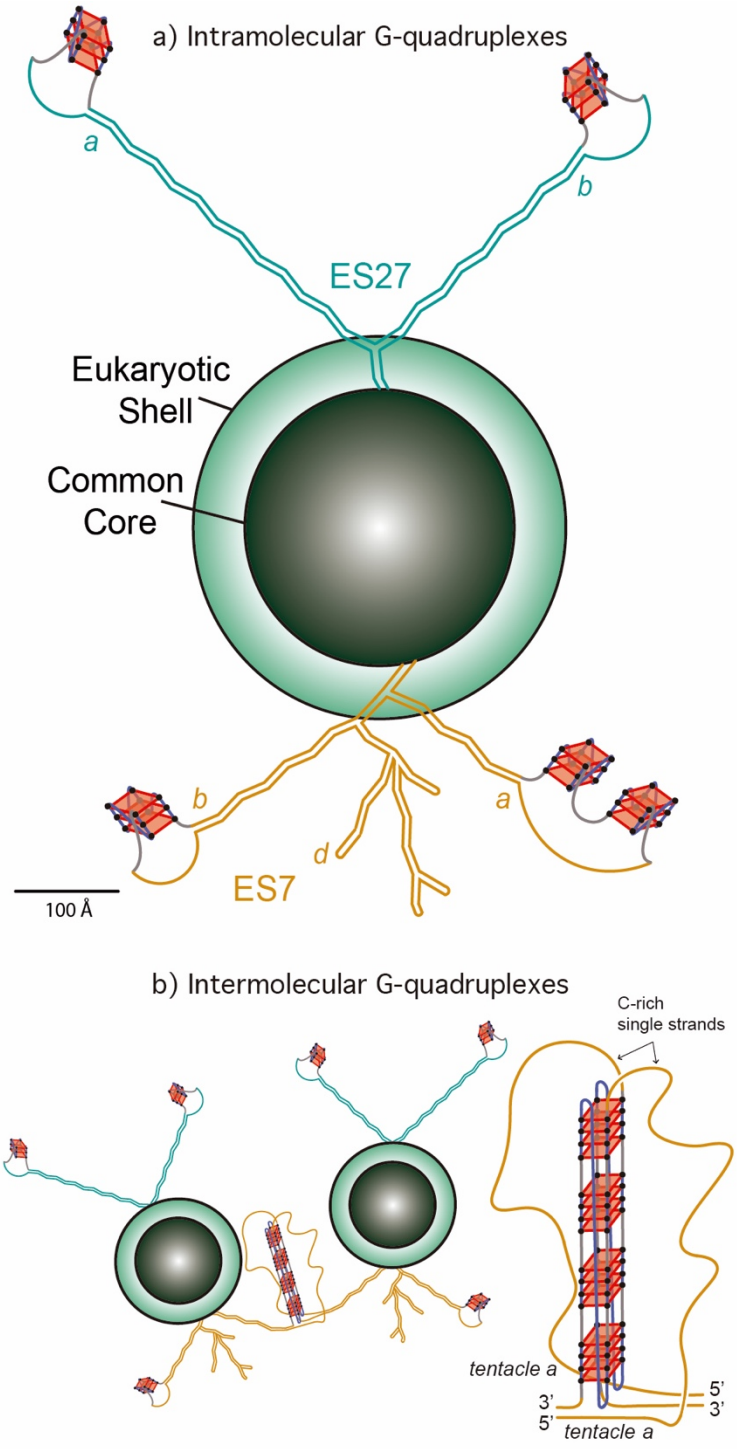


Figure 2.7. a) Schematic representation of the common core, the eukaryotic shell and the tentacles of the LSU of the *Homo sapiens* ribosome. G-quadruplexes are indicated on ES7 and ES27. The lengths of $ES7_{HS}$ (orange) and $ES27_{HS}$ (green) tentacles are roughly scaled to the size of the common core. The G-quadruplexes represented in *tentacle b* of $ES27_{HS}$ do not fall within the $G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}N_{1\text{-}7}G_{\geq 3}$ motif and are speculative. The G-quadruplex region found in Helix 63 of $ES27_{HS}$ is not indicated here. b) Schematic representation of interactions between ribosomes via intermolecular G-quadruplexes. G-tracts on ES7 tentacle *a* from different ribosomes contribute to the formation of G-quadruplexes (an expanded view is presented on the right).

The preferential localization of rRNA G-quadruplexes near the termini of rRNA tentacles suggests these regions are loci for association of specific cytosolic proteins or for assembly of ribosomes. Here we identified multiple human RNA helicases and other RNA

remodeling proteins that bind to rRNA G-quadruplexes. (Figure 2.6) These proteins could be participants in G-quadruplex regulation on ribosomes.

Our observation here that polysomes appear to form more extensive G-quadruplexes than monomer ribosomes suggests a role for inter-ribosomal G-quadruplexes in closely associated ribosomes (Figure 2.7b). Our work points to the possibility that, inside cells, ribosomes present polymorphic tentacles that can switch between unimolecular and multimolecular G-quadruplexes and duplex forms. In this model, surfaces of ribosomes contain multivalent docking sites for G-quadruplex-specific proteins and for nucleic acid assemblies, including in polyribosomes. It has been shown that G-quadruplexes can form phase separated RNA gels (49). It is conceivable that densely packed ribosomes on the rough ER and in polysomes are surrounded by phase separated G-quadruplexes gels, composed of rRNA tentacles.

G-quadruplex-forming sequences have been described in genes encoding rRNA, where they are proposed to influence transcription (98) and bind to the nucleolar protein nucleophosmin (99). These studies have focused on the external and internal transcribed regions (ETS and ITS) and are not part of the assembled ribosome. Moore and co-workers reported tetramerization of an oligomer containing a single G-tract, derived from *E. coli* 5S rRNA (100). The results here highlight potential ribosomal functionality associated with the large rRNA tentacles of mammals and birds. The conservation of rRNA G-quadruplex-forming sequences throughout Chordata, albeit at various locations in rRNA tentacles, suggests significant functions.

2.5. Methods

$ES7_{HS}$ spans nucleotides 436 to 1311 of the *H. sapiens* LSU rRNA and contains sequences that we call GQES7-a (nts 583-652) and GQES7-b (nts 825-853). RNAs corresponding to $ES7_{HS}$, GQES7-a and GQES7-b were synthesized *in vitro* by transcription (HiScribe™ T7 High Yield RNA Synthesis Kit, New England Biolabs). For cation-dependent experiments, precautions were used to remove contaminating cations that stabilize G-quadruplexes. RNAs were ethanol precipitated from 800 mM LiCl, 10 mM Tris-HCl, pH 7.5. The RNA pellets were resuspended in 1 mM LiCl, 10 mM Tris-HCl, pH 7.5 were dialyzed extensively against the same buffer using Slide-a-Lyzer dialysis cassettes (MWCO 3,500, Pierce) at 4°C. *mt*ES7-a and *mt*ES7-b were ordered as RNA oligomers. Baker's yeast tRNAs were purchased from Roche. RNA purity was monitored by 8 M Urea 5% acrylamide gel in TBE buffer. Complete sequences of $ES7_{HS}$, GQES7-a, GQES7-b, *mt*ES7-a and *mt*ES7-b are contained in Table A.1.

2.5.1. HEK293T 28S rRNA extraction and purification.

HEK293T cells were grown to 60% confluency after which total RNA was extracted with TRI Reagent® (Sigma-Aldrich). 28S rRNA was extracted from an agarose gel by running the rRNA into wells in the center of the gel, where the rRNA was extracted with a pipette. The rRNA was precipitated in 5 M ammonium acetate-acetic Acid, pH 7.5 with excess ethanol. 28S rRNA purity was monitored on 1% agarose gels (Figure A.1).

2.5.2. Thioflavin T (ThT) fluorescence.

RNAs were prepared at a final concentration of 1 μM (strand) and annealed in 150 mM KCl, NaCl or LiCl, 10 mM Tris-HCl, pH 7.5, 2 □M ThT by cooling from 90°C to 25°C at 1°/min. RNAs were incubated at 4°C for 10 min and were loaded onto a Corning® 384 Well Flat Clear Bottom Microplate. Fluorescence from 300-700 nm, exciting at 440 nm were acquired on a BioTek Synergy™ H4 Hybrid plate reader. When appropriate, pyridostatin (PDS) was added to the desired concentration after the RNA was annealed.

2.5.3. Circular dichroism.

RNA at 1 μM (strand) in 150 mM KCl or LiCl and 10 mM Tris-HCl (pH 7.5) was annealed as described above. CD spectra were acquired at 20 °C on a Jasco J-810 spectropolarimeter using 1 mm cuvettes. Data from 200-320 nm was acquired at a rate of 100 nm/min with 1 sec response, a bandwidth of 5 nm, and averaged over three measurements. The buffer spectrum was subtracted. Smoothing was performed with Igor Pro. The observed ellipticity (θ, mdeg) was normalized (101) using the expression $\Delta\varepsilon = \theta/(32{,}980 \times c \times l)$, where c is the molar strand concentration of the RNA and l is the path length of the cuvette in centimeters.

2.5.4. UV thermal melting.

Absorbance measurements were collected at 295 nm using a Varian Cary-1E UV spectrophotometer. RNA samples (800 μL, final OD_{260} of 0.50 units) in 10 mM Tris-HCl, pH 7.5 and 100 mM KCl or LiCl were annealed as described above and added to 1 cm path-length quartz cuvettes. Samples were then heated from 15°C to 90°C and cooled at the same rate at 0.5°C/min. Data was recorded every 0.5°C.

2.5.5. EMSAs.

The anti-G-quadruplex BG4 antibody was purchased from Absolute Antibody (Catalog #: Ab00174-1.1). GQES7-a (3 μM) rRNA or the negative control *mt*ES7-a RNA were annealed in 20 mM Hepes-Tris, pH 7.5, 50 mM KCl. GQES7-a rRNA or *mt*ES7-a RNA were combined with various concentrations of BG4 at a final RNA concentration of 1 μM RNA (strand). RNA-protein mixtures were incubated at room temperature for 20 min in 50 mM KCl. RNA-protein interactions were analyzed by 5% native-PAGE. Gels were visualized following a dual fluorescent dye protocol (102) with a Azure imager c400 (Azure Biosystems).

2.5.6. rRNA - BG4 antibody dot blotting.

RNAs were annealed in the presence of 50 mM KCl and were diluted 1x, 2x and 4x. GQES7-a, GQES7-b, *mt*ES7-b, tRNA: 3.2 μM, 1.6 μM, 0.8 μM. $ES7_{HS}$: 1.4 μM, 0.7 μM, 0.35 μM. 28S rRNA: 55 nM, 27.5 nM, 13.7 nM. RNAs were loaded onto nitrocellulose membranes and dried at room temperature for 30 min. The membranes were blocked for 1 h at room temperature. BG4 antibody was added (1:2,000 dilution) and incubated with gentle rocking for sixty min at room temperature. The membrane was washed for ten min twice with 1X TBST and incubated for sixty min with an appropriate fluorescent secondary antibody anti-mouse (1:10,000 dilution) (Biotium, #20065-1). The membrane was washed for ten min twice with 1X TBST and was imaged on a Li-Cor Odyssey Blot Imager. Intact 80S ribosomes and polysomes were purified from HEK293, which were incubated 5 min in 10 μg/mL cycloheximide at 37°C. Lysis buffer (10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.5, 1% Triton X-100, 1% sodium

deoxycholate, 0.2 U/mL DNase I, RNase inhibitor, 1mM dithiothreitol, 10 μg/mL cycloheximide) was used to scrap the cells. Nuclei and cell debris were removed by centrifugation and the supernatant was transferred to a 15-50% sucrose gradient containing 100 mM NaCl, 10 mM $MgCl_2$, 30 mM Tris-HCl, pH 7.5 and centrifuged by ultracentrifugation. Purified 80S ribosomes and polysomes were then incubated at room temperature for 15 min in the presence of 50 mM KCl with or without 10 □M PDS. Ribosomes or polysomes were added iteratively in 30-min intervals to the same site on a nitrocellulose membrane (0.9 μg, 2.7 μg, 4.5 μg). The membrane was then treated as described above. BG4 was added to a final dilution of 1:1,000 and the secondary antibody was added to a final dilution of 1:5,000.

2.5.7. Mung bean nuclease (MBN) probing.

$ES7_{HS}$ and tRNA were prepared at 100 ng/□L and annealed in the presence/absence of 100 mM KCl, 15 mM Tris-HCl (pH 7.5) by cooling from 90°C to 25°C, at 1°/min. PDS was added to the annealed RNA to a final concentration of 2 μM. One unit of MBN was added per μg of RNA and samples were incubated at 30°C for 30 min. SDS was added to a final concentration of 0.01% to denature the nuclease and RNA was purified by ethanol precipitation. The extent of RNA cleavage was determined on an 8 M urea 5% acrylamide (19:1 acrylamide/bisacrylamide) gel stained with ethidium bromide.

2.5.8. ES7 secondary structures.

*S*econdary structures of human and *D. melanogaster* ES7 were obtained from RiboVision (103)**.** Nucleotides of G-quadruplex regions in *P. troglodytes*, *M. musculus* and

G. gallus (Table 2.1) were numbered as in Bernier (64), subtracting the nucleotides from the 5.8S rRNA.

2.5.9. Phylogeny and Multiple Sequence Alignments.

The SEREB MSA (64) was used as a seed to align additional eukaryotic ES7 sequences to increase the density of eukaryotic species in the MSA. The 28S rRNA sequences in the SEREB MSA were used to search (104) the NCBI databases (105) for LSU rRNA sequences. The SEREB database has sequences from 10 chordate species; seven additional chordate species from 7 new orders were added to the ES7 *tentacle a* MSA (Figure 2.7, Table A.3). Sequences without intact ES7 *tentacle a* were excluded. Sequences with partial 28S rRNA were marked as partial. Sequences inferred from genomic scaffolds were marked as predicted (Table A.3). The extended database was queried for G-quadruplex-forming sequences.

Sequences were incorporated into the SEREB-seeded MSA using MAFFT (106) and adjusted manually using BioEdit (107). Manual adjustments incorporated information from available secondary structures. In some cases, the positions of G-tracts in sequences with large gaps relative to *H. sapiens* are not fully determined, as they can be aligned equally well with flanking G-tracts in the MSA. Alignment visualization was done with Jalview (108). The phylogenetic tree and the timeline of clade development were inferred from TimeTree (109).

Analysis of the entire LSU was performed on SEREB sequences, which are highly curated and always complete. This procedure ensured that negative results indicate absence of G-quadruplex-forming sequences from intact rRNAs rather than absence from rRNA

fragments that lack the appropriate regions. G-quadruplex-forming sequences are not detected in any of the 20 non-chordate members of the SEREB database.

2.5.10. SILAC.

HEK293T cells were cultured in SILAC media - “heavy” or “light” Dulbecco's Miodified Eagle Media (DMEM) (Thermo Scientific) supplemented with 10% dialyzed fetal bovine serum (FBS) (Corning) and 1% penicillin-streptomycin solution (Sigma) in a humidified incubator at 37 ˚C with 5% carbon dioxide. The heavy media contained 0.798 mM L-lysine ($^{13}C_6$ and $^{15}N_2$, Cambridge Isotope Laboratories) and 0.398 mM L-arginine ($^{13}C_6$, Cambridge Isotope Laboratories). The light media had the same concentrations of normal lysine and arginine (Sigma). Media were supplemented with 0.2 mg/mL proline (Sigma) to prevent arginine-to-proline conversion. Heavy and light cells were grown for at least six generations. Once the confluency reached 80%, cells were harvested by scraping, washed twice with ice-cold PBS (Sigma), lysed in a buffer containing 10 mM HEPES pH=7.4, 200 mM potassium chloride, 1% Triton X-100, 10 mM magnesium chloride (all from Sigma) and 1 pill/10 mL cOmplete ULTRA tablet protease inhibitor (Roche), and incubated on an end-over-end shaker at 4 ºC for 1 hour. Lysates were centrifuged at 25,830 g at 4 ºC for 10 minutes, and the supernatants were collected and kept on ice.

Ten μg of GQES7-a-Biotin RNA was annealed as described above in the presence of 10 mM Tris-HCl, pH 7.5, and 100 mM KCl. Twenty μL of magnetic streptavidin-coated beads (GE Healthcare) were washed with the lysis buffer (10 mM HEPES, pH 7.4, 200 mM KCl, 1% Triton X-100, 10 mM $MgCl_2$, protease inhibitors). Annealed RNA was then added to the washed beads and incubated at 4ºC for 30 min with gentle shaking. For control

experiments, no RNA was added. SILAC cell lysates were incubated with 0.5 mg *E. coli* tRNA per 1 mg protein at 4°C for 30 min with gentle shaking. "RNA+Beads" and control "Beads Only" samples were transferred into the SILAC cell lysates: "RNA+Beads" were added to the Heavy cell lysate and "Beads Only" was added to "Light" HEK293T cell lysate. As a replicate, "RNA+Beads" was added "Light" cell lysate and "Beads Only" was added to "Heavy" cell lysate. 200 U/mL of RNasin was added and the lysates were incubated at 4 °C for 2 hrs with gentle shaking. Samples were centrifuged, the supernatant was discarded, and the pelleted beads were washed with lysis buffer with increasing KCl concentrations (0.4 M, 0.8 M, 1.6 M). After the three washes, 100 µL of the elution buffer (100 mM Tris-HCl, pH 7.4, 1% SDS, 100 mM DTT) was added to one of the two samples and then combined with the beads of the corresponding sample. "RNA+Beads" in "Heavy" lysates were combined with "Beads Only" in "Light" lysates and "RNA+Beads" in "Light" lysates were combined with "Beads Only" in "Heavy" lysates. The combined samples were boiled and then briefly centrifuged. Beads were discarded and samples were analyzed with an online LC-MS system.

2.5.11. Mass spectrometry.

Eluted proteins were diluted 10 times with 50 mM HEPES pH=7.4 and were alkylated with 28 mM iodoacetamide (Sigma) for 30 minutes at room temperature in the dark. Proteins were precipitated by methanol-chloroform, and the pellets were resuspended in digestion buffer containing 50 mM HEPES pH=8.8, 1.6 M urea, and 5% acetonitrile (ACN) (all from Sigma). After digestion with sequencing-grade modified trypsin (Promega) at 37 °C for 16 hours, reactions were quenched with 1% trifluoroacetic acid (TFA, Fisher Scientific) and purified with StageTip. Peptides were dissolved in 10 µL 5%

ACN and 4% FA solution, and 1 µL was loaded to a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) with a microcapillary column packed in-house with C18 beads (Magic C18AQ, 3 µm, 200 Å, 75 µmx16 cm). A 110-minute gradient of 3-22% ACN containing 0.125% FA was used. The peptides were detected with an LTQ Orbitrap Elite Hybrid Mass Spectrometer (Thermo Fisher Scientific) controlled by Xcalibur software (version 3.0.63). MS/MS analysis was performed with a data-dependent Top20 method. For each cycle, a full MS scan in the Orbitrap with the automatic gain control (AGC) target of 10^6 and the resolution of 60,000 at 400 m/z was followed by up to 20 MS/MS scans in the ion trap for the most intense ions. Selected ions were excluded from further sequencing for 90 seconds. Ions with singly or unassigned charge were not sequenced. Maximum ion accumulation times were 1,000 ms for each full MS scan and 50 ms for each MS/MS scans. The spectra were searched against a human protein database downloaded from UniProt using the SEQUEST algorithm (version 28) (110). The following parameters were used: 20 ppm precursor mass tolerance; 1.0 Da fragment ion mass tolerance; trypsin digestion; maximum of 3 missed cleavages; differential modifications for methionine oxidation (+15.9949 Da), heavy lysine (+8.0142 Da), and heavy arginine (+6.0201 Da); fixed modification for cysteine carbamidomethylation (+57.0215 Da). The false discovery rates (FDR) were evaluated and controlled by the target-decoy method. Linear discriminant analysis (LDA) was used to filtered the peptides to <1% FDR based on parameters such as XCorr, ΔC_n, and precursor mass error. An additional filter was used to control the protein FDR to <1%. For SILAC quantification, the S/N ratios of both heavy and light peptides must be greater than 3. Otherwise, one of the two versions of the peptides (heavy or light) must have the S/N ratio greater than 10. Other peptides that did not pass these criteria were

removed. The final protein ratio was calculated from the median value of the peptides from each parent protein. The raw files are publicly accessible at http://www.peptideatlas.org/PASS/PASS01260, Username: PASS01260, Password: TL3854zn.

2.5.12. Western Blotting.

Samples were electrophoresed on 12% SDS-PAGE gels and transferred to a nitrocellulose membrane overnight. Membranes were blocked for 1 hour at room temperature with gentle shaking and then incubated for another hour with primary antibodies: 1:200 dilution of FIP1 (mouse monoclonal, sc-398392), DDX3 (mouse monoclonal, sc-81247), FUS (mouse monoclonal, sc-47711), or hnRNP H (mouse monoclonal, sc-32310). Antibodies were obtained from Santa Cruz Biotechnology. Membranes were washed three times with 1X TBST and secondary antibody CF680 goat anti-mouse IgG (H+L) (Biotium, 20065) was added (1:5,000 dilution). Membranes were washed three times with 1X TBST and imaged on a Li-Cor Odyssey Blot Imager.

CHAPTER 3

PROFUSION OF G-QUADRUPLEXES ON BOTH SUBUNITS OF METAZOAN RIBOSOMES

This chapter is adapted from the previously published work: Mestre-Fos, S., Penev, P. I., Richards, J. C., Dean, W. L., Gray, R. D., Chaires, J. B., Williams, L. D. "Profusion of G-quadruplexes on both Subunits of Metazoan Ribosomes", *PLOS ONE.* 14, (2019). The author of this document contributed to this work by conceiving, designing and performing all experiments except the phylogenetic analysis and analytical ultracentrifugation and CD melting experiments, making all figures except Figure 3.1a and 3.3b, making all tables, making all supplementary materials except Figures B.1, B.3, and B.4, and co-writing the manuscript.

3.1. Abstract

Mammalian and bird ribosomes are nearly twice the mass of prokaryotic ribosomes in part because of their extraordinarily long rRNA tentacles. Human rRNA tentacles are not fully observable in current three-dimensional structures and their conformations remain to be fully resolved. In previous work we identified sequences that favor G-quadruplexes *in silico* and *in vitro* in rRNA tentacles of the human large ribosomal subunit. We demonstrated by experiment that these sequences form G-quadruplexes in vitro. Here, using a more recent motif definition, we report additional G-quadruplex sequences on

surfaces of both subunits of the human ribosome. The revised sequence definition reveals expansive arrays of potential G-quadruplex sequences on LSU tentacles. In addition, we demonstrate by a variety of experimental methods that fragments of the small subunit rRNA form G-quadruplexes in vitro. Prior to this report rRNA sequences that form G-quadruplexes were confined to the large ribosomal subunit. Our combined results indicate that the surface of the assembled human ribosome contains numerous sequences capable of forming G-quadruplexes on both ribosomal subunits. The data suggest conversion between duplexes and G-quadruplexes in response to association with proteins, ions, or other RNAs. In some systems it seems likely that the integrated population of RNA G-quadruplexes may be dominated by rRNA, which is the most abundant cellular RNA.

3.2. Introduction

rRNA expansion segments (ES's) decorate the surfaces of eukaryotic ribosomes. Ribosomal ES's of complex eukaryotes, especially in birds and mammals, contain long rRNA tentacles that appear to extend for 100's of Å from the ribosomal surface. We previously reported that rRNA tentacles of chordates can form G-quadruplexes (111). rRNA tentacles of *Homo sapiens* contain multiple sequences that form G-quadruplexes of unusually high stability *in vitro*.

In our previous work, rRNAs were scrutinized for G-quadruplex sequences using the classic 3 x 4 motif of at least three contiguous guanines and a short spacer, repeated four times [$(G_{\geq 3}N_{1-7})_{n\geq 4}$] (111). A G-quadruplex is composed of four guanine columns surrounding a central cavity that sequesters monovalent cations ($K^+ > Na^+ > Li^+$) (32, 112). The guanine columns are linked by Hoogsteen hydrogen bonds between co-planar

guanines. The 3 x 4 sequence criteria identified four potential G-quadruplex forming regions in ES7 and ES27 of the rRNA of the large ribosomal subunit (LSU) of humans. rRNA of the human small ribosomal subunit (SSU) rRNA appeared to lack G-quadruplex sequences.

However, the RNA sequence space of G-quadruplexes has recently been re-evaluated: both shortened and bulged G-tracts are now seen to form stable G-quadruplexes (18-21, 34, 113). Here, we extended our rRNA sequence search using a revised motif of four repeats of *two or mor*e adjacent guanines connected by a spacer (2 x 4, $(G_{\geq 2}N_{1-7})_n$). Using the 2 x 4 criterion, we identify numerous additional G-quadruplex forming sequences on the human LSU rRNA, and for the first time, detect potential G-quadruplex forming sequences in SSU rRNAs (Figure 3.1). We experimentally confirm formation *in vitro* of G-quadruplexes by rRNA fragments derived from SSU rRNA as well as by native human SSU rRNA.

The revised sequence definition reveals expansive arrays of potential G-quadruplex sequences on LSU tentacles of expansion segments 7 and 27 (ES7 and ES27). The G-quadruplex forming regions of ES7 are increased from 10 to 20 G-tracts (tentacle *a*, regions 1 and 2) and from 4 to 23 G-tracts (tentacle *b*, regions 3 and 4) (Figure 3.1B). For ES27, the increases are from 6 to 25 G-tracts (tentacle *a*, regions 10 and 11) and from 0 to 32 G-tracts (tentacle *b*, regions 12, 13 and 14). rRNA sequences on the surfaces of the *H. sapiens* LSU and SSU that meet the 2 x 4 criteria are shown in Figure 3.1B and listed in Tables 3.1 (SSU) and 3.2 (LSU). G-quadruplex sequences in the SSU are located mainly near the termini of expansion segments (Figure 3.1B); es's are considerably smaller than ES's (lower case es indicates SSU and uppercase ES indicates LSU).

Here, we experimentally characterize SSU 2 x 4 G-quadruplex sequences of es3 and es12. Our data suggest that the 2 x 4 sequence of es3 forms a G-quadruplex *in vitro* while that of es12 forms a hairpin. The es12 sequence converts from hairpin to G-quadruplex in the presence of a G-quadruplex stabilizer or at elevated temperature. We also observe that 2 x 4 G-quadruplex sequences on the surface of the SSU are conserved over phylogeny of warm-blooded animals.

The combined results suggest that G-quadruplexes can be formed by multiple surface-exposed sequences on both the LSU and SSU of *H. sapiens*. The results are consistent with a model in which human ribosomal surfaces are structurally polymorphic, with complex liquid-liquid phase behavior, mediated in part by G-quadruplex formation. In metazoans, the integrated population of RNA G-quadruplexes (the RNA G-quadruplexome) may be dominated by rRNA, which is more abundant than other cellular RNAs (10, 114) and presents more expansive arrays of G-tracts.

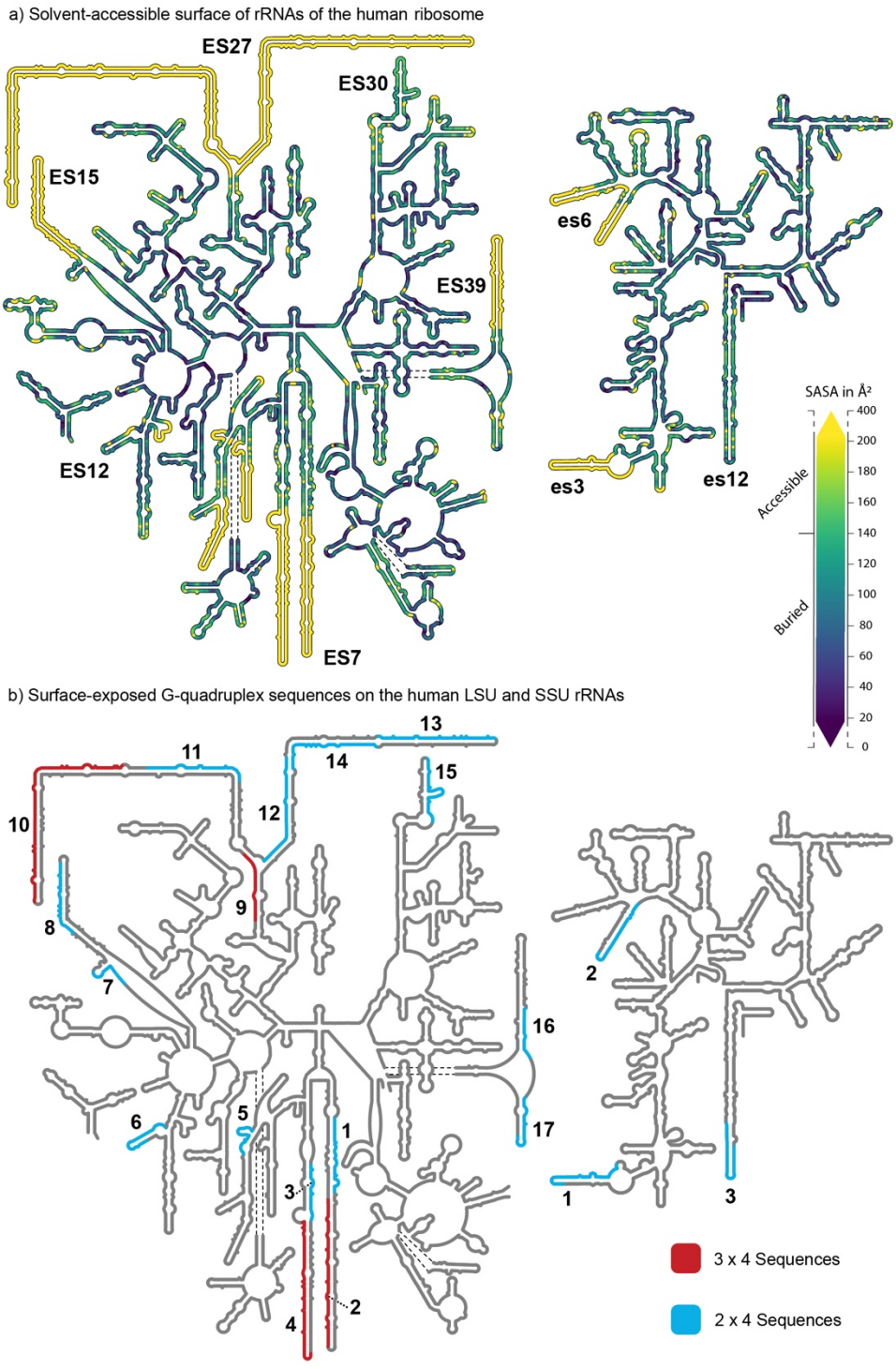


Figure 3.1. Secondary structures of the LSU and SSU rRNAs of *Homo sapiens*. (a) Nucleotide level solvent-Accessible Surface Area of the human LSU and SSU determined from intact ribosomes (PDB ID: 4UG0 (115)). (b) G-quadruplex forming sequences identified by the 3 x 4 motif are highlighted in red and those identified with the 2 x 4 motif are highlighted in blue. Only G-quadruplex forming sequences located on the ribosomal surface are shown. These images were generated with RiboVision (103).

3.3. Results

3.3.1. Computation

3.3.1.1. Solvent accessible surface of the ribosome.

Here we presume that rRNA G-quadruplexes are more likely to form on the surface regions of the ribosome and not in the interior. The interior of the ribosome is engaged in intense inter- and intra-molecular interactions that constrain the structure and prevent conversion between different conformational states. To quantitate and visualize surface exposed rRNA, we calculated the Solvent-Accessible Surface Area (SASA) for all ribosomal nucleotides of the intact *H. sapiens* ribosome, and then mapped the data onto the two-dimensional structure of the rRNA (Figure 3.1A). The SASA varies between 0 $Å^2$ (fully buried) to 400$Å^2$ (fully exposed) per nucleotide. The color gradient shown in Figure 3.1A covers the range between 20$Å^2$ and 200$Å^2$ since the great majority of nucleotides fall in this range (Figure B.1). These ranges are consistent with previous definitions of exposed and buried nucleotides (116). Nucleotides with SASA below 20$Å^2$ are purple, nucleotides with SASA above 200$Å^2$ are yellow. Nucleotides with SASA above 140$Å^2$ are considered as solvent accessible. Multiple contiguous nucleotides with SASA above 140$Å^2$ indicate rRNA that is on the ribosomal surface. Surface rRNA indicated by SASA was confirmed by inspection of the three-dimensional ribosomal structure. Nucleotides that are not resolved in the structure but are known to be on the surface are also yellow. These regions of the rRNA are of particular interest because conformational heterogeneity, possibly conversion between duplex and G-quadruplex forms, may be the ultimate source of the

diffuse and smeared electron density during structure determination. Overall, the results confirm that rRNA expansion segments are in solvent-accessible regions of the ribosome.

3.3.1.2. Identification of G-quadruplex sequences.

The 2 x 4 sequence definition significantly extends the repertoire of putative rRNA G-quadruplexes in the human rRNA. All 3 x 4 sequences are located on solvent-accessible ribosomal regions whereas several of the 2 x 4 sequences are buried in the interior of the ribosome. The interior 2 x 4 sequences would appear to be locked in fixed non-G-quadruplex structures and were not investigated further. Canonical secondary structures of the human LSU and SSU rRNAs illustrating surface-exposed 2 x 4 and 3 x 4 sequences are shown in Figure 3.1B. Both surface-exposed and buried 2 x 4 and 3 x 4 sequences are shown in Figure B.2. Several of the 3 x 4 sequences in the human LSU have already been characterized (111). Here, we concentrate on the 2 x 4 sequences of the human SSU rRNA, whose G-quadruplex forming capabilities have not been reported to our knowledge.

We identified three 2 x 4 sequences on the surface of the human SSU rRNA. These sequences are located in expansion segments es3, es6, and es12. The propensity of each of these 2 x 4 sequences to form G-quadruplexes was estimated with the program QGRS Mapper, which outputs G-scores (117). The greatest SSU G-score (21) corresponds to the G-tracts of es3. Four 2 x 4 sequences are located on the buried interior of the ribosome. These interior 2 x 4 sequences are found in helices h33, h34, and in the junctions of helices h11 and h12 as well as helices h17 and h18 (Figure B.2).

G-tracts are ‘polarized’ in all LSU tentacles and in both es3 and es6 on the SSU. Polarized G-tracts are confined to one strand of the hairpin form of the tentacle with

complementary C-rich sequences on the opposing strand (Figure 3.1B). In general, polarization is an indicator of proximal rather than dispersed G-tracts. Proximal G-tracts are more likely to form stable G-quadruplexes that dispersed G-tracts.

3.3.2. Experiment.

We produced RNA oligomers (GQes3 and GQes12, Figure 3.2A and Table B.1) that contain the 2 x 4 sequences found in tentacles of es3 and es12. We also investigated intact 18S rRNA (SSU rRNA) extracted from human cells. As negative controls, we produced mutants *mut*es3 and *mut*es12 with the same nucleotide composition as GQes3 and GQes12, respectively, but with disrupted G-tracts (G-scores: 0, Table B.1).

The G-tracts of es12 are distributed between both strands of the hairpin form of the tentacle (Figure 3.2A). We anticipate that in isolated oligonucleotides, polarized G-tracts are more likely to form G-quadruplexes than non-polarized G-tracts because for non-polarized G-tracts the hairpin form competes with G-quadruplex formation. Here, we experimentally investigate the 2 x 4 sequences of a polarized SSU tentacle (es3) and a non-polarized SSU tentacle (es12). The combined data, described below, suggest that GQes3 forms G-quadruplexes while GQes12 forms a hairpin at low temperature in the absence of G-quadruplex stabilizers. At elevated temperatures and in the presence of G-quadruplex stabilizers GQes12 converts from hairpin to G-quadruplex.

Table 3.1. Sequences and G-scores of 2 x 4 G-quadruplex regions located on the ribosomal surface of the *Homo sapiens* 18S rRNA (SSU).

Name	Region	SSU rRNA Sequence (5' to 3')	Highest G-score
GQes3	es3	**GG**CCCC**GG**CC**GGGGGG**C**GGG**CGCC**GGCGG**CUUU**GG***	21
GQes12	es12	**GGGG**UC**GG**CCCAC**GG**CCCU**GGCGG**	20
	es6	**GG**AGC**GGGC****GGG**C**GG**UCCGCCGCGA**GG**	20

*GQes3 rRNA oligomer does not contain the last 9 nucleotides of the es3 sequence shown in Table 1

Table 3.2. Sequences and G-scores of 2 x 4 and 3 x 4 G-quadruplex regions located on the ribosomal surface of the *Homo sapiens* 28S rRNA (LSU).

#	Region	LSU rRNA Sequence (5' to 3')	Highest G-score
1	ES7-a	**GGCGGCGGG**UCC**GG**CCGUGUC**GGCGGC**CCC**GGCGG**	20
2	ES7-a	**GGGGGCGGG**CUCC**GGCGGG**UGC**GGGGGU****GGGCGGGC****GGGG**CC**GGGGGU****GGGG**UC**GGC** **GGGGG**	60
3	ES7-b	**GGCGGGGAAGGUGG**CUC**GGGGGG**	19
4†	ES7-b	**GGGAGGG**CGCGC**GGG**UC**GGGGC**GGCGGCGGCGGCGGCGGU**GGC**GGCGGCGGCGGC**GG** C**GGCGGG**	38
5	ES7-g	**GGGCCCGGGGGAGG**UUCUCUC**GGGG**	19
6	ES12	**GG**CUCGCU**GG**CGU**GG**AGCC**GGGC**GU**GG**	20
7	ES15	**GGAC****GGGAGC****GGCGGGGGC****GG**	21
8	ES15	**GGAGGGCGGCGGCGGCGGCGGCGGCGGGGG**UGU**GGGG**	21
9	H63	**GGGCUGGGUCGGUCGGGCUGGGG**	38
10†	ES27-a	**GGGGG**AGCGCCGCGU**GGGGGC****GGC****GGGGGGG**AGAA**GGG**UC**GGGGC****GGCAGGGGC**C **GGCGGCGG**CCCGCCGC**GGGG**CCCC**GGCGGCGGGGG**CAC**GG**	40
11	ES27-a	**GGGGGGCCCGGGC**ACCC**GGGGGGC****GGCGGCGGCGGC**GACUCU**GG**	37
12*	ES27-b	**GGCGGGC**GUCGC**GG**CCGCCCCC**GGGG**AGCCC**GGCGGGC**GCC**GG**	20
13	ES27-b	**GGGGGCGGGG**AGC**GG**UC**GGGCGGCGGCGG**UC**GGCGGGCGGC**GGGGCGGGGCGG**	21
14	ES27-b	**GG**CGCGC**GGCGGCGGCGGCGG**CA**GGCGGC****GGAGGGG**CCGC**GGGCCGG**	21
15	ES30	**GGGGCCC****GGGGC****GGGG**UCCGCC**GG**CCCUGC**GGG**CCGCC**GG**	21
16	ES39	**GGG**ACC**GGGG**UCC**GG**UGC**GG**	20
17	ES39	**GGGAAAC****GGGGC**GC**GGCCGG**AGA**GGCGG**	20

*This region contains a loop longer than 7 nucleotides
† These regions meet both 3 x 4 and 2 x 4 criteria.

3.3.2.1. Circular dichroism (CD) spectroscopy.

CD has been used extensively for inferring whether RNAs or DNAs form G-quadruplexes. RNA can form both parallel and antiparallel G-quadruplex topologies (44, 45). The CD spectra of GQes3 and GQes12, with positive peaks at 260 nm and troughs at around 240 nm (Figure 3.2B), are consistent with RNA G-quadruplexes (118). However, the spectra are not definitive because A-form DNA and G-quadruplexes present similar

CD spectra (119). The CD spectra of negative controls *mut*es3 and *mut*es12 are characteristic of A-C rich single-stranded RNA (120).

Thermal denaturation of GQes3 (T_m > 95 ºC) or GQes12 ($T_{m,1}$ =50 ºC, $T_{m,2}$> 90ºC) was monitored by CD (Figure 3.2C) to reveal that both form exceptionally stable structures; neither is fully melted at the highest obtainable temperatures. Since neither completely unfolds, a detailed thermodynamic analysis is not possible.

When monitored at 265 nm the melting of GQes3 appears simple and is typical of G-quadruplex melting (Figure 3.2C and Figure B.3)**.** However, the melting of GQes12 is complex, with a positive inflection near 50ºC, followed by a separate denaturation transition with T_m > 90 ºC. This biphasic behavior suggests that GQes12 may form one structure (perhaps a duplex hairpin, see below) at low temperature. Upon melting of the first structure, the strand appears to fold into a second structure (a G-quadruplex) that denatures with further increase in temperature. A probable hairpin structure for GQes12 is shown in Figure B.4 (calculated using the mFold RNA server).

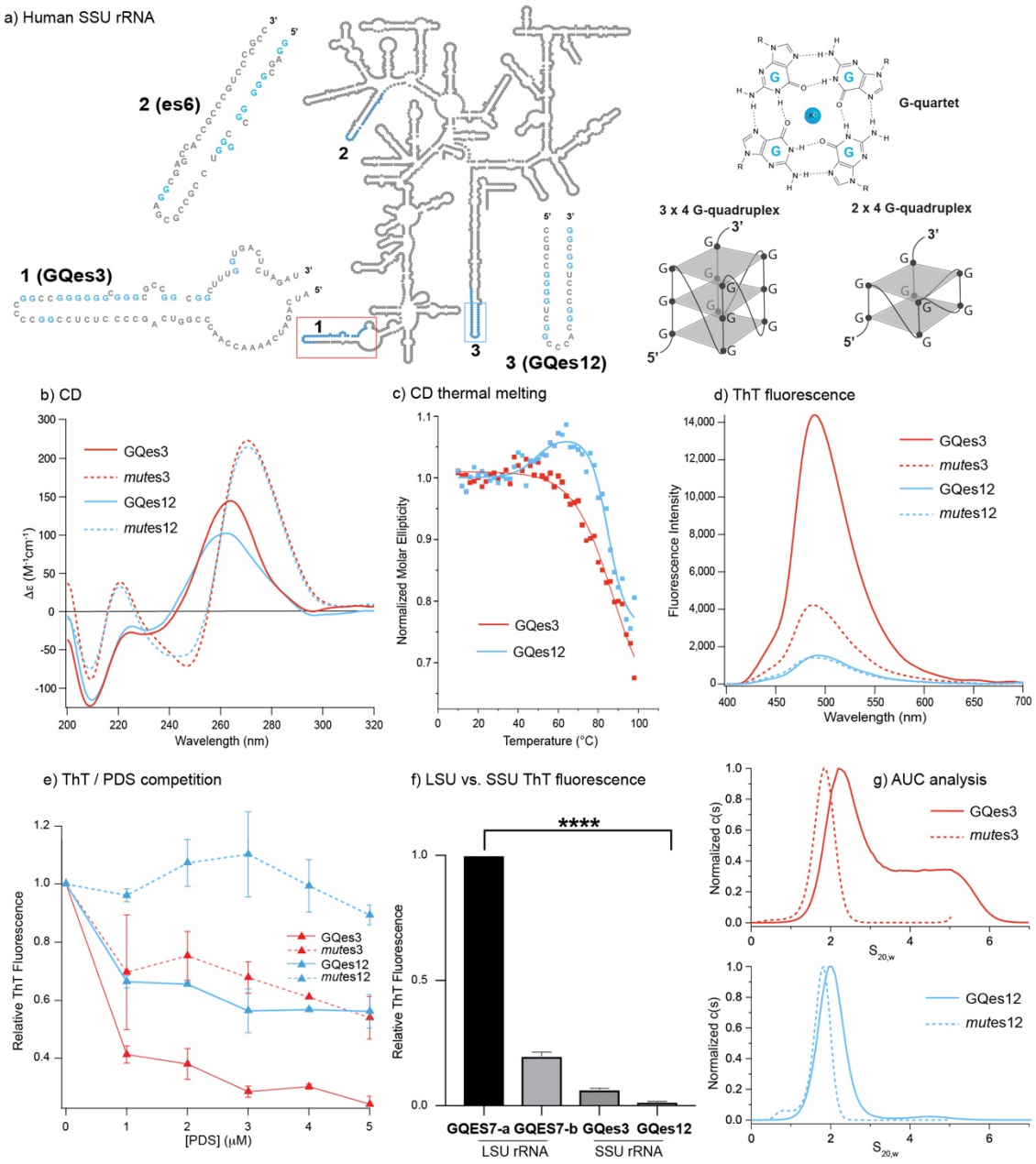


Figure 3.2. (a) The expanded secondary structure of the G-quadruplex regions on es3 or es12 of the human SSU rRNA. Schematic diagram of a single G-quartet, a 3 x 4 G-quadruplex (with tracts of at least three guanine residues), and a 2 x 4 G-quadruplex (with tracts of at least two guanine residues). (b) CD spectra of SSU G-quadruplex regions GQes3 and GQes12 and control RNAs *mut*es3 and *mut*es*12*. (c) Changes in CD amplitude at 260 nm as a function of temperature for GQes3 and GQes12. (d) ThT fluorescence in solutions of GQes3, GQes12, *mut*es3 or *mut*es*12* annealed in the presence of potassium ions. (e) ThT/PDS competition assay. After annealing, the rRNA oligomers were incubated with increasing concentrations of PDS and then fluorescence was recorded at 490 nm. (f) ThT fluorescence of LSU oligomers GQES7-a and GQES7-b and SSU oligomers GQes3

and GQes12. (g) Sedimentation velocities of GQes3, GQes12, *mut*es3 or *mut*es*12*. Data are shown as c(s) plots that show the distribution of sedimentation coefficients obtained by analysis using the program SEDFIT (121). The statistical significance relative to GQES7-a-ThT fluorescence is indicated by asterisks using an ordinary one-way ANOVA with Dunnett's post-hoc test. **** $P < 0.0001$.

3.3.2.2. Thioflavin T (ThT) fluorescence.

ThT interacts specifically with G-quadruplexes, and in doing so fluoresces at 490 nm (122). Our results show that ThT in the presence of GQes3 gives a significantly stronger fluorescence signal than in the presence of the negative control *mut*es3, consistent with formation of G-quadruplexes by GQes3 (Figure 3.2D). However, GQes12-ThT fluorescence is not significantly higher than the negative control, consistent with formation of hairpin instead of G-quadruplex.

G-quadruplex stability is cation dependent ($K^+ > Na^+ > Li^+$). The cation dependence of ThT fluorescence is consistent with formation of G-quadruplexes by GQes3. ThT fluorescence is significantly more intense when annealed in K^+ than in Li^+ (Figure B.5). GQes12-ThT gives a stronger fluorescence in the presence of Li^+ than in K^+, consistent with our model in which GQes12 forms a hairpin and not a G-quadruplex.

3.3.2.3. ThT - pyridostatin (PDS) competition.

PDS is a G-quadruplex stabilizer that binds with greater affinity than ThT to G-quadruplexes and can displace ThT from G-quadruplexes (123). The results here show that G-quadruplex-induced ThT fluorescence of GQes3 is attenuated by addition of PDS, consistent with displacement of ThT from G-quadruplexes by PDS. Although ThT fluorescence in the presence of either GQes3 or GQes12 decreases as PDS concentration

is increased (Figure 3.2E), the effect on GQes3 is greater than the effect on GQes12. Under the conditions of our experiment, for GQes3 the maximum fluorescence intensity decrease is ~80% while the weaker signal of the *mut*es3 control decreases by approximately 40%. For GQes12, the fluorescence intensity decreases by ~40% while the signal for *mut*es12 control remains constant. The data are consistent with a model in which GQes3 forms G-quadruplexes in solution whereas GQes12 forms a hairpin in the absence of a G-quadruplex stabilizer. The G-quadruplex stabilizer shifts the GQes12 equilibrium to favor G-quadruplexes.

3.3.2.4. LSU and SSU G-quadruplexes.

We compared the ThT fluorescence of 3 x 4 sequences from the LSU with 2 x 4 sequences GQes3 and GQes12 from the SSU, anticipating that the 3 x 4 LSU signals would be more intense than the 2 x 4 SSU signals. We previously characterized rRNA oligomers and polymers containing 3 x 4 sequences from the tentacles of *H. sapiens* LSU rRNA (111). These sequences form highly stable G-quadruplexes *in vitro*. That work focused on 3 x 4 sequences within *tentacles a* and *b* of ES7. The LSU sequences are GQES7-a, sequence **1** in Table 3.2 and GQES7-b, the first four G-tracts of sequence **3** of Table 3.2. The G-scores of GQES7-a and GQES7-b are much greater than those of GQes3 and GQes12 (GQES7-a, 60; GQES7-b, 38; GQes3, 21; GQes12, 20).

The SSU sequences GQes3 and GQes12 give significantly weaker ThT fluorescence than the LSU sequences GQES7-a and GQES7-b. ThT fluorescence of GQES7-b is approximately 20% of that of GQES7-a, while GQes3-ThT fluorescence is around 6% and GQes12-ThT fluorescence is ~1.5% (Figure 3.2F). Overall, the

experimental and computational results are self-consistent: the 3 x 4 G-quadruplexes in the human LSU rRNA appear to be considerably more stable and more extensive than the 2 x 4 G-quadruplexes of the SSU rRNA.

3.3.2.5. Analytical Ultracentrifugation (AUC).

The results of sedimentation velocity experiments (Figure 3.2G) are consistent with and support results described above. The control sequence *mut*es3, designed to be an unstructured single-strand, shows a single AUC species with an S value of 1.8 S. G-quadruplex sequence. GQes3 has a more complex c(s) distribution with major species of 2.5 S and 4.7 S (Figure 3.2G, top panel). The elevated baseline between these species suggests an interacting self-associating system, probably a mono-dimer equilibrium. The molecular weights of the two species observed for GQes3 indicates that they most likely correspond to monomer and dimer forms in equilibrium.

The difference in sedimentation coefficients between the GQes3 monomer and *mut*es3 indicates that monomeric GQes3 folds into a compact form, consistent with G-quadruplex formation. For comparison, the change in sedimentation of a DNA oligomer upon G-quadruplex formation is a similar to the change observed for GQes3. Specifically, the folding of a 24 nucleotide DNA oligomer into a G-quadruplex results in a change from 1.4S (unfolded) to 2.0S (G-quadruplex) (124). Additional sedimentation velocity experiments (data not shown) showed that as the concentration of GQes3 is decreased, the fractional of 2.5S monomer increases, consistent with equilibrium between monomer and a dimer. The monomer-dimer concentration dependence of GQes3 mobility is confirmed by gel electrophoresis experiments (Figure B.6).

The AUC behavior of GQes12 (Figure 3.2G, bottom panel) differs from that of GQes3. The *mut*es12 control (unstructured) shows a single species with 1.75S, while under folding conditions GQes12 shows a single species of 2.03 S. S values are determined with 0.1S precision, so the difference is significant, but is smaller than expected for folding of GQes12 into a G-quadruplex. The data indicate folding of GQes12, presumably to a hairpin, which is less compact than a G-quadruplex.

3.3.2.6. Association of SSU rRNA with BioTASQ.

G-quadruplex formation by the *H. sapiens* SSU rRNA was probed using BioTASQ, a biotin-linked small molecule that binds specifically to G-quadruplexes (26, 27). BioTASQ associates with streptavidin beads and appears to provide an unambiguous assay for G-quadruplex formation; RNA that associates with BioTASQ is prevented by the beads from entering the gel during electrophoresis. As a negative control, we used *E. coli* SSU rRNA, which lacks GQes3, GQes12 or other potential surface-exposed G-quadruplex forming sequences. Our results indicate the *H. sapiens* SSU rRNA but not *E. coli* SSU rRNA binds to BioTASQ (Figure 3.3A) consistent with G-quadruplex formation by *H. sapiens* SSU rRNA. The simplest interpretation of this data, consistent with the other experiments presented here, is that *in vitro* GQes3 forms G-quadruplexes within *H. sapiens* rRNA.

We also investigated binding of GQes3 and GQes12 to BioTASQ. GQes3 binds well to BioTASQ while GQes12 binds poorly (Figure 3.3A). Treatment of cells with G-quadruplex stabilizers has been shown to significantly increase the efficiency of BioTASQ to pull down G-quadruplex forming RNAs (27). Here, we observe that the addition of the

G-quadruplex stabilizer PDS increases BioTASQ binding for all putative G-quadruplex forming rRNAs, including GQes12, but not for the negative controls. The results of these experiments suggest that GQes3 readily forms G-quadruplexes while GQes12 forms G-quadruplexes under stabilizing conditions. GQes3, GQes12 and possibly other SSU 2 x 4 sequences like the one found in es6 **(**Figure 3.2A**)** are probably responsible for the formation of G-quadruplexes within the intact *H. sapiens* SSU rRNA.

BioTASQ also binds human 28S rRNA but not *T. thermophilus* 23S rRNA (Figure 3.3A) which lacks surface-exposed G-quadruplex regions, indicating the formation of these secondary structures in the human LSU rRNA and corroborating our initial findings (111).

3.3.2.7. Conservation of 2 x 3 sequences of the mammalian SSU rRNA.

We explored the phylogeny of G-quadruplex forming sequences in es3 throughout the Eukaryotic domain. Multiple sequence alignments (MSAs) indicate that all mammalian species contain es3 sequences with the potential to form 2 x 3 G-quadruplexes (Figure 3.3B). G-tract polarization is conserved in mammalian es3 sequences. This pattern of conserved polarization was observed previously within expansion segments of the LSU. Using polarization as a marker for G-quadruplex formation, these results suggest that the last mammalian common ancestor had G-quadruplex forming sequences within es3.

3.3.2.8. Parallel evolution of es3 within the ray-finned fishes.

Our initial SSU MSA included only a single representative of the ray-finned fishes (*Actinopterygii*) – *D. rerio*. The *D. rerio* sequence appeared anomalous in that es3 is long, comparable in length to es3 of mammals (Figure 3.3B). No 2 x 4 G-quadruplex forming

sequences are observed in es3 of *D. rerio*. To investigate the evolution of es3 within the ray-finned fishes, we incorporated additional species to our SSU MSAs. The results show that es3 length is variable within ray-finned fish. Two of the ray-finned fish species in our alignment contain 2 x 4 G-quadruplex forming sequences. However, the 2 x 4 G-quadruplexes of ray-finned fish are not polarized.

The combined results suggest an uneven process of sequence evolution of es3 within the ray-finned fishes compared within mammals. The prospect of parallel evolution of G-quadruplex forming sequences between mammals and ray-finned fishes eliminates some possibilities for the underlying evolutionary pressures. The data limit the basis for G-quadruplex formation to characteristics beyond those exclusive to mammalian lineages but rather to those that are common to all chordates.

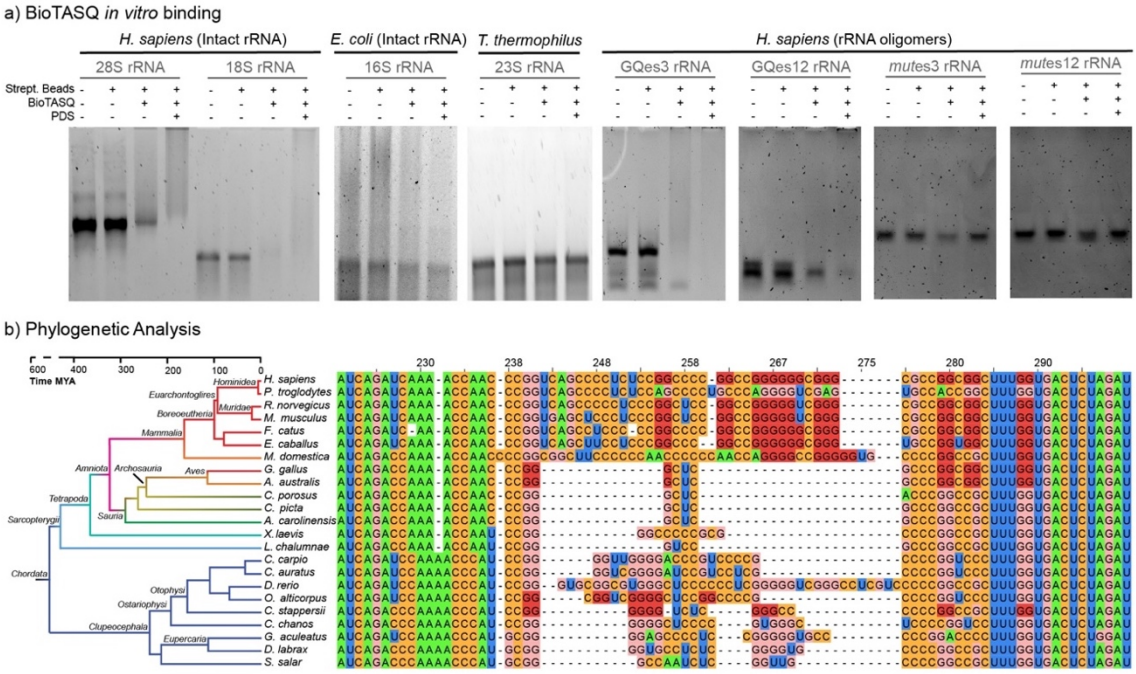


Figure 3.3. (a) BioTASQ binds to human LSU and SSU RNAs *in vitro* but not to *E. coli* or T. *thermophilus* rRNAs. BioTASQ binds G-quadruplex forming fragments of human SSU RNA but not to control RNA (b) Multiple Sequence Alignment of es3 for chordate species.

G-tracts are common in chordates, specifically in mammals. Guanines from G-quadruplex forming sequences are highlighted with dark red, all other guanines are pink. All nucleotides are numbered in accordance with *H. sapiens* 18S rRNA. Human 28S (LSU), 18S (SSU) and *E. coli* 16S (SSU) rRNAs were extracted from cells. *T. thermophilus* 23S rRNA was synthesized *in vitro*. rRNAs were annealed in the presence of potassium and magnesium at pH 7.5, followed by incubation with BioTASQ and streptavidin-coated beads.

3.4. Discussion

Previously we identified sequences in the human LSU rRNA that form G-quadruplexes in vitro (111). G-quadruplex sequences are a general feature of tentacles of chordate ribosomes. We identified G-quadruplex forming sequences in rRNA using the canonical 3 x 4 sequence motif. Sequences falling within this established motif were experimentally demonstrated to form G-quadruplexes in vitro. Recently however, others have shown that RNA sequences containing short tandem G-tracts that do not meet the 3 x 4 motif form G-quadruplexes (18-21, 34, 113). A relaxed 2 x 4 criterion extends the repertoire of G-quadruplex forming regions, for example, to the UTRs of mRNAs encoding for polyamine synthesis proteins (18). The development of next generation RNA sequencing strategies for mapping G-quadruplexes has significantly increased the ability to identify G-quadruplex sequences across the human transcriptome.

Here, applying the 2 x 4 criteria to human rRNA we identified seven potential G-quadruplex sequences in the SSU. Three of these are located on the ribosomal surface near the termini of the rRNA tentacles of es3, es6, and es12 (Figure 3.1). Our experimental results indicate that rRNA oligomers derived from these regions, as well as the native 18S rRNA, can form G-quadruplexes *in vitro*. This report represents the first evidence of G-quadruplex formation of any SSU rRNA.

The surfaces of both the LSU and the SSU of the human ribosome contain a sufficiently large number of G-quadruplex forming regions (Figure 3.4) that it seems possible that in some environments G-quadruplexes might dominate interactions of ribosomes with other cellular components. Extended arrays of G-tracts on solvent exposed regions of rRNA suggest roles in protein recruitment and polysome assembly. The association of the protein FUS (125, 126) with rRNA tentacles (111) is consistent with a model in which ribosomes participate in RNA mediated liquid-liquid phase separation (49).

Our phylogenetic analysis suggests that 2 x 4 sequences within the chordate SSU rRNAs have a complex evolutionary history. The data are consistent with a model in which 2 x 4 sequences evolved in parallel in distant chordate species. The phylogenetic conservation G-quadruplex sequences in warm-blooded animals and their surface localization suggest conserved function.

In yeast, ES27 has been shown to recruit specific proteins to the ribosome (13, 127). In mammals, ES7 is extended by tentacles, reaching a size zenith in humans. Our results suggest LSU tentacles confer increased complexity compared with protists in part via ability to form G-quadruplexes. rRNA tentacles appear to be dynamic, switching between G-quadruplex and duplex conformations depending on environment or protein association.

rRNA makes up more than 80% of the cellular RNA, suggesting that most of the G-quadruplexes in cells at any given time could be contributed by rRNA and hence that the RNA G-quadruplexome could be ribosome-centered. Historically, study of RNA G-quadruplexes has been directed to poly-adenylated mRNAs and long non-coding RNAs. Most transcriptome-wide studies have explicitly excluded rRNAs from their analyses (27,

34). The discovery of G-quadruplex regions on rRNA was made only recently (111). The results presented in this study corroborate these initial findings and extend the G-quadruplex forming capability of rRNA to the *H. sapien*s SSU.

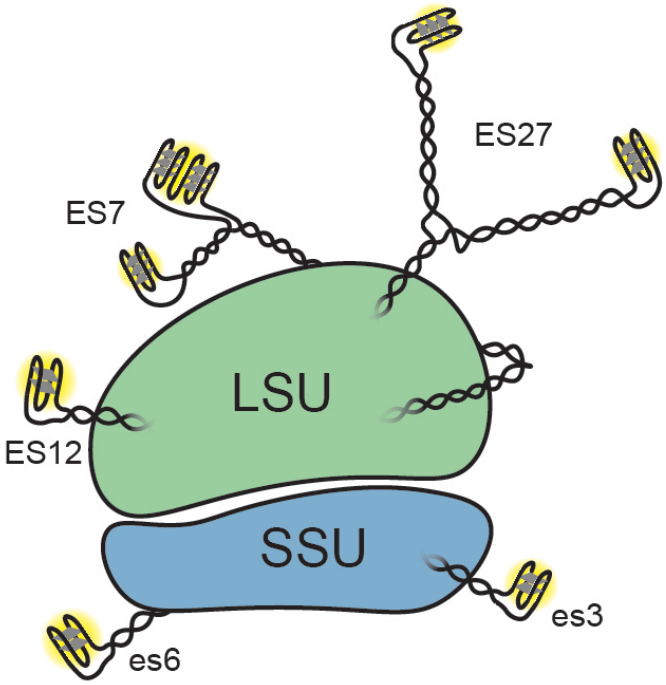


Figure 3.4. Schematic representation of the *Homo sapiens* ribosome with G-quadruplexes. ES lengths are not drawn to scale. This simplified schematic does not indicate the possibility of inter-ES and inter-ribosome G-quadruplexes or of the interconversion between G-quadruplexes and duplexes.

3.5. Methods

3.5.1. RNAs.

GQes3, GQes12, *mut*es3, and *mut*es12 were purchased from Integrated DNA Technologies. GQES7-a and GQES7-b were synthesized *in vitro* by transcription. *H. sapiens* 28S and 18S rRNAs were extracted from HEK293 cells. Briefly, HEK293 cells were grown to 60% confluency and total RNA was extracted with TRI reagent®. Intact rRNAs were extracted with a pipette from an agarose gel by running the rRNA into wells in the center of the gel. The rRNA was then precipitated with 5 M Ammonium Acetate-Acetic Acid (pH 7.5) with excess ethanol. *E. coli* 16S rRNA was extracted from DH5α *E. coli* strain using the same method. RNA sequences are listed in Table B.1.

3.5.2. RNA Annealing.

Before any experiment, RNAs were annealed by heating at 95 °C for 5 min and then cooled to 25 °C at 1°C/min and incubated for 10 min at 4°C.

3.5.3. CD Spectroscopy.

RNAs solutions were prepared at a final concentration of 10 µM (strand) and annealed as described above in the presence of 150 mM KCl and 10 mM Tris-HCl, pH 7.5. Spectra were acquired from 320 nm to 200 nm at a constant temperature of 20 °C on a Jasco J-815 spectropolarimeter using 1-mm cuvettes, at a rate of 100 nm/min with 1-s response, a bandwidth of 5 nm, averaged over three measurements. The same buffer minus RNA was used as the baseline. Igor Pro software was used to smooth the data. The expression $\Delta\varepsilon = \theta/(32{,}980 \times c \times l)$, where c is the molar concentration and l is the

cuvette path length was used to obtain the molar ellipticity from the observed ellipticity (θ, mdeg).

3.5.4. ThT Fluorescence.

RNAs solutions were prepared at a final concentration of 10 μM (strand) and annealed as described above in the presence of 150 mM KCl, 10 mM Tris-HCl, pH 7.5, and 2 μM ThT. For cation dependency experiments, either 50 mM KCl or LiCl was used. For LSU vs. SSU rRNA G-quadruplex-formation comparison (Figure 3.2F), RNAs were prepared at a final strand concentration of 1 μM. After annealing, RNAs were loaded onto a Corning® 384 Well Flat Clear Bottom Microplate and fluorescence was recorded from 300 nm to 700 nm, exciting at 440 nm. Fluorescence data were acquired on a BioTek Synergy™ H4 Hybrid plate reader.

For ThT/PDS competition assays, PDS was added after RNA annealing at final concentrations of 1 μM, 2 μM, 3 μM, 4 μM, or 5 μM. Mixtures were allowed to sit at room temperature for 10 min before data acquisition.

3.5.5. Analytical Ultracentrifugation (AUC).

RNAs were prepared at a final OD_{260} of 1.0 in the presence of 50 mM KCl and 10 mM Tris-HCl, pH 7.5. Sedimentation velocity measurements were carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0 °C and at 40,000 rpm in standard 2 sector cells. Data (200 scans collected over a 10 hour centrifugation period) were analyzed using the program Sedfit (121) in the continuous c(s) mode or by a model assuming discrete, noninteracting species

(www.analyticalultracentrifugation.com). Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0 °C and buffer viscosity was measured on an Anton Paar Automated Microviscometer AMVn. For the calculation of frictional ratio, 0.55 mL/g was used for partial specific volume and 0.3 g/g was assumed for the amount of water bound.

3.5.6. BioTASQ Binding.

RNAs solution were prepared at a final concentration of 15 nM (*H. sapiens* 28S, 18S rRNAs; *E. coli* 16S rRNA; *T. thermophilus* 23S rRNA) and 1 μM (GQes3, GQes12 or *mut*es3, *mut*es12) in the presence of 50 mM KCl, 1 mM $MgCl_2$ and 10 mM Tris-HCl, pH 7.5, and were annealed by heating at 75 °C for 1 min and cooling to room temperature at 1 °C/min. For (+) PDS samples, PDS was added to a final concentration of 5 μM after RNAs were annealed and allowed to sit at room temperature for 10 min. BioTASQ was added to the annealed RNA samples to a final concentration of 20 μM. Samples were allowed to mix by rocking at room temperature for 1 hr. Streptavidin-coated magnetic beads (GE Healthcare) were washed three times with 50 mM KCl and 10 mM Tris-HCl, pH 7.5. Then, 1.5 μg of the beads was added to the RNA-BioTASQ samples and allowed to mix overnight by rocking at room temperature. rRNAs were subsequently analyzed by native agarose electrophoresis (18S and 16S rRNAs) or 5% native-PAGE (GQes3, GQes12, *mut*es3 and *mut*es12).

3.5.7. Phylogeny and Multiple Sequence Alignments.

The SEREB MSA (64) was used as a seed to align additional eukaryotic es3 sequences. The SSU rRNA sequences in the SEREB MSA were used to search (128) the

NCBI databases (129) for SSU rRNA sequences. The SEREB database has sequences from 10 chordate species; sixteen additional chordate species were added to the es3 MSA (Figure 3.3). Sequences were incorporated into the SEREB-seeded MSA using MAFFT (130) and adjusted manually using Jalview (131). Manual adjustments incorporated information from available secondary structures. In some cases, the positions of G-tracts in sequences with large gaps relative to H. sapiens are not fully determined, as they can be aligned equally well with flanking G-tracts in the MSA. Alignment visualization was done with Jalview (131). The phylogenetic tree and the timeline of clade development were inferred from TimeTree (132).

CHAPTER 4

HUMAN RIBOSOMAL G-QUADRUPLEXES REGULATE HEME BIOAVAILABILITY

This chapter is adapted from the manuscript in review: Mestre-Fos, S., Ito, C., Moore, C. M., Reddi, A. R., Williams, L. D. "Human Ribosomal G-Quadruplexes Regulate Heme Bioavailability". The author of this document contributed to this work by conceiving, designing and performing all experiments except the HS1 transfection experiment, making all figures except Figure 4.4d, making all tables, making all supplementary materials, and co-writing the manuscript.

4.1. Abstract

The *in vitro* formation of stable G-quadruplexes (G4s) in human ribosomal RNA (rRNA) was recently reported. However, their formation in cells and their cellular roles have not been resolved. Here, by taking a chemical biology approach that integrates results from immunofluorescence, G4 ligands, heme affinity reagents, and a genetically encoded fluorescent heme sensor, we report that human ribosomes can form G4s *in vivo* that regulate heme bioavailability. Immunofluorescence experiments indicate that the vast majority of extra-nuclear G4s are associated with rRNA. Moreover, titrating human cells with a G4 ligand alters the ability of ribosomes to bind heme and disrupts cellular heme

bioavailability as measured by a genetically encoded fluorescent heme sensor. Overall, these results suggest ribosomes are central hubs of heme metabolism.

4.2. Introduction

Cells tightly control heme concentration and bioavailability (55-57) because it is essential but potentially cytotoxic. Proteins that regulate heme concentration are relatively well understood; structures and mechanisms of all eight heme biosynthetic enzymes and the heme degrading heme oxygenases are known (55-57). However, regulation of heme bioavailability, including intracellular trafficking from sites of synthesis in the mitochondrial matrix or uptake at the plasma membrane, is poorly understood. Current paradigms for heme trafficking and mobilization involves heme transfer by unknown proteinaceous factors and largely ignore contributions from nucleic acids. Given that the first opportunity for protein hemylation occurs during or just after translation, ribosomal RNA (rRNA) or proteins (rProteins) may be critical for shepherding labile heme to newly synthesized proteins.

We hypothesized that intracellular heme bioavailability is regulated in part by rRNA quadruplexes (G4s). G4s are nucleic acid secondary structures that are composed of four guanine columns surrounding a central cavity that sequesters monovalent cations. Our hypothesis is based on the high affinity of heme for G4s (K_D ~10 nM) (50-52), our work demonstrating that rRNA forms extensive G-tracts *in vitro* (17, 111), the extreme stabilities of rRNA G4s *in vitro* (17, 111) and the extraordinary abundance of rRNA *in vivo* (10).

DNA G4s are proposed to help regulate replication (133), transcription (134), and genomic stability (135). In RNA, G4s are associated with untranslated regions of mRNA

and have been proposed to regulate translation (35, 36, 40). However, the *in vivo* folding state and functional roles of G4s are under debate. Eukaryotic cells contain helicases that appear to unfold RNA G4s (31) although counter arguments have been put forth (27, 32). The density of G4 sequences on surfaces of the human ribosome, which is extremely abundant, is high, with 17 G4 sequences in the 28S rRNA and 3 in 18S rRNA (Figure 4.1A). Previous to this report, it was not known if human ribosomes form G4s *in vivo* or what their functions might be.

Indeed, herein we present evidence that rRNA forms G4s *in vivo* that regulate cellular heme homeostasis. Results of immunofluorescence experiments with a G4 antibody, RNA pulldowns and competition experiments with G4 ligands provide strong support for *in vivo* formation of G4s by rRNA tentacles. We find that G4s on ribosomes bind heme *in vitro* (Figure 4.1B) and that perturbation of G4s *in vivo* with G4 ligands affects *in vivo* heme interactions and heme bioavailability, as measured by heme affinity reagents and genetically encoded heme sensors. Taken together, the results here indicate that surface-exposed rRNA G4s interact with heme in cells and suggest that ribosomes are hubs for cellular heme metabolism.

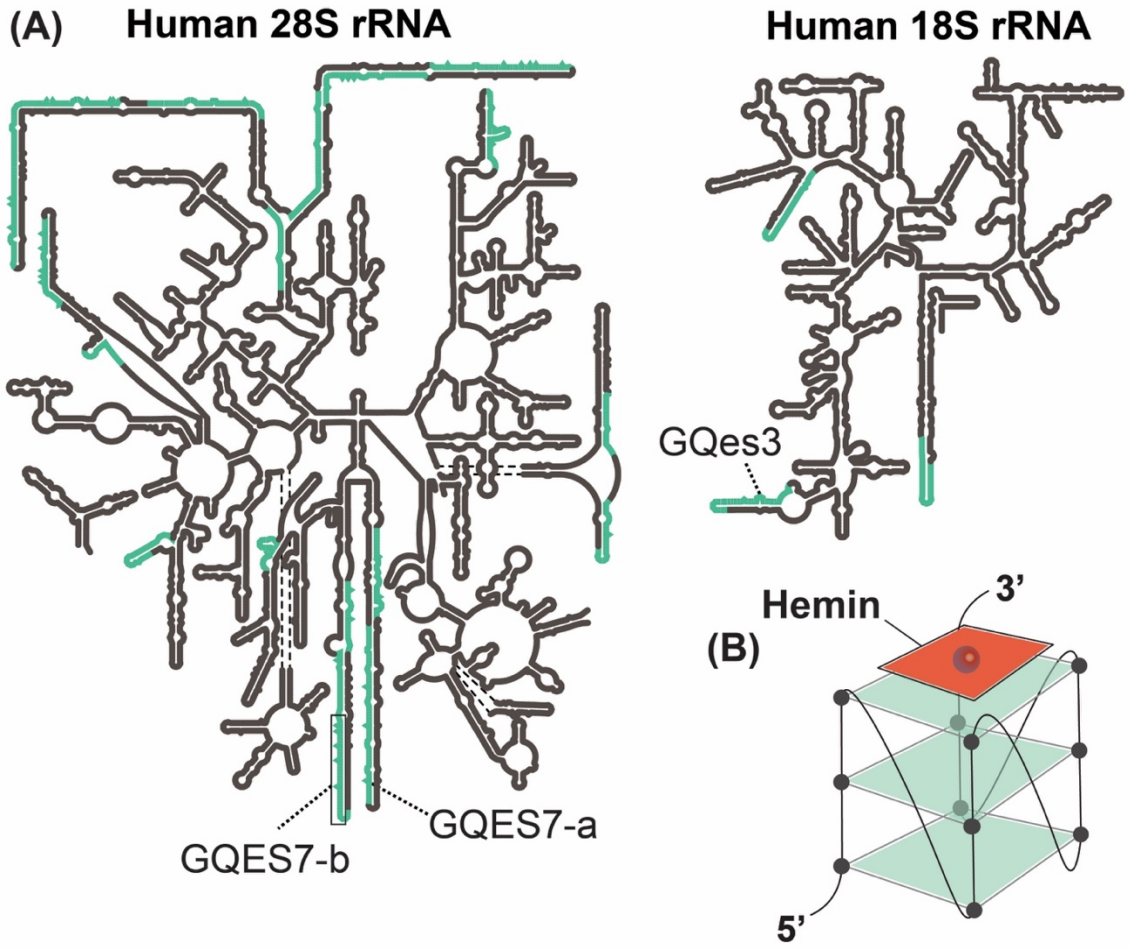


Figure 4.1. (A) Secondary structures of the human LSU rRNAs (5.8S and 28S) and SSU rRNA (18S). G4 sequences are highlighted in green. rRNA-based oligomers from the LSU (GQES7-a, GQES7-b) and from the SSU (GQes3) are indicated. (B) Schematic representation of a hemin-G4 complex.

4.3. Results

4.3.1. Ribosomal RNA forms G4s in vivo.

Confocal microscopy and G4-pulldowns were used to determine if human ribosomes form G4s *in vivo*. For confocal microscopy, we used the BG4 antibody, which selectively targets G4s (24, 25) and has been broadly used for visualizing DNA G4s and non-ribosomal RNA G4s in cells. (25, 136-138) Our method of permeabilizing cells for

antibody treatment does not permeabilize the nuclei (139). Therefore, DNA G4s were not anticipated or observed. To identify ribosome associated G4s, we determined the extent to which antibodies to rProtein L19 (eL19) and to G4s colocalize and how this is altered when cells are subjected to RNase or G4 ligand PhenDC3, which are expected to modulate G4-L19 colocalization. Prior to antibody addition, cells were crosslinked with paraformaldehyde to lock G4s *in situ*. This procedure is intended to prevent induction of G4s by the antibody and has been shown to reduce levels of detection of G4s (27). The extent of L19 and G4 antibody colocalization suggests that a fraction of ribosomes form G4s (Figure 4.2A,C) and that most G4s are associated with ribosomes. Specifically, we find that ~83% of BG4 pixels colocalize with L19, indicating that the vast majority of G4s *in vivo* are associated with ribosomes (Figure 4.2C, green bar) and are therefore rRNA G4s. Conversely, only 5% of L19 pixels colocalize with BG4 (Figure 4.2C, WT red bar), indicating that only a specialized fraction of ribosomes contains G4s. Similar results were obtained using an antibody against rProtein uL4 instead of L19 (not shown).

PhenDC3, which is known to induce and stabilize G4s, (140, 141) appears to increase ribosomal G4 formation *in vivo*; treating cells with PhenDC3 increases L19-BG4 colocalization from 5 to ~24% (Figure 4.2C). The increase in colocalization upon PhenDC3 treatment supports formation of G4s by ribosomes. By contrast, treating cells with RNase A abolishes the L19-BG4 colocalization signal (Figure 4.2C). Together, these results indicate the colocalized BG4 signal is coming from a G4 forming RNA in close proximity to L19.

mRNA in the cytosol, in the unlikely event that they form G4s at high frequency (31), may confound our ability to selectively detect rRNA G4s. The high density of

ribosomes on the surface of the endoplasmic reticulum (ER) and the lower abundance of mRNA in this location as compared to the cytosol (142) motivated us to investigate if G4s colocalize with the ER. Toward this end, we determined the extent to which BG4 colocalizes with an antibody against an ER membrane protein (calnexin) (Figure 4.2B). Indeed, we find that ~45% of the BG4 signal colocalizes with the ER marker (Figure 4.2D, green bar), indicating a significant presence of RNA G4s at the ER membrane. As with L19, the fraction of the ER signal that colocalizes with G4s (~2%) is completely abolished by RNase (undetectable) and enhanced by PhenDC3 (12%) (Figure 4.2D). Altogether, the data are consistent with formation of RNA G4s by ER-bound ribosomes.

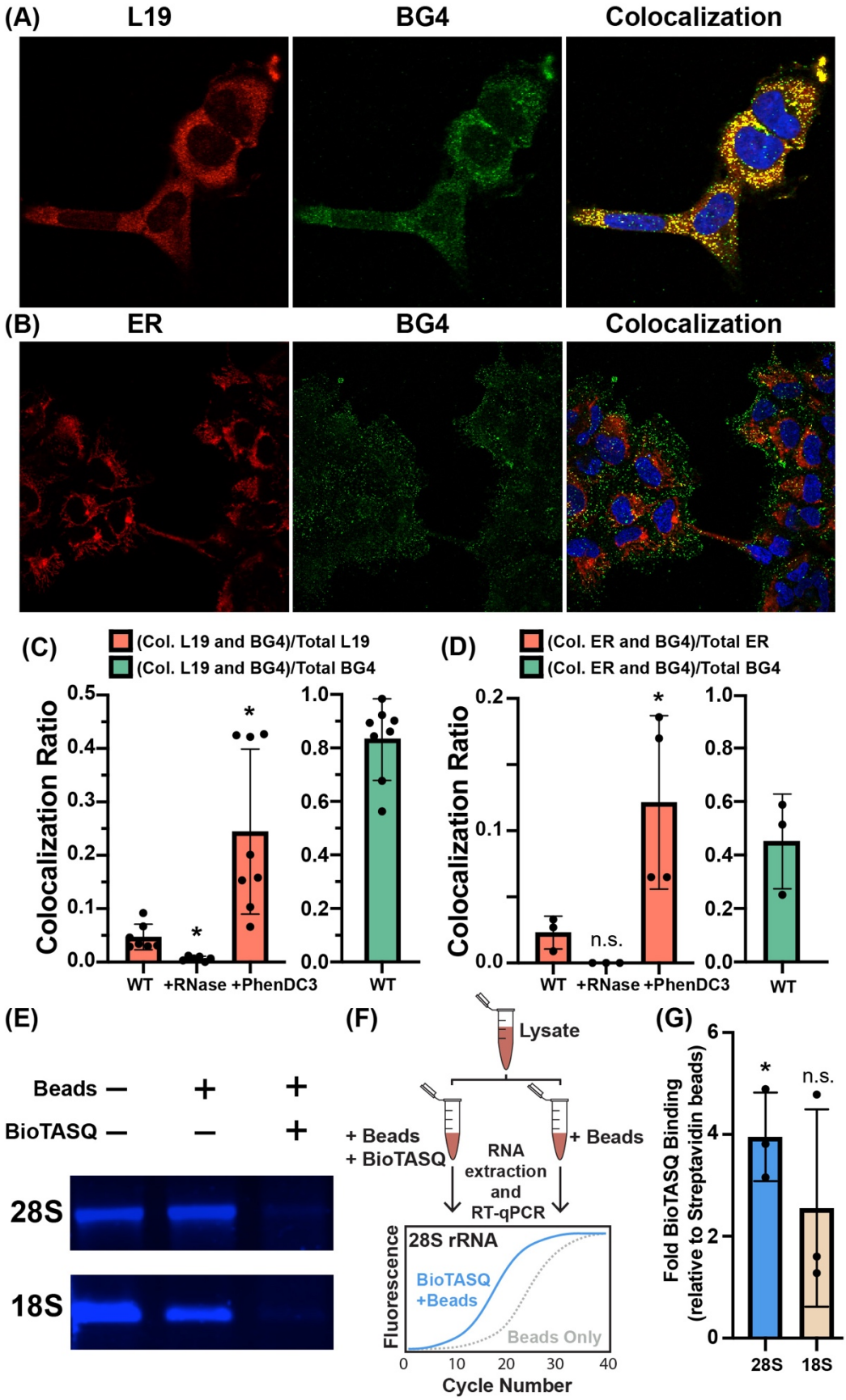

(A)
L19
BG4
Colocalization
(B)
ER
BG4
Colocalization
(C)
(Col. L19 and BG4)/Total L19
(Col. L19 and BG4)/Total BG4
Colocalization Ratio
WT
+RNase
+PhenDC3
WT
(D)
(Col. ER and BG4)/Total ER
(Col. ER and BG4)/Total BG4
Colocalization Ratio
n.s.
WT
+RNase
+PhenDC3
WT
(E)
Beads
BioTASQ
28S
18S
(F)
Lysate
+ Beads
+ BioTASQ
+ Beads
RNA extraction and RT-qPCR
28S rRNA
BioTASQ +Beads
Beads Only
Fluorescence
Cycle Number
(G)
Fold BioTASQ Binding (relative to Streptavidin beads)
n.s.
28S
18S

Figure 4.2. Ribosomal G4s in HEK293 cells. Colocalization of (A) ribosomal protein L19 or (B) endoplasmic reticulum (red) with RNA G4s (green). Nuclei were stained with DAPI (blue). (C) Extent of colocalization is quantitated as the ratio of colocalized pixels over total L19 pixels (red bars) or as the ratio of colocalized pixels over total BG4 pixels (green bar). Same analysis was done for ER-BG4 colocalization (D). The statistical significance relative to WT is indicated by asterisks using an ordinary one-way ANOVA with Dunnett's post-hoc test. Each dot represents a biological replicate. (E) G4 ligand BioTASQ binds to 28S and 18S rRNAs. In the presence of BioTASQ and streptavidin beads, human rRNAs do not enter the agarose gel. (F) Schematic representation of the BioTASQ pulldown protocol. (G) RT-qPCR analysis of rRNAs pulled down by BioTASQ. The statistical significance relative to a fold enrichment value of 1 is indicated by asterisks using a one sample *t* and Wilcoxon test. Each dot represents a biological replicate. Data in (G) are represented as RNA enrichment under "BioTASQ + streptavidin beads" conditions relative to control streptavidin beads. * $P < 0.05$. n.s. = not significant.

In an orthogonal approach, we pulled down RNA with BioTASQ (26, 27), which is a G4 ligand linked to biotin. BioTASQ captures G4s. We previously used BioTASQ to demonstrate that human rRNA forms G4s *in vitro* (Figure 4.2E) (17). Here, we captured rRNA G4s from crosslinked HEK293 cells by methods summarized in Figure 4.2F. BioTASQ captures 28S rRNA from cell lysates (Figure 4.2G), in agreement with our previous *in vitro* BioTASQ data and with observations of G4-L19 colocalization above. BioTASQ also captures 18S rRNA although the signal is significantly weaker. Taken together, our immunofluorescence and BioTASQ pulldown experiments provide strong evidence that human ribosomes form G4s *in vivo*.

4.3.2. Human ribosomes bind hemin in vitro

It has been suggested that G4s might associate with heme *in vivo* (63). *In vitro*, heme binds with high affinity to G4s by end-stacking (143-145) (Figure 4.1B). We used UV-visible spectroscopy to assay the binding of hemin to human rRNA. rRNA oligomers GQES7-a (Figure 4.3A), GQES7-b (Figure C.1A) or GQes3 (Figure C.1B) were titrated

into fixed amount of hemin. GQES7-a and GQES7-b are fragments of expansion segment 7 of human LSU rRNA (111). GQes3 is a fragment of expansion segment 3 of human SSU rRNA (17). Each of these oligonucleotides is known to form G4s and each caused a pronounced increase in the Soret band of hemin at 400 nm. The binding is specific for G4s as a mutant oligonucleotide, *mut*es3, that lacks G-tracts does not induce a change in the hemin Soret band (Figure C.1C). Larger human ribosomal components also bind heme. Intact 28S and 18S rRNAs extracted from human cells (Figure C.1D-E), assembled large (LSU) (Figure 4.3B) and small (SSU) (Figure C.1F) ribosomal subunits, and polysomes (Figure 4.3C) all induce changes in the hemin Soret bands, which is indicative of heme-rRNA interactions. The combined data are consistent with a model in which rRNA tentacles of human ribosomes bind to hemin *in vitro*.

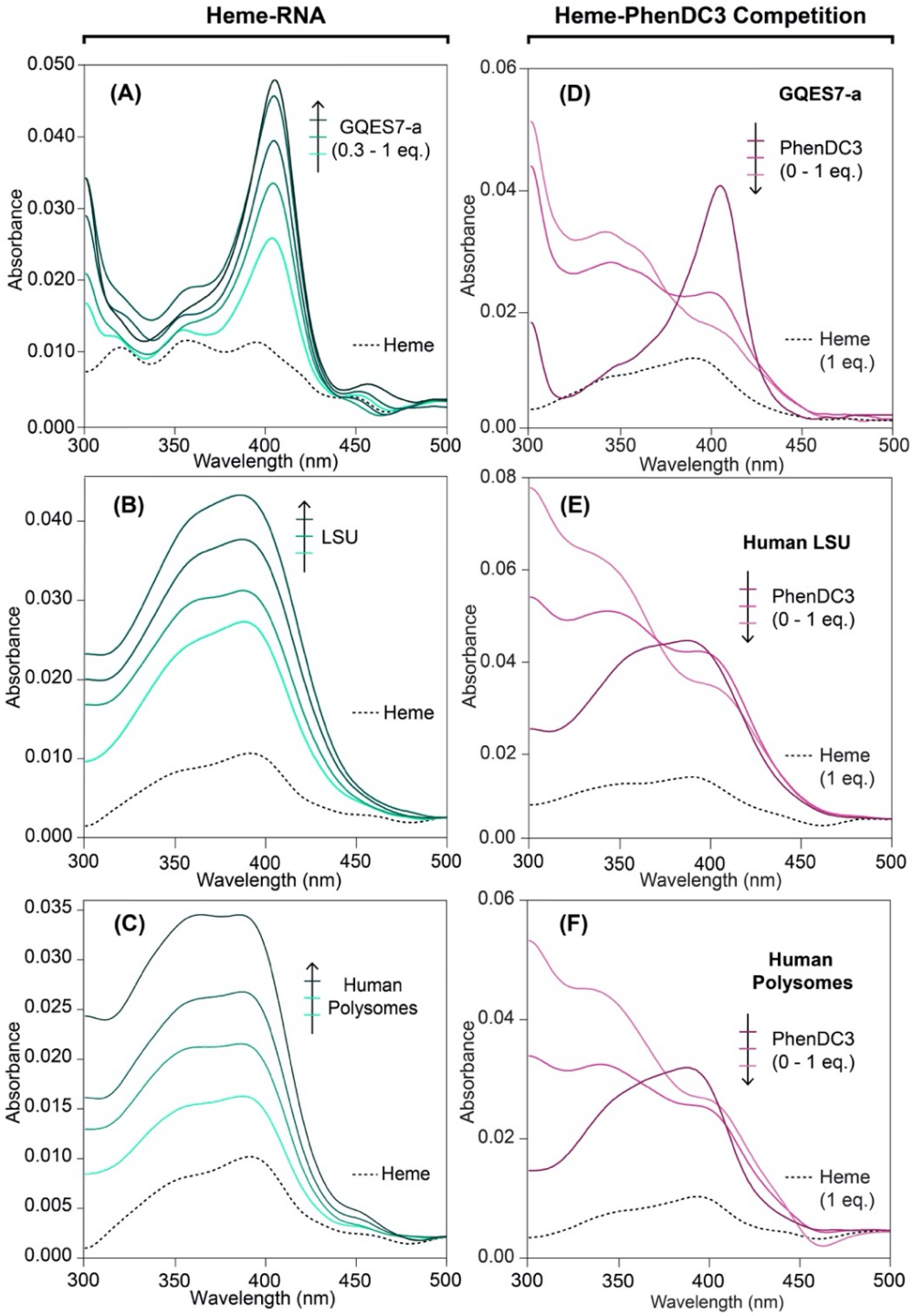


Figure 4.3. Human rRNA G4s bind heme *in vitro*. (A) UV-Vis spectra of heme during a titration with GQES7-a, (B) during a titration with the assembled LSU, and (C) during a titration with polysomes. (D) UV-Vis spectra of constant heme/GQES7-a during a titration with PhenDC3, (E) heme/LSU during a titration with PhenDC3, and (F) heme/polysomes during a titration with PhenDC3.

PhenDC3 was used to confirm binding of hemin to ribosomal G4s. PhenDC3, like hemin, end-stacks on G4s (63) and therefore competes with heme for binding to G4s. With fixed GQES7-a and hemin, addition of PhenDC3 causes a decrease in the intensity of the hemin Soret peak (Figure 4.3D) due to dissociation of heme. The same phenomenon is observed with assembled ribosomal particles (LSU: Figure 4.3E, SSU: Figure C.2A) and with polysomes (Figure 4.3F). Hemin that is associated with purified 28S and 18S rRNAs is also dissociated by PhenDC3 (Figures C.2B-C). Solutions of hemin with *mut*es3, however, do not show a change in the Soret peak upon the addition of PhenDC3 (Figure C.2D). Addition of PhenDC3 causes a slight increase in absorbance at 350 nm (Figures 4.3D-F). This phenomenon is intrinsic to PhenDC3, which absorbs at this wavelength (Figure C.2E). Taken together, the results here provide strong support for association of hemin with G4s on human ribosomes *in vitro*.

4.3.3. Human ribosomes bind heme in vivo.

We developed an assay that exploits differential interactions with hemin-agarose, an agarose resin covalently linked to heme, to report *in vivo* heme binding to ribosomes and rRNA. The degree to which any biomolecule interacts with heme in cells is inversely correlated with the extent to which it interacts with hemin-agarose upon lysis due to competition between endogenous heme and hemin-agarose for the heme-binding site. Therefore, the effects of heme binding factors *in vivo* can be monitored by determining if their interaction with hemin-agarose changes upon depletion of intracellular heme.

Accordingly, HEK293 cells were conditioned with and without succinylacetone (SA (60)), an inhibitor of heme biosynthesis. Lysates of these cells were incubated with

hemin-agarose, and hemin-agarose interacting rRNA was quantified by RT-qPCR. Consistent with previous work (146), treatment with 0.5 mM SA for 24 hours caused a 7-fold decrease in total cellular heme in HEK293 cells (results not shown). The results reveal that rRNA binding to hemin-agarose relative to control agarose lacking heme increases by ~4-fold in cells depleted of heme (Figure 4.4A). This result suggests that, under heme-depleted conditions, a greater fraction of rRNA heme binding sites are free and available to bind hemin agarose. In short, the data are consistent with a model in which ribosomal RNAs associate with endogenous heme.

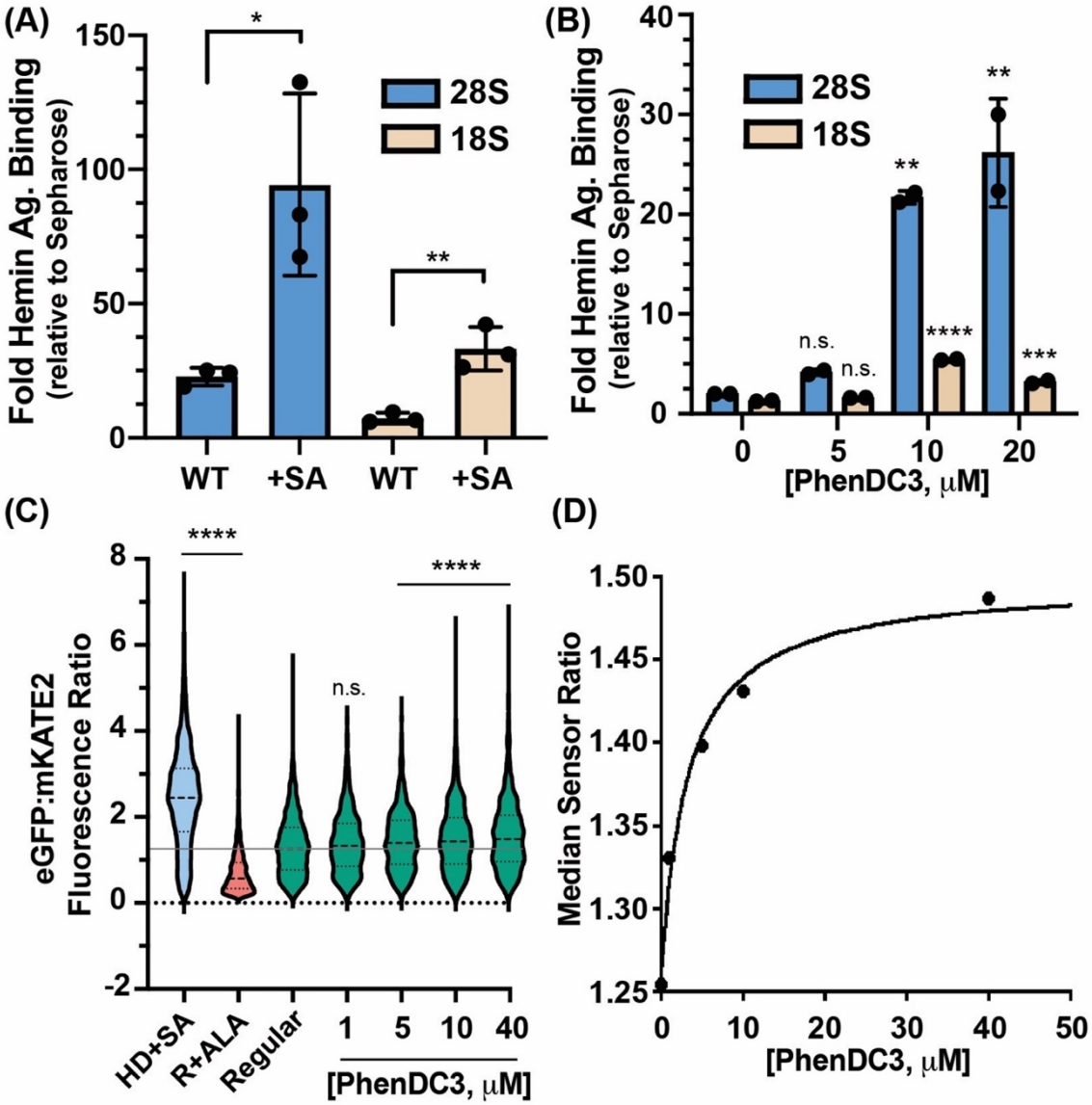


Figure 4.4. Ribosomes appropriate heme *in vivo* through rRNA G4s. (A) RT-qPCR analysis from untreated (WT) and SA-treated human cells. Statistical significance relative to WT is represented by asterisks using Student's *t*-test. Each dot represents a biological replicate. (B) RT-qPCR analysis from PhenDC3-treated HEK293 cells. Statistical significance relative to no treatment conditions is represented by asterisks using ordinary one-way ANOVA with Dunnett's post-hoc test. Each dot represents a technical replicate coming from individual biological replicates. The experiment was performed a total of 2 times with similar dose-dependent trends (Fig. C.3A). Data in (A) and (B) are represented as RNA enrichment in hemin agarose beads relative to control sepharose beads. (C) Single cell analysis of HS1-transfected HEK293 cells grown in heme deficient media containing succinylacetone (HD+SA), regular media containing 5-aminolevulinic acid (R +ALA), or

regular media (regular) with the indicated concentrations of PhenDC3. Statistical significance relative to regular conditions is represented by asterisks using the Kruskal-Wallis ANOVA with Dunn's post-hoc test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; n.s. = not significant; ($n \cong 1500$ cells). (D) Median HS1 sensor ratios obtained in (C) as a function to PhenDC3 concentration.

4.3.4. In vivo PhenDC3 increases binding of ribosomes to hemin agarose.

To determine if rRNA G4s bind heme *in vivo*, we treated HEK293 cells with the G4 ligand PhenDC3 (48 hrs at 37 °C). PhenDC3 and heme compete for binding to G4 rRNA *in vitro* (Figure 4.3). Thus, if rRNA G4s bind heme *in vivo*, PhenDC3 is expected to displace any rRNA bound heme. After cell lysis, excess hemin agarose is expected to out compete rRNA bound PhenDC3. Indeed, RT-qPCR reveals that PhenDC3 treatment of HEK293 cells causes a dose-dependent increase in binding of the LSU to hemin agarose (Figure 4.4B). A corresponding, but weaker signal is seen for the SSU, in agreement with the higher abundance of G4 regions in the LSU than in the SSU (Figure 4.1A). Treatment of HEK293 cells with carrier DMSO does not result in hemin agarose enrichment (Figure C.3B) or in heme-G4 interaction alterations (Figure C.3C). Taken together, these results indicate that G4s in rRNA bind heme in cells.

4.3.5. rRNA G4s regulate heme bioavailability in vivo.

To determine if rRNA G4s regulate heme homeostasis, we deployed a previously described genetically encoded ratiometric fluorescent heme sensor, HS1. HS1 is a tri-domain fusion protein consisting of a heme binding domain, cytochrome b_{562}, fused to fluorescent proteins, eGFP and mKATE2, whose fluorescence is quenched or unaffected by heme, respectively. Thus, the eGFP:mKATE2 fluorescence ratio is inversely correlated

with bioavailable heme as measured by HS1. HS1 was previously used to characterize heme homeostasis in yeast, bacteria, and mammalian cells, and was instrumental in identifying new heme trafficking factors and signals that alter heme biodistribution and dynamics (60, 147, 148). We asked if cytosolic heme bioavailablity is altered in response to PhenDC3 (60). As indicated in Figure 4.4C, single cell analysis of a population of ~1500 HEK293 cells per condition indicate the median HS1 eGFP/mKATE2 ratio increases upon heme depletion in heme deficient media containing SA (HD+SA) and decreases upon increasing intracellular heme when cells are conditioned with the heme biosynthetic precursor 5-aminolevulinic acid (ALA) to drive heme synthesis. Titration of PhenDC3 results in a dose dependent increase in HS1 sensor ratio, indicating heme is less bioavailable when it is displaced from G4s in rRNA. The fractional heme saturation of HS1 decreases by ~15% (Figure 4.4D). Together, our data indicate that rRNA G4s bind heme and regulate intracellular heme bioavailability.

4.4. Discussion

The results here provide strong evidence that ribosomal tentacles form G4s in human cells, and that these G4s are involved in appropriating heme. Immunofluorescence experiments with BG4 and L19 antibodies suggest a specialized fraction of cytosolic ribosomes (~5%) form G4s and that most extra-nuclear G4s (~83%) are ribosomal. The small fraction of ribosomes observed to form G4s *in vivo* contrasts with the high stability of ribosomal G4s *in vitro* (17, 111). This difference is consistent with Guo and Bartel, who propose that eukaryotic cells have a robust machinery that unfolds G4s (31). However, the high concentration of rRNA acts in opposition to the low frequency per ribosome, so the

RNA G4s are reasonably abundant. The RNA G-quadruplexome appears to be ribosome-centric.

We previously reported that surfaces of both the SSU and the LSU contain G4 sequences (17, 111). A broad variety of data are consistent with more extensive formation of G4s on the LSU than on the SSU. These data include:

- more abundant and more expansive G-tracts on the LSU than the SSU (17),
- greater conservation over phylogeny of LSU G-tracts than SSU G-tracts (17, 111),
- higher thermodynamic stability of LSU G4s than SSU G4s (17, 111),
- greater heme binding to G4 oligomers from the LSU than those from the SSU (Figure 4.3A, Figure C.1B),
- greater enrichment of LSU than SSU particles in BioTASQ pulldowns (Figure 4.2D),
- greater enrichment of LSU than SSU particles in hemin-agarose pulldowns (Figure 4.4A), and
- greater effect of *in vivo* PhenDC3 treatment on LSU than on SSU rRNA in hemin-agarose pulldowns (Figure 4.4B).

Our findings that rRNA G4 forms complexes with heme *in vivo* has major implications for the physiology of G4-heme interactions. Decades of *in vitro* biophysical and chemical characterization of G4-heme complexes have found that they interact with high affinity (K_D ~ 10 nM) and are potent redox catalysts, facilitating peroxidase and peroxygenase reactions (50-52). However, it remained unclear if heme-G4 complexes are formed *in vivo* and if heme-G4 catalyzed reactions were physiologically relevant.

Gray and coworkers (63) recently proposed that heme binds to G4s *in vivo*, based on the transcriptional response of cells to PhenDC3. PhenDC3 up-regulates heme degrading enzymes like heme oxygenase, and other iron and heme homeostatic factors. These responses were interpreted to support a model in which G4s sequester and detoxify heme in cells (63). Here, by exploiting differential interactions of *apo*-rRNA and heme-rRNA with hemin-agarose, we developed more direct methods to establish *in vivo* heme-G4 interactions, with a focus on rRNA G4s. Our results are consistent with the work of Gray *et. al.* in that we too conclude that G4s bind hemin *in vivo*. Moreover, our observation of rRNA-heme interactions *in vivo* supports a physiological role of rRNA G4-heme complexes in redox chemistry (149, 150). Indeed, work by Sen concurrent with ours has established G4-heme interactions *in vivo* by exploiting the peroxidase activity of G4-hemin complexes to self-biotinylate G4s in RNA and DNA using a phenolic-biotin derivative.

We propose that heme-rRNA G4 interactions may be important for protein hemylation reactions or buffering cytosolic heme. Indeed, PhenDC3, which competes for heme binding in G4s, causes a decrease in heme bioavailability as measured by the heme fluorescent sensor, HS1. This could be due to displacement of heme from rRNA G4s to a site that is less exchange labile, resulting in the observed decrease in HS1 heme binding. Alternatively, the upregulation in heme oxygenase due to PhenDC3 treatment (63) may decrease cellular heme levels, giving rise to the observed decrease in bioavailable heme. Overall, our results indicate rRNA G4s are sites of exchangeable heme in cells that may be available for heme dependent processes and hemylation reactions at the ribosome.

Taken together, the results here suggest that structural features of the human ribosome coupled with its high cytosolic abundance facilitate association with endogenous

hemin. These results provide potentially new insights into the molecules and mechanisms underlying intracellular heme trafficking and bioavailability, which are currently poorly understood (55-57, 151). Our results suggest that ribosomes, and G4 containing rRNA in particular, are central hubs of heme metabolism, acting to buffer intracellular heme and possibly regulate heme trafficking and cotranslational hemylation. The ribosome as a hub for heme is consistent with its role as an abundant and versatile sink for ions and small molecules, including antibiotics, (152) platinum-based drugs, (153-155) certain metabolites (156), and metal cations Mg^{2+}, Ca^{2+}, Mn^{2+}, and Fe^{2+}, K^{+} (157-163).

4.5. Methods

4.5.1. Cell culture.

HEK293 cells were cultured in Dulbecco's Modified Eagle Media (DMEM) containing 4.5 g/L Glucose without Sodium Pyruvate and L-Glutamine (Corning) supplemented with 10% fetal bovine serum (FBS) (Corning) and 2% penicillin-streptomycin solution (Gibco) in a humidified incubator kept at 37 ^{o}C with a 5% carbon dioxide atmosphere.

4.5.2. RNAs.

GQES7-a and GQES7-b were synthesized *in vitro* by transcription (HiScribe™ T7 High Yield RNA Synthesis Kit; New England Biolabs). GQes3 and *mut*es3, were purchased from Integrated DNA Technologies. Human 28S and 18S rRNAs were extracted from HEK293 cells with TRIzol (Invitrogen). Intact rRNAs were isolated by pipetting from a native agarose gel after running the rRNA into wells in the center of the gel. The rRNA

was then precipitated with 5 M ammonium acetate-acetic acid (pH 7.5) with excess ethanol. RNA sequences are listed in Table C.1.

4.5.3. RNA annealing.

RNAs were annealed by heating at 95˚C for 5 min and cooled to 25˚C at 1˚C/min and incubated for 10 min at 4˚C.

4.5.4. UV-Visible Absorbance Heme-RNA Binding.

Stock solutions of hemin chloride (1mM) were prepared in DMSO. Prior to use, the hemin chloride solution was sonicated for 10 min. RNAs (GQES7-a, GQES7-b, GQes3, *mut*es3) were annealed as described above in 50 mM KCl and 10 mM Tris-HCl, pH 7.5 in increasing RNA concentrations (for rRNA oligomers: from 0.3 to 1 equivalents of heme). The annealing buffer for intact 28S and 18S rRNAs and assembled ribosomal subunits and polysomes was the same as that of the rRNA oligomers except for the inclusion of 10 mM $MgCl_2$. After binding, heme was added to a final concentration of 3 μM. Solution were allowed to stand at room temperature for 30 min then loaded onto a Corning® 384 Well Flat Clear Bottom Microplate. Absorbances were recorded from 300 nm to 700 nm on a a BioTek Synergy™ H4 Hybrid plate reader.

4.5.5. UV-Visible Absorbance, Heme-PhenDC3 Competition Assay.

For heme - PhenDC3 competition assaya, RNAs were annealed and allowed to bind to heme as above. Final heme concentration was 3 μM. Final RNA concentrations were: GQES7-a (3 μM), intact human 18S rRNA (65 nM), intact human 28S rRNA (22 nM). After solutions were inclubated for 30 min at room temperature, PhenDC3 or carrier

DMSO was added to final concentrations consisting of 1.5 μM, 3 μM, and 6 μM. Samples were allowed to stand at room temperature for 15 minutes and were loaded onto a Corning® 384 Well Flat Clear Bottom Microplate. Absorbance was recorded from 300 nm to 700 nm.

4.5.6. Total Heme Quantification of Untreated and SA-treated HEK293 cells.

Heme was quantified as described (164). Briefly, HEK293 cells were seeded in complete DMEM media at an initial confluency of 10% and incubated at 37 ºC for 48 hrs. Media for SA-treated cells was replaced by DMEM supplemented with 10% heme-depleted FBS and 0.5 mM SA. Heme depletion of serum performed as described (165). Media for untreated cells was replaced by complete media (supplemented with 10% regular FBS) and allowed to seed at 37 ºC for 24 hrs. Cells were harvested by scrapping and counted using an automated TC10 cell counter (Bio-Rad). Then, $2.5x10^4$ cells per condition were treated with 20 mM oxalic acid and incubated at 4 ºC overnight in the dark. An equal volume of 2 M oxalic acid was added to the cell suspensions. Samples were split, with half incubated at 95 ºC for 30 min and half incubated at room temperature for 30 min. Samples were centrifuged at 21,000*g* for 2 min, and 200 μL of each was transferred to a black Greiner Bio-one flat bottom fluorescence plate, Porphyrin fluorescence (ex: 400 nm, em: 620 nm) was recorded on a Synergy Mx multi-modal plate reader. Heme concentration was calculated from a standard curve prepared by diluting a 0.1 μM hemin chloride stock solution in DMSO and treated as cell suspensions above. To calculate heme concentration, the fluorescence of the unboiled samples is taken as the background level of protoporphyrin IX and it is subtracted from the fluorescence of the boiled sample, which is used as the free

base porphyrin produced upon the release of the heme iron. Using this method, our data suggest SA-treatment of HEK293 cells results in a 7-fold decrease in the total cellular heme concentration.

4.5.7. Hemin agarose binding.

HEK293 cells were seeded onto a 6-well plate at an initial confluency of 20% in Dulbecco's modified Eagle's medium (DMEM) with 10% Fetal Bovine Serum (FBS) and allowed to seed for 48 hrs at 37 °C. Media was then replaced for DMEM with 10% heme-depleted FBS supplemented with 0.5 mM succinyl acetone (for SA-treated cells). For untreated cells, media was changed for DMEM in 10% regular FBS. Both treated and untreated samples were allowed to incubate at 37 °C for 24 hrs. Cells were then collected by scrapping and lysed using 1.5 mm zirconium Beads (Benchmark). Lysates were quantified by Bradford assay. In the meantime, hemin agarose beads and sepharose beads were equilibrated 3 times by centrifugation with Lysis buffer (0.1% Triton X-100, 10 mM Sodium Phosphate, 50 mM KCl, 5 mM EDTA, pH 7.5, 1X protease arrest, RNasin RNase Inhibitor (Promega)). 100 □L of beads (50 μL bed volume) were used per biological replicate. After bead equilibration, each lysate was divided into two and 10 μg were loaded to hemin agarose and 10 μg to sepharose beads. Mixtures were allowed to bind for 60 min, rotating at 20 rpm at room temperature. Then, three washes were performed using Lysis buffer and supernatants were discarded. Each wash consisted of 10 min incubation at room temperature with 20 rpm rotation followed by centrifugation at 700*g* for 5 min. Bead bound fractions were eluted by a 15 min incubation at room temperature with 20 rpm rotation in 50 μL of 1M imidazole in Lysis buffer followed by centrifugation at max. speed for 2 min

and supernatants were collected. RNA was then extracted from eluted fractions with TRIZOL using the manufacturer's protocol. For the PhenDC3 titration in HEK293 cells experiment, the same protocol was followed with the difference that cells were allowed to seed for 24 hrs (20% initial confluency) and then PhenDC3 was added in increasing concentrations (5 μM, 10 μM, 20 μM). DMSO carrier treatment was performed the same way but with equivalent DMSO volumes. Cells were left at 37 °C for 48 hrs and collected and lysed as described above.

4.5.8. RT-qPCR.

The sets of primers used can be found in Table C.2. Luna Universal One-Step RT-qPCR kit (New England Biolabs) was used following the manufacturer's protocol. Fold enrichments were calculated by comparing the C(t) values obtained from RNAs extracted from hemin agarose to RNAs extracted from sepharose beads. Three biological replicates were performed for all the RT-qPCR experiments. For BioTASQ experiments, fold enrichments were calculated by comparing the C(t) values obtained from the lysates containing BioTASQ + beads with those containing beads only.

4.5.9. Heme Bioavailability Assay using the HS1 Sensor.

HEK293 cells were plated and transfected in polystyrene coated sterile 6 well plates (Grenier) for flow cytometry. The cells were plated in basal growth medium Dulbecco's modified eagle medium (DMEM) containing 10% fetal bovine serum. At 30% confluency cells were transfected with the heme sensor plasmid pEF52α-hHS1 using Lipofectamine LTX according to the manufacturer's protocols. After 48 hours of treatment with transfection reagents, cells treated with PhenDC3 (1 mM stock) in fresh DMEM 10% FBS

for 24 hours prior to harvesting. Heme depleted cells were treated with 500 μM succinylacetone (SA) in DMEM containing 10% heme depleted FBS for 72 hours prior to harvesting. Heme sufficient cells were treated with 350 μM 5-aminolevulinic acid (ALA) in DMEM 10% FBS for 24 hours. Cells were harvested in 1X PBS for flow analysis. Flow cytometric measurements were performed using a BD FACS Aria Ill Cell Sorter equipped with an argon laser (ex 488 nm) and yellow-green laser (ex 561 nm). EGFP was excited using the argon laser and was measured using a 530/30 nm bandpass filter, mKATE2 was excited using the yellow- green laser and was measured using a 610/20 nm bandpass filter. Data evaluation was conducted using FlowJo v10.4.2 software. Single cells used in the analysis were selected for by first gating for forward scatter (FSC) and side scatter (SSC), consistent with intact cells, and then by gating for cells with mKATE2 fluorescence intensities above background were selected. The fraction of sensor bound to heme may be quantified according to the following equation (60):

$$\%\,\mathrm{Bound} = 100 \times \left(\frac{\mathrm{R} - \mathrm{R_{min}}}{\mathrm{R_{max}} - \mathrm{R_{min}}}\right)$$

where R is the median eGFP/mKATE2 fluorescence ratio in regular media and R_{min} and R_{max} are the median sensor ratios when the sensor is depleted of heme or saturated with heme. R_{min} and R_{max} values are derived from cells cultured in heme deficient media conditioned with succinylacetone (HD+SA) or in media conditioned with ALA (60). The plot in Figure 4D was obtained by fitting the median sensor ratios in Figure 4C to the following 1-site binding model (60, 166):

$$\text{Ratio} = \text{initial ratio} + \Delta\text{ratio} \times \left(\frac{x}{K_d + x}\right)$$

where x is the independent variable, [PhenDC3].

4.5.10. BG4 purification.

pSANG10-3F-BG4 was a gift from Shankar Balasubramanian (Addgene plasmid # 55756; http://n2t.net/addgene:55756; RRID:Addgene_55756). BL21 cells transformed with this plasmid were grown in room temperature and induced overnight with 0.1mM IPTG. Cells were pelleted, then resuspended in xTtractor (Takara) supplemented with Protease arrest (G-protein), lysozyme and DNase I. Sonicated cell lysate was combined with Ni-NTA resin (Invitrogen) and purified via the his-tag. BG4 was further purified by FPLC using a Superdex75 size exclusion column (GE Healthcare).

4.5.11. Immunofluorescence.

Immunofluorescence was performed by standard protocols. HEK293 cells were seeded onto Poly-L-lysine coated cover glass two days before the experiment and fixed in 4% formaldehyde for 15 min. Cells were permeabilized with 0.1% Triton X-100 for 3 min and blocked with 5% donkey serum (Jackson ImmunoResearch), followed by incubation with antibodies for 1 hr at room temperature or overnight at 4 °C. Antibodies used here are: BG4, rabbit anti-FLAG (Cell Signaling Technology, 14793S), mouse anti-L19 (Santa Cruz Biotechnology, sc-100830), mouse anti-rRNA (Santa Cruz Biotechnology, sc-33678), mouse anti-Calnexin (Santa Cruz Biotechnology, sc-23954), Alexa Fluor 488 conjugated donkey anti-rabbit (Jackson ImmunoResearch, 711-545-152), Rhodamine Red-X conjugated donkey anti-mouse (Jackson ImmunoResearch, 715-295150). After staining

cells were carefully washed with DPBS supplemented with 0.1% tween-20. Nuclear DNA was stained with 4',6-diamidino-2-phenylindole (DAPI). Images were acquired with a Zeiss 700 Laser Scanning Confocal Microscope. PhenDC3 treatment consisted of incubation at 37 °C overnight at 10 μM final PhenDC3 concentration prior to cell fixation. Determination of colocalization ratios was performed as described in Zen software (Zeiss). No primary antibody controls as well as RNase A and PhenDC3 treated images are reported in Figures C.4 and C.5. The "Colocalization" image in Figure 4.2A,B is showing the G4 signal that colocalizes with L19 and with the ER (yellow pixels) and the one that does not colocalize (green pixels). "L19", "ER", and "BG4" images only present their respective fluorescence signals.

4.5.12. BioTASQ capture of cellular RNAs.

BioTASQ experiments followed published protocols *in vitro* (17) and *in vivo* (27). Briefly, HEK293 cells were seeded onto a 6-well plate at 20% confluency and allowed to incubate at 37 °C for 48 hrs. Cells were then crosslinked with 1% paraformaldehyde/PBS for 5 min at room temperature. Crosslinking was stopped by incubating cells with 0.125 M glycine for 5 min at room temperature. Cells were harvested by scrapping and resuspended in Lysis Buffer (200 mM KCl, 25 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.5 mM DTT, 1% Triton X-100, RNasin RNAse Inhibitor, 1X protease arrest). Cells were lysed by sonication (30% amplitude, 10 sec. on and off intervals, 2 min sonication time). The lysate was then split: BioTASQ was added at a final concentration of 100 μM to one of the samples, the other one was left untreated. Lysates were incubated at 4 °C overnight with gentle rotation. Sera-Mag magnetic streptavidin-coated beads (GE Healthcare) were washed three times

with wash buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M KCl). Each wash was followed by centrifugation at 3,500 rpm for 5 min at 4 °C. Beads were then treated with Buffer 1 (0.1 M NaOH, 0.05 M KCl in RNase/DNase-free water) two times at room temperature for 2 min and then centrifuged at 3,500 rpm at 4 °C, 5 min, and washed with Buffer 2 (0.1 M KCl in RNase/DNase-free water). Lastly, to block, beads were treated with 1 μg/mL BSA and 1 μg/mL yeast tRNA and allowed to incubate at 4 °C overnight with gentle rotation.

After incubation overnight with BioTASQ, cell lysates were treated with 1% BSA for 1 hr at 4 °C. Washed magnetic beads were added to the lysates (20 μg beads /sample) and allowed to mix with gentle rotation at 4 °C for 1 hr. Beads were then washed three times with Lysis buffer for 5 min and then crosslinking was reversed by incubating the beads at 70 °C for 1 hr. Finally, TRIZOL was used to extract RNAs, for analysis by RT-qPCR.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1. Conclusions

Historically, the ribosome has been approached as an essential but passive macromolecular machine that translates the nucleic acid code carried by mRNAs into proteins. Therefore, the canonical function of ribosomes is protein synthesis. However, across phylogeny not all ribosomes are the same: eukaryotic ribosomes present expansion segments that, in cases like *Homo sapiens*, extend for hundreds of Angstroms from the ribosomal surface in the form of rRNA tentacles. Bacterial ribosomes lack these rRNA tentacles and hence are nearly twice as small as human ribosomes. As explained in Chapter 1 of this thesis, the length of rRNA expansion segments correlates well with species complexity.

The work described in this thesis has been focused on the discovery of non-canonical G-quadruplex structures in specific rRNA tentacles of the human ribosome and their involvement in non-canonical ribosomal functions. This work is described in detail in chapters 2, 3, and 4. Taken together, this work describes (i) the identification of the sequences with G-quadruplex potential on human 28S and 18S rRNAs, (ii) the formation of these rRNA structures *in vitro* and *in vivo*, (iii) their conservation across phylogeny, and (iv) their physiological roles. We have data supporting their role in inter-molecular association of human ribosomes and in recruiting specific proteins to the ribosome (Chapter 2). Our *in vivo* experiments suggest that a small fraction of specialized ribosomes

form G-quadruplexes (~5%) and that most extra-nuclear G-quadruplexes are associated with ribosomes (~83%), indicating the RNA G-quadruplexome is ribosome-centric. The work described in Chapter 4 provides strong evidence that one of the cellular roles of rRNA G-quadruplexes is to regulate heme bioavailability. The ribosome as a hub for heme is consistent with its role as an abundant and versatile sink for ions and small molecules, including antibiotics, platinum-based drugs, certain metabolites, and metal cations such as Mg^{2+}, Ca^{2+}, Mn^{2+}, and Fe^{2+}, K^{+}.

Overall, the results presented in this thesis represent the first report on the *in vitro* and *in vivo* formation of non-canonical G-quadruplex structures on any ribosomal RNA and provide strong evidence on the involvement of these structures in a non-canonical ribosomal function: heme appropriation and regulation. Prior to this thesis work, ribosomal RNA had never been reported to contain sequences with G-quadruplex potential. Taking into account that the vast majority of RNA is cells is ribosomal (>80%), our results imply that most RNA G-quadruplexes are ribosomal. In addition, the work described in Chapter 4 has major implications for the physiology of G-quadruplex-heme interactions. Current studies on the identification of heme trafficking factors are mainly centered on proteinaceous factors and largely ignore contributions from nucleic acids. Our results provide strong evidence that human ribosomes, and G-quadruplex containing rRNAs in particular, appropriate and buffer cytosolic heme. Given that the first opportunity for protein hemylation occurs during or just after translation, rRNA and rProteins may be critical for directing labile heme to newly synthesized proteins.

5.2. Future directions: investigating the roles of rProteins in heme regulation

The work performed in this thesis has been solely performed on rRNA. However, ribosomes are ribonucleoproteins that contain over 50 rProteins. Therefore, it is conceivable that rProteins are also involved in regulating heme bioavailability. Towards this end, we have started performing experiments with rProteins. By treating human ribosomes purified from human cells with protease K we observed a significant decrease (~30%) on their ability to bind heme *in vitro*, strongly suggesting specific rProteins also contain heme-binding properties (Figure 5.1A).

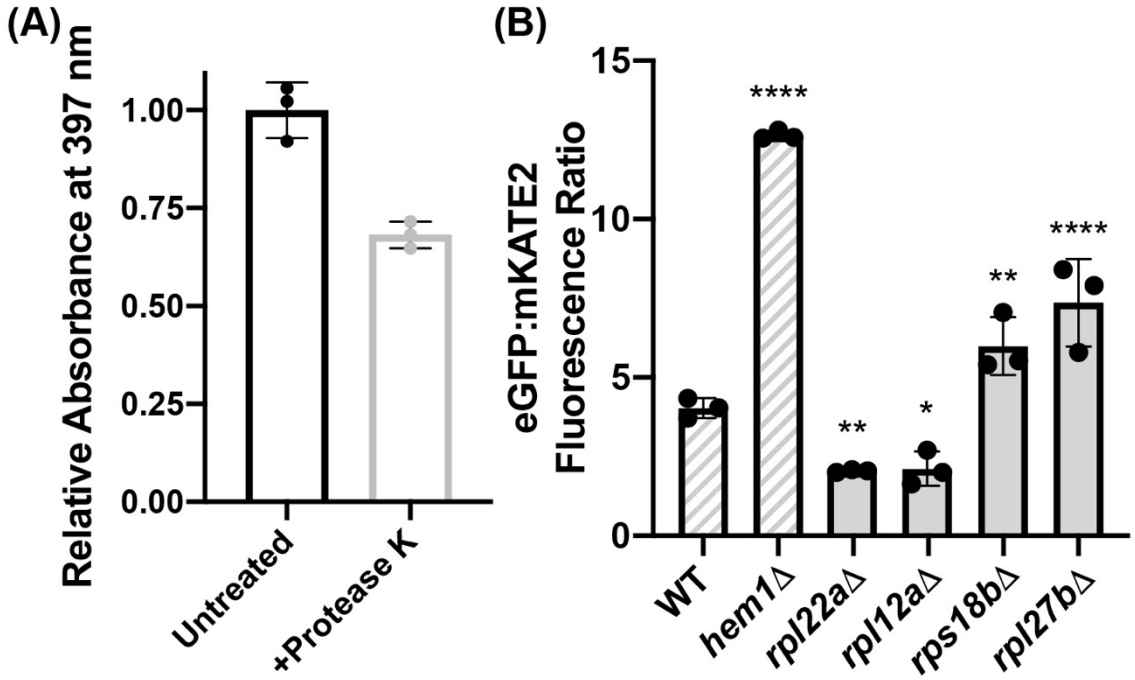


Figure 5.1. Role of rProteins on heme binding and heme bioavailability. (A) Protease K digestion of human ribosomes *in vitro*. (B) L22a, L12a, S18b and L27b regulate cytosolic heme bioavailability. Fluorimetric determination of eGFP/mKATE2 fluorescence ratios of the indicated yeast strains expressing heme sensor HS1-M7A. Statistical significance relative to WT strain is represented by asterisks using ordinary one-way ANOVA with Dunnett's post-hoc test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

In addition, previous experiments performed both at the genome- and proteome-wide levels in the laboratory of professor Dr. Amit Reddi at the Georgia Institute of Technology suggested several rProteins presented potential heme-binding properties. Some of this previous work consisted of transforming the heme sensor HS1 described in detail in Chapter 4 to all available yeast knockout strains from the yeast genome library to identify candidate genes involved in heme bioavailability regulation. Several rProtein strains presented promising results. We screened for these candidate rProtein knockouts and confirmed that 4 ΔrProtein strains exhibit different bioavailable levels of heme with respect to WT strains (Figure 5.1B). Specifically, deletion of rProteins L22 (eL22) and L12 (uL11) results in an increase in the bioavailable levels of cytosolic labile heme, whereas deletion of rProteins L27 (eL27) and S18 (uS13) results in a decrease in the bioavailable levels of cytosolic labile heme. These results represent the first link between rProteins and heme bioavailability and, together with our rRNA-heme results, support a model in which ribosomes act as central hubs for heme metabolism through both rRNAs and rProteins. The most logical next step would be to study the roles of ribosomes in heme trafficking throughout cellular compartments and the potential involvement of ribosomes in hemylation of nascent proteins.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Table A.1. DNA and RNA sequences encoding *Homo sapiens* ES7, GQES7-a, GQES7-b, *mt*ES7-a and *mt*ES7-b.

Gene	Sequence (5' to 3')
H. sapiens ES7	GAATTC*TAATACGACTCACTATAGGG*CGGGTGGGGTCCGCGCAGT CCGCCCGGAGGATTCAACCCGGCGGCGGGTCCGGCCGTGTCGG CGGCCCGGCGGATCTTTCCCGCCCCCCGTTCCTCCCGACCCCTC CACCCGCCCTCCCTTCCCCCGCCGCCCCTCCTCCTCCTCCCCGGA GGGGGCGGGCTCCGGCGGGTGCGGGGGTGGGCGGGCGGGGCC GGGGGTGGGGTCGGCGGGGGACCGTCCCCCGACCGGCGACCGG CCGCCGCCGGGCGCATTTCCACCGCGGCGGTGCGCCGCGACCG GCTCCGGGACGGCTGGGAAGGCCCGGCGGGGAAGGTGGCTCGG GGGGCCCCGTCCGTCCGTCCGTCCGTCCTCCTCCTCCCCCGTCTC CGCCCCCCCGGCCCCGCGTCCTCCCTCGGGAGGGCGCGCGGGTC GGGGCGGCGGCGGCGGCGGCGGTGGCGGCGGCGGCGGCGGCG GCGGGACCGAAACCCCCCCCGAGTGTTACAGCCCCCCCGGCAG CAGCACTCGCCGAATCCCGGGGCCGAGGGAGCGAGACCCGTCG CCGCGCTCTCCCCCCTCCCGGCGCCCACCCCCGCGGGGAATCCC CCGCGAGGGGGGTCTCCCCCGCGGGGGCGCGCCGGCGTCTCCT CGTGGGGGGGCCGGGCCACCCCTCCCACGGCGCGACCGCTCTC CCACCCCTCCTCCCCGCGCCCCCGCCCCGGCGACGGGGGGGGT GCCGCGCGCGGGTCGGGGGGCGGGGCGGACTGTCCCCAGTGCG CCCCGGGCGGGTCGCGCCGTCGGGCCCGGGGGAGGTTCTCTCG GGGCCACGCGCGCGTCCCCCGAAGAGGGGGACGGCGGAGCGA GCGCACGGGGTCGGCGGCGACGTCGGCTACCCACCCG**TCGATC CGGTTCGCCGGATCCAAATCGGGCTTCGGTCCGGTTC**AAGC TT
GQES7-a	GAATTC*TAATACGACTCACTATAGGG*CGGA**GGGGG**C**GGG**CTCCG GC**GGG**TGC**GGGGG**T**GGG**C**GGG**C**GGGG**CC**GGGGG**T**GGGG**TC GGC**GGGGG**ACCGAAGCTT
*mt*ES7-a	CGGA**GAGAG**C**AGU**CUCCGGC**AGA**UGC**GAUAG**U**GAU**C**AGA**C**G UUG**CC**GUAUG**U**GAAG**UCGGC**GAAUG**ACCG
GQES7-b	GAATTC*TAATACGACTCACTATAGGG*CCTC**GGG**A**GGG**CGCGC**GG G**TC**GGGG**CGGCAAGCTT
*mt*ES7-s	CCUC**AGU**A**UUG**CGCGC**AAG**UC**GAUG**CGGC

- The T7 promoter region is in italics, restriction sites are underscored (EcoRI 5', HindIII 3'), and the Weeks cassette (167) (used to performed SHAPE footprinting, data not published) is in bold. G-tracts and G-tract mutations are underscored and in bold.

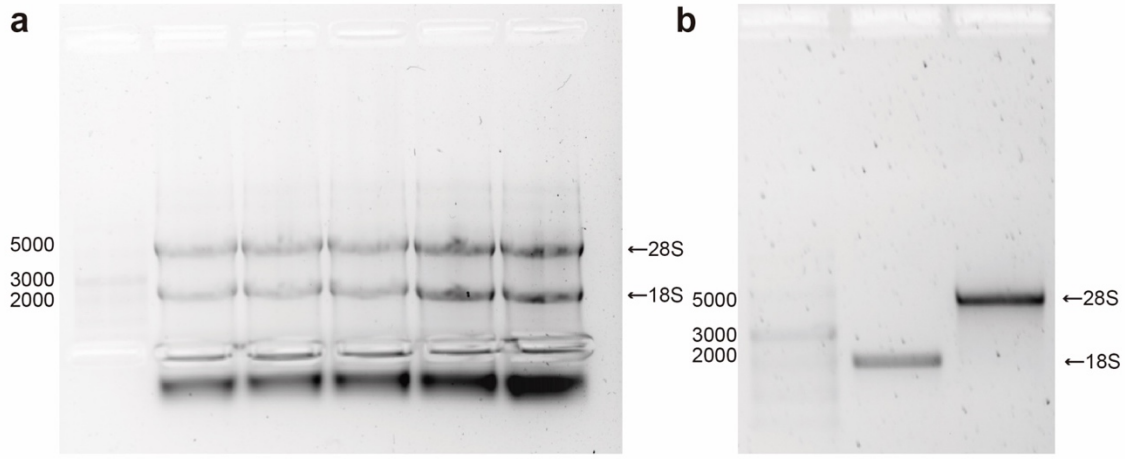


Figure A.1. Total RNA was extracted from HEK293T cells using TRI reagent. a) Total RNA loaded on a 1% agarose gel. When the 18S and 28S rRNAs passed through the second set of wells (in the middle of the gel) they were extracted by pipetting and further purified by ammonium acetate/ethanol precipitation; b) The final integrity and purity of the 18S and 28S rRNAs are indicated on an agarose gel.

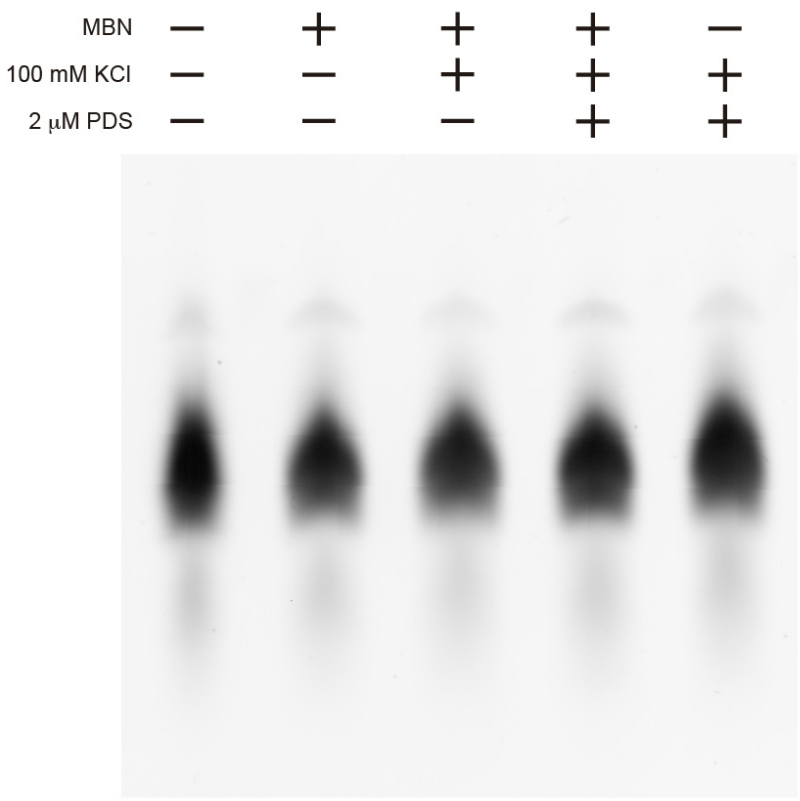


Figure A.2. tRNA from baker's yeast was folded under the indicated conditions followed by incubation with MBN before denaturing gel electrophoresis.

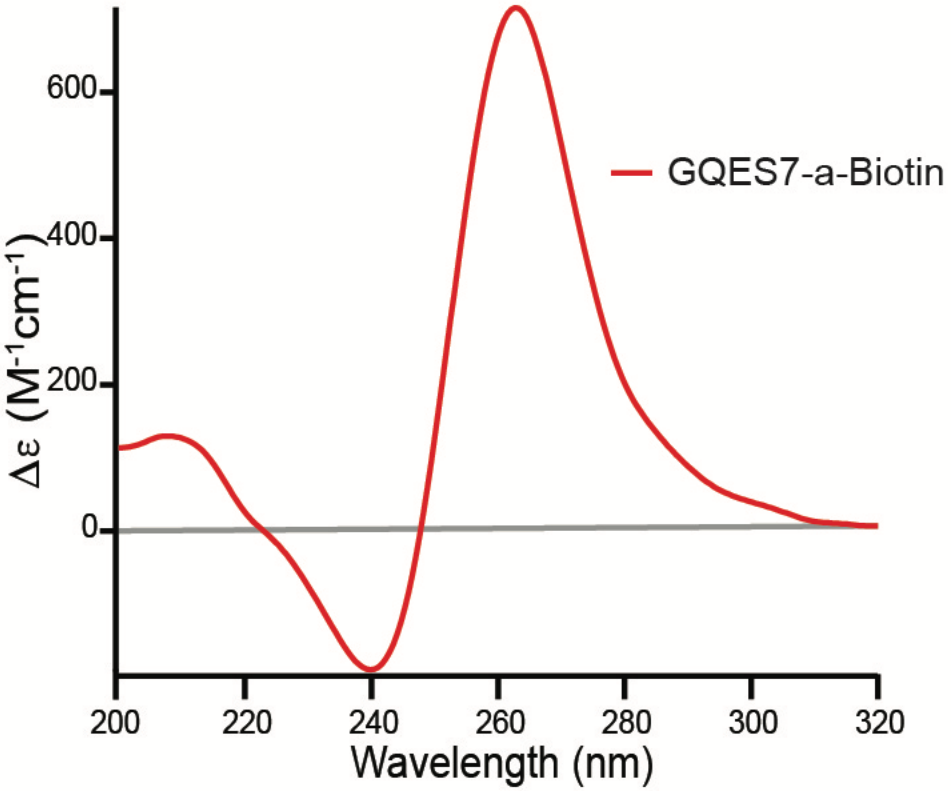


Figure A.3. CD spectrum of GQES7-a-Biotin confirms that the biotinylation of GQES7-a does not disrupt G-quadruplexes.

Table A.2. Distribution of G-quadruplex-forming sequences across LSU ES tentacles of Eukaryotes from the SEREB database. G-scores were calculated with QGRS Mapper (72).

Species	ES tentacle	Position	Sequence	G-score
H. sapiens	ES7a	587-612	GGGGGCGGGCUCCGGCGGGUGCGGG	38
		615-632	GGGCGGGCGGGGCCGGG	41
	ES7b	829-849	GGGAGGGCGCGCGGGUCGGG	38

	ES27	2896-2918	GGGCUGGGUCGGUCGGGCUGGG	38
	ES27a	3110-3140	GGGGGCGGCGGCGGGGGGAGAAGGGUCGGG	38
P. troglodytes	ES7a	583-608	GGGGGCGGGCUCCGGCGGGUGCGGG	38
		611-628	GGGCGGGCGGGGCCGGG	41
	ES7b	816-836	GGGAGGGCGCGCGGGUCGGG	38
	ES27	2890-2912	GGGCUGGGUCGGUUGGGCUGGG	38
	ES27b	3255-3283	GGGCAGGGAGUGAUUGGGUGUCGGUGGG	37
	ES30	3829-3853	GGGGGUGGGGUGGGGUCCUGUGGG	39
R. norvegicus	ES7a	633-662	GGGCGCGGGGUGUGGUGGGGGCGCGCGGG	39
		663-687	GGGGCCGGGGGUGGGGUCGGCGGG	40
	ES7f	1041-1067	GGGGGUCCGGGGGCCCGGGGGGCGGG	42
	ES27	2736-2758	GGGCUGGGUCGGUCGGGCUGGG	38
	ES27a	2935-2952	GGGGGGGACGGGGCGGG	41

	ES27b	3214-3240	GGGGUCGGGGGUUCCCGGGGUUCGGG	41
M. musculus	ES7a	627-646	GGGCGGGGCCGGGGGUGGG	41
	ES7d	907-926	GGGCGGGCGUGGGGGUGGG	40
	ES27	2663-2685	GGGCUGGGUCGGUCGGGCUGGG	38
M. domestica	ES7a	572-594	GGGGCGGCGGGCGAGGGCCGGG	40
	ES7b	729-759	GGGGGUGGCGGGGGCCAGAGGGGGCCCGGG	41
	ES7d	995-1019	GGGCCCGGGUUCCCGGGGGACGGG	40
	ES27	2726-2748	GGGCUGGGUCGGUCGGGCUGGG	38
	ES27a	2932-2949	GGGCGGGGGGGGACGGG	41
		2949-2978	GGGGGCGCGCGGGGGGUCGGGGGCGCGGG	41
		3045-3061	GGGGGGGCAGGGGGGG	41
		3123-3149	GGGGUCCGGGGUUCGGGCCCGGCGGG	40
	ES27b	3242-3266	GGGCCCGGGGAGCCCGGGGGCGGG	40

	ES39a	4728-4749	GGGGGGGCCUCGGGGGCCGGG	39
G. gallus	ES7a	557-577	GGGGCGGGGCGGGCCCAGGG	40
		577-594	GGGGCGGGCGGGCCGGG	41
	ES7d	933-953	GGGGCGCGGGGGCGGGGGGG	40
	ES27	2651-2673	GGGCUGGGUCGGUCGGGCUGGG	38
	ES27a	2755-2774	GGGGCGCCGGGGGGGGGGG	39
	ES27b	3018-3040	GGGCGGGGCGGUCCCGGGCGGG	36
A. carolinensis	ES7a	537-559	GGGGAAGGGGUUCCCGGGAGGG	38
	ES27	2365-2387	GGGCUGGGCCGGUCGGGCUGGG	38
X. laevis	ES7d	793-814	GGGGCGGGGAAGGGGGAAGGG	42
	ES9	1085-1111	GGGGGCGGGGGGGGGGCGCCGGCGGG	42
	ES27	2412-2434	GGGCUGGGUCGGUCGGGCUGGG	38
	ES27a	2532-2555	GGGGGGCCGCGGGGGCGGGGGGG	39

L. chalumnae	ES27	2260-2282	GGGCUGGGUCGGUCGGGCCGGG	38
D. rerio	ES7d	751-773	GGGGGGGGCUCGGAGGGACGGG	38
	ES27a	2514-2537	GGGCUCGGGGUGGGUGUUGCGGG	38
		2566-2587	GGGUCGGGGCCUGCGGGGGGG	38
	ES27b	2695-2715	GGGGGGAGGGCACGGGCGGG	40
	ES30	3122-3153	GGGAGGCCCUGGGGCCCUCGGGCCCAGCGGG	21

Table A.3. Species used in phylogenetic analysis of ES7.

Species names	**Species alias**	**Common name**	**Taxonomy identifier**	**Order**	**Comment**
Homo sapiens	*H. sapiens*	*Human*	9606	Primates	SEREB Database (64)
Pan troglodytes	*P. troglodytes*	*Common chimpanzee*	9598	Primates	SEREB
Rattus norvegicus	*R. norvegicus*	*Common rat*	10116	Rodentia	SEREB
Mus musculus	*M. musculus*	*House mouse*	10090	Rodentia	SEREB
Felis catus	*F. catus*	*Domestic cat*	9685	Carnivora	Predicted
Equus caballus	*E. caballus*	*Horse*	9798	Perissodactyla	Partial
Dasypus novemcinctus	*DASNO*	*Nine-banded armadillo*	9361	Cingulata	Predicted; partial
Monodelphis domestica	*M. domestica*	*Gray short-tailed opossum*	13616	Didelphimorphia	SEREB
Caloenas nicobarica	*C. nicobarica*	*Nicobar pigeon*	187106	Columbiformes	

Gallus gallus	*G. gallus*	*Chicken*	9031	Galliformes	SEREB
Apteryx australis	*A. australis*	*Common kiwi*	8822	Apterygiformes	Predicted; partial
Crocodylus porosus	*C. porosus*	*Saltwater crocodile*	8502	Crocodilia	Predicted
Chrysemys species	*Chrysemys sp.*	*Painted turtle*	307036	Testudines	Predicted
Anolis carolinensis	*A. carolinensis*	*American anole*	28377	Squamata	SEREB
Xenopus laevis	*X. laevis*	*African clawed frog*	8355	Anura	SEREB
Latimeria chalumnae	*L. chalumnae*	*Coelacanth*	7897	Coelacanthiformes	SEREB
Danio rerio	*D. rerio*	*Zebrafish*	7955	Cypriniformes	SEREB

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

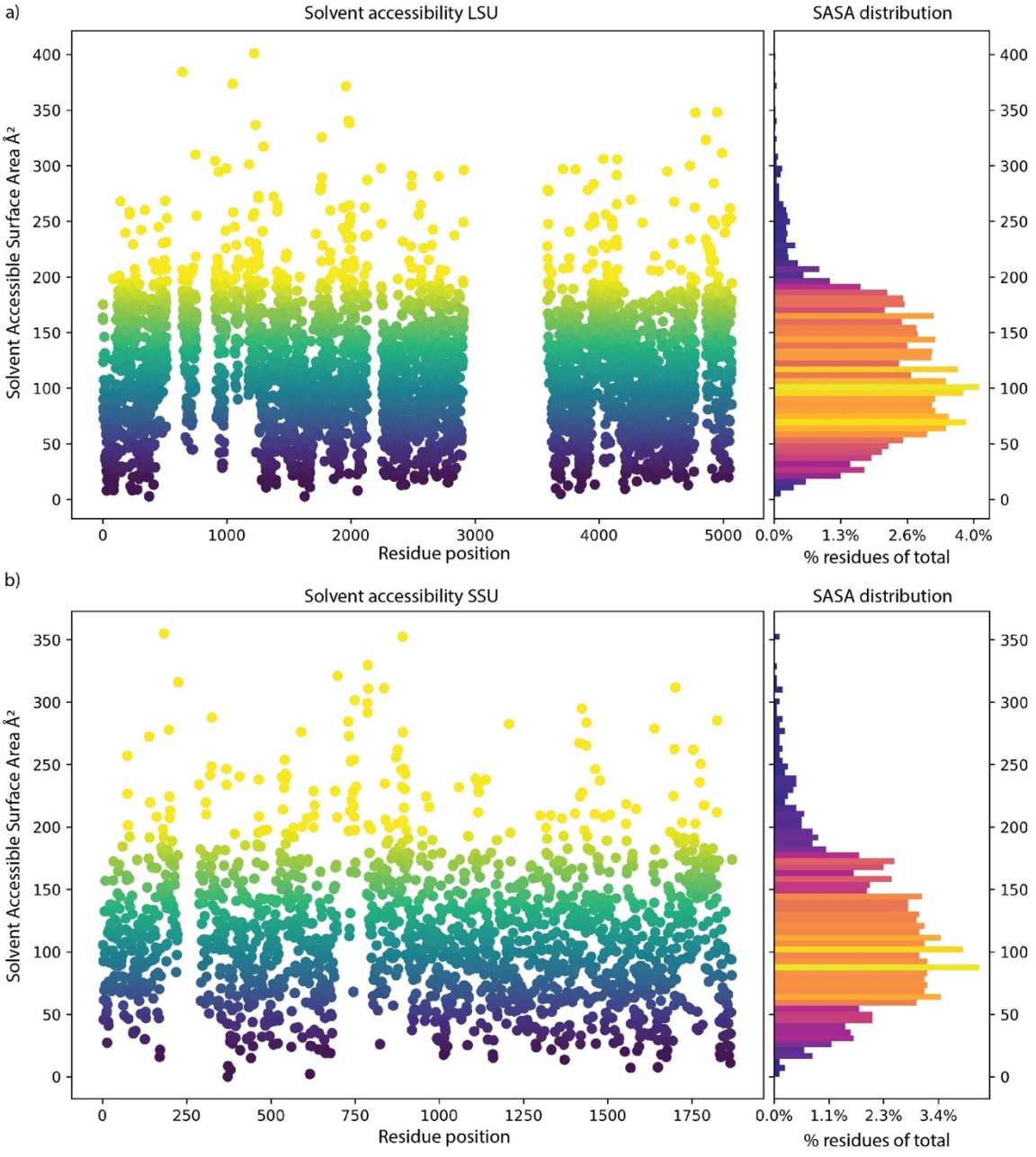


Figure B.1. Distribution of solvent-accessible surface area for each residue of the a) human LSU rRNA and b) human SSU rRNA.

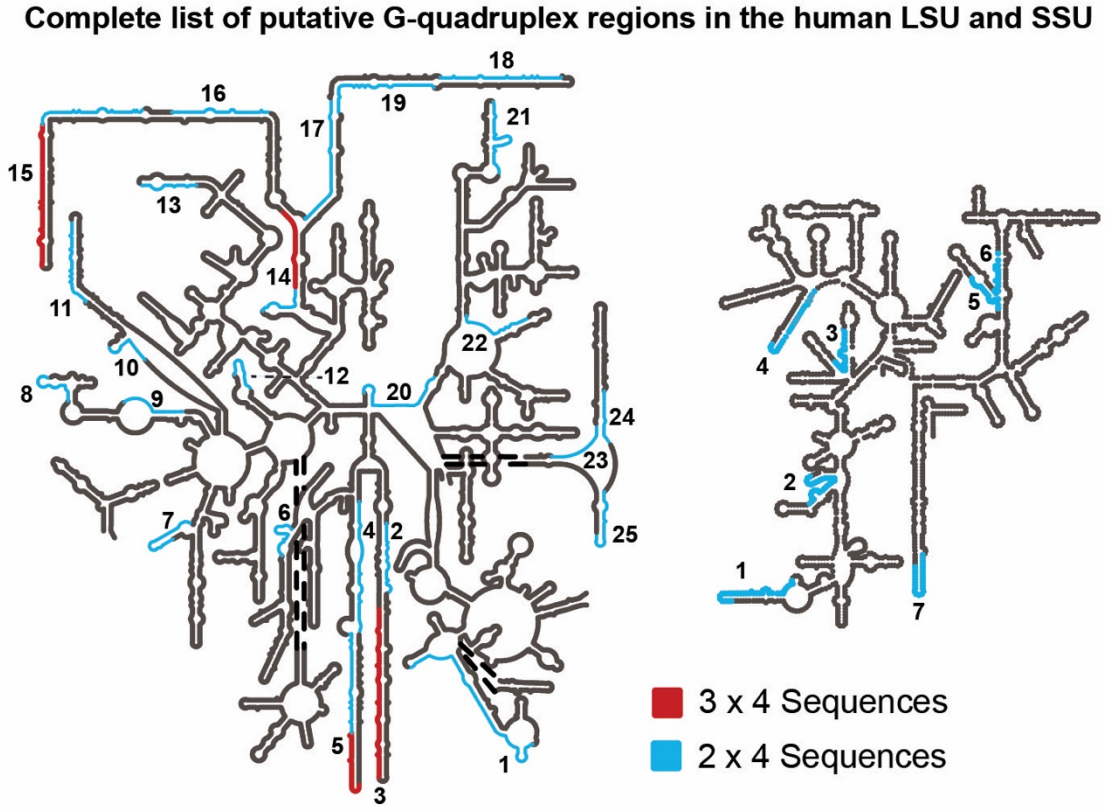


Figure B.2. Secondary structures of the human LSU and SSU rRNAs with all identified 3 x 4 (red) and 2 x 4 (blue) G-quadruplex regions.

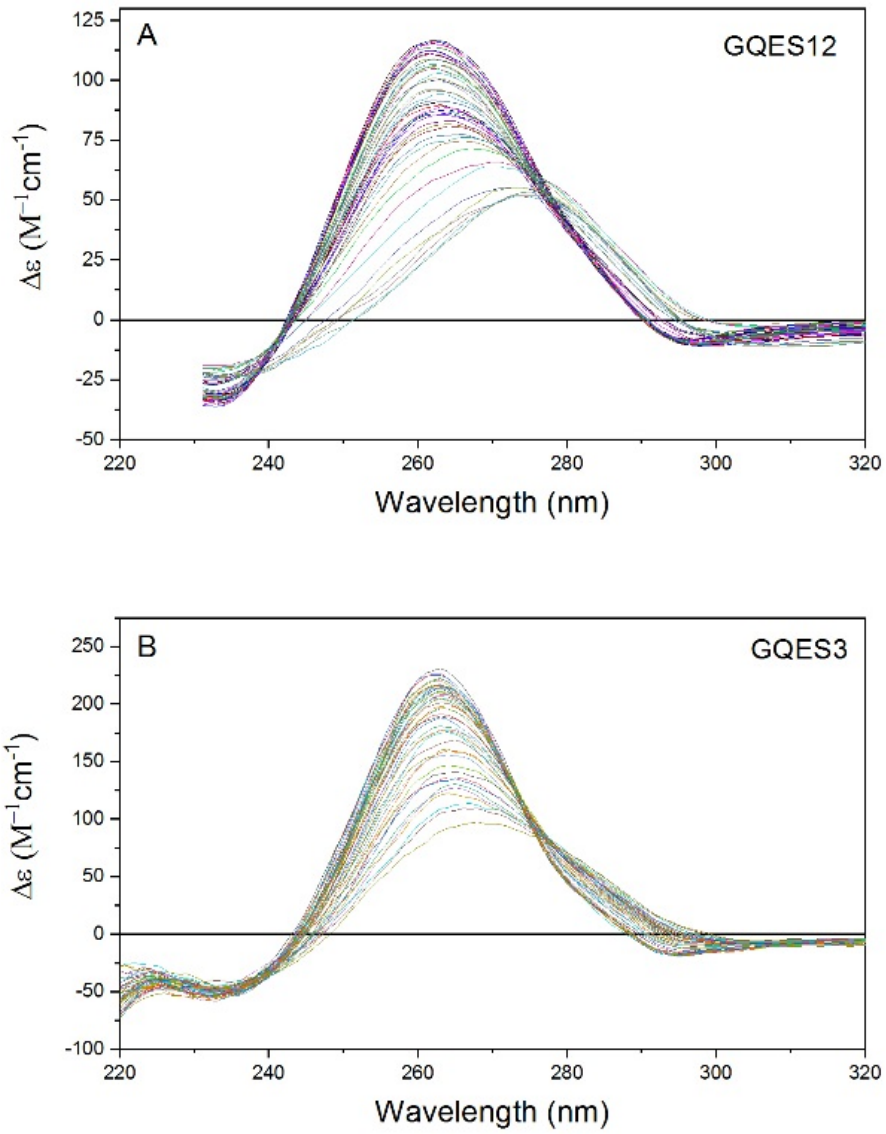


Figure B.3. Thermal denaturation of GQes12 (A) and GQes3 (B) as monitored by changes in circular dichroism. Panels A and B show the decrease in amplitudes in CD spectra as temperature is increased.

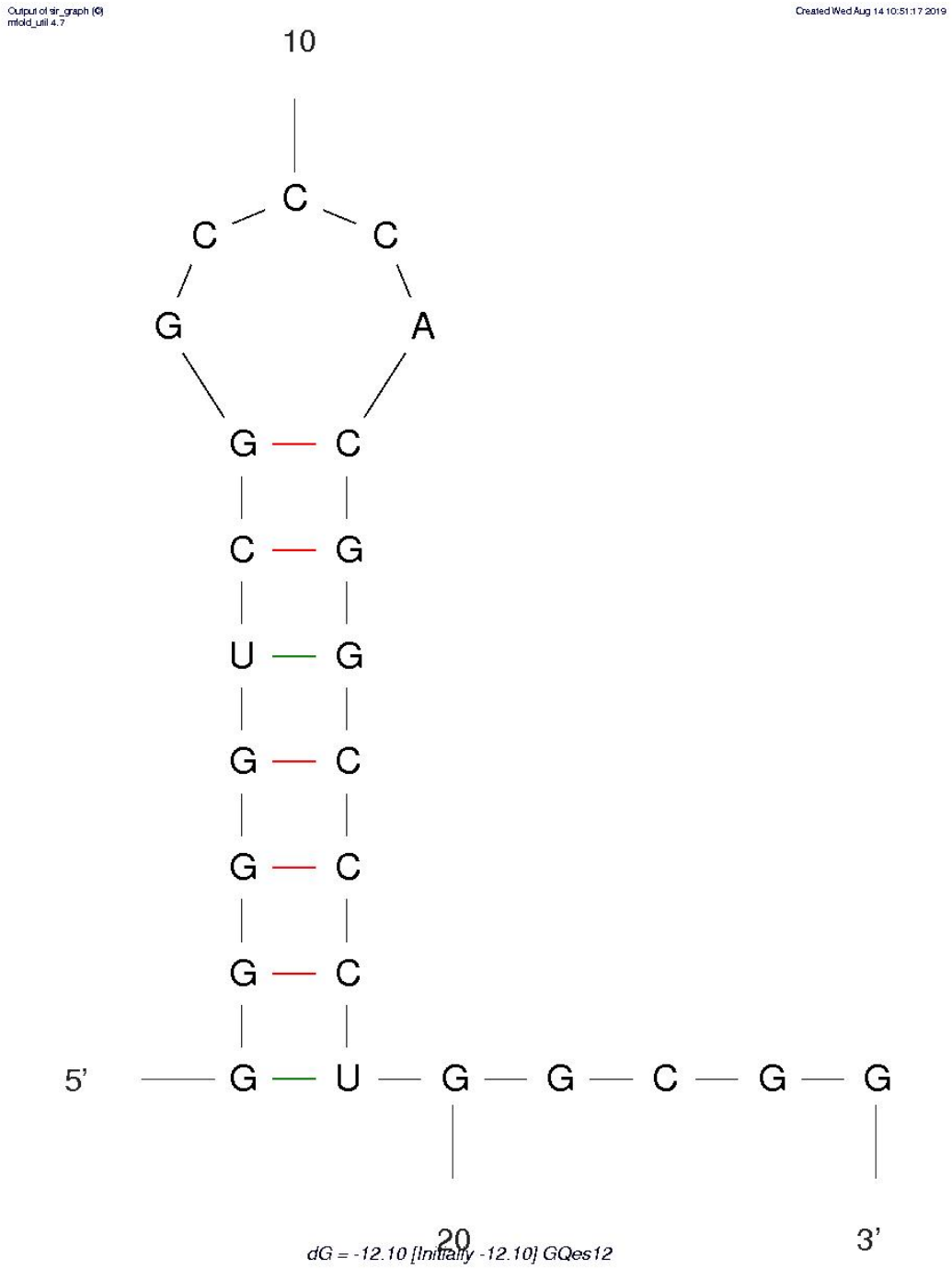


Figure B.4. Possible hairpin structure of GQes12 calculated using the mFold RNA server.

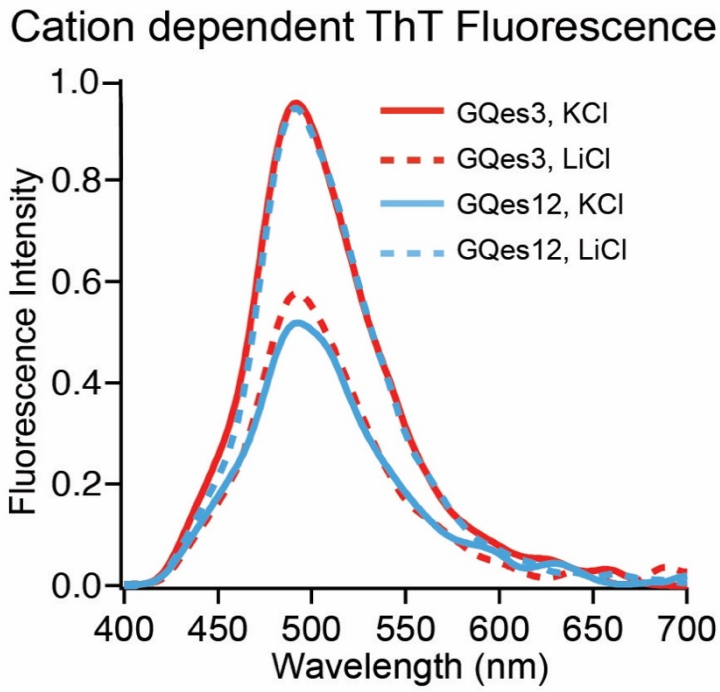


Figure B.5. Relative ThT fluorescence spectra of the GQes3 and GQes12 rRNA oligomers annealed in the presence of potassium or lithium ions.

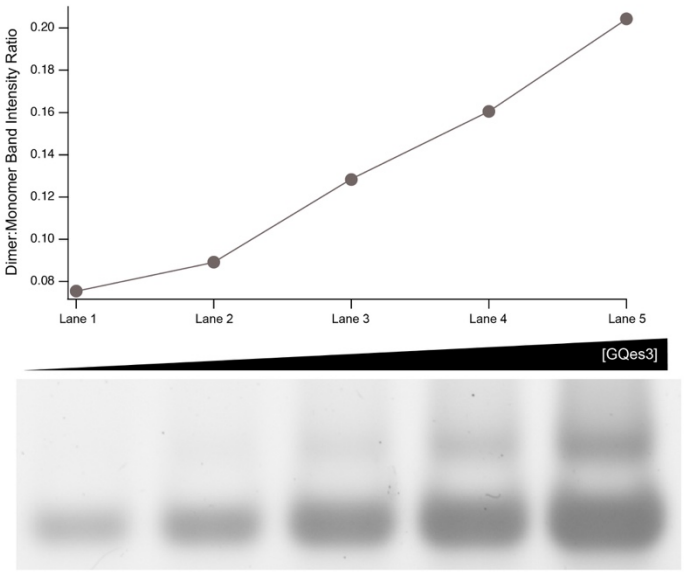


Figure B.6. Increasing concentrations of GQes3 result in a shift from the monomer to the dimer species. Band intensities were quantified using ImageJ and the ratio of dimer to monomer was plotted. The increase in the ratio indicates that the monomer-dimer equilibrium is shifted to the latter upon the increase in the RNA concentration. RNA was resolved on a 6% Native PAGE.

Table B.1. DNA and RNA sequences encoding RNAs used.

Gene	Sequence (5' to 3')
GQES7-a	GAATTC*TAATACGACTCACTATAGGG*CGGA**GGGGG**C**GGG** CTCCGGC**GGG**TGC**GGGGG**T**GGG**C**GGG**C**GGGG**CC**GGG** **GG**T**GGGG**TCGGC**GGGGG**ACCGAAGCTT
GQES7-b	GAATTC*TAATACGACTCACTATAGGG*CCTC**GGG**A**GGG**CGC GC**GGG**TC**GGGG**CGGCAAGCTT
GQes3	**GG**CCCC**GG**CC**GGGGGG**C**GGG**CGCC**GG**
*mut*es3	**AA**CCCC**AA**CC**GAAAAG**C**GAA**GCC**AA**
GQes12	**GGGG**UC**GG**CCCAC**GG**CCCU**GG**C**GG**
*mut*es12	**AAAG**UC**AA**CCCAC**AA**CCCU**AA**C**GG**

- The T7 promoter region is in italics, restriction sites are underscored (EcoRI 5', HindIII 3'). G-tracts and G-tract mutations are underscored and in bold.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

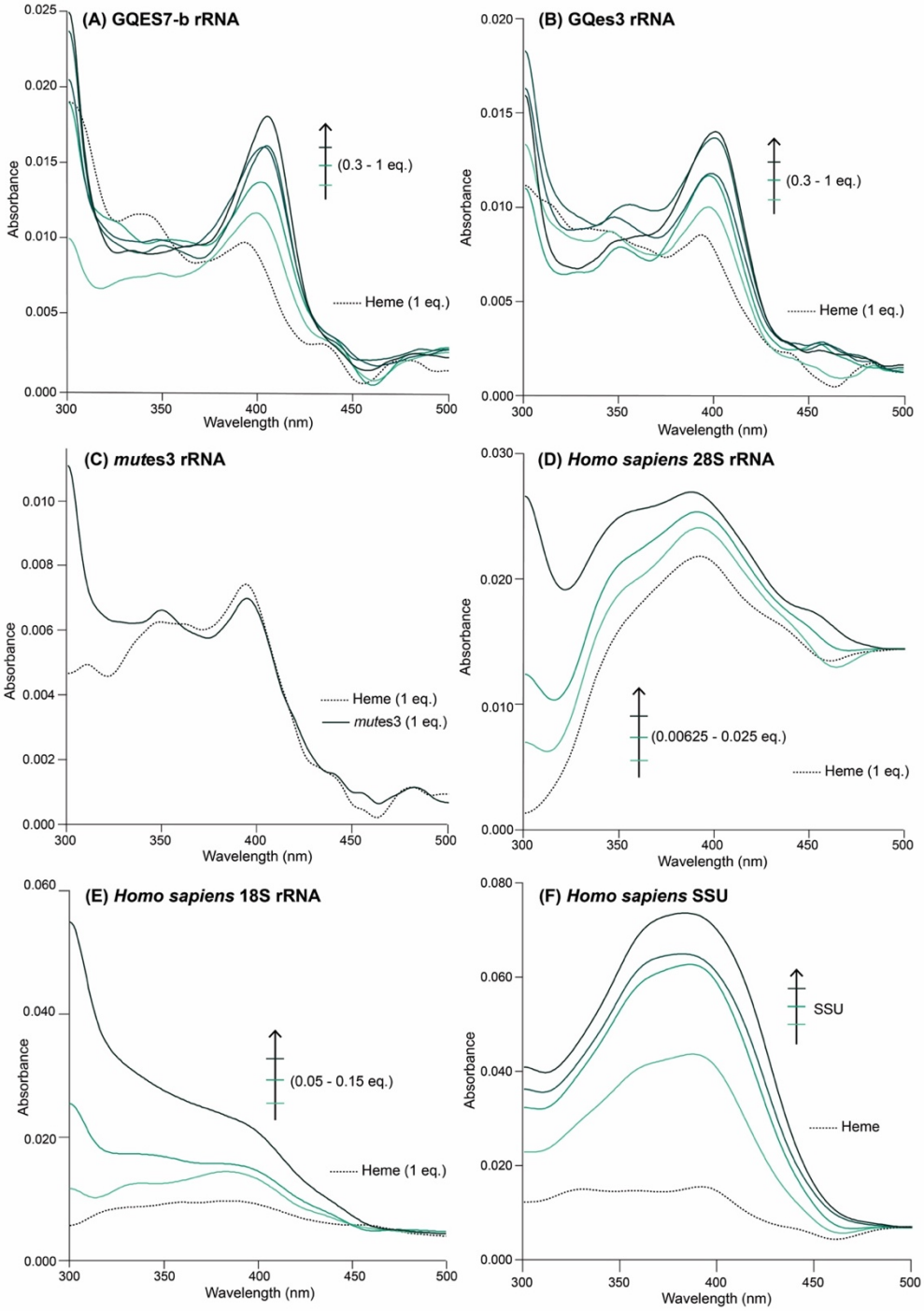


Figure C.1. UV-Vis spectra of increasing concentrations of GQES7-b (A), GQes3 (B), *mut*es3 (C), intact human 28S rRNA (D), intact human 18S rRNA (E), and assembled small subunit (F) to a constant concentration of heme.

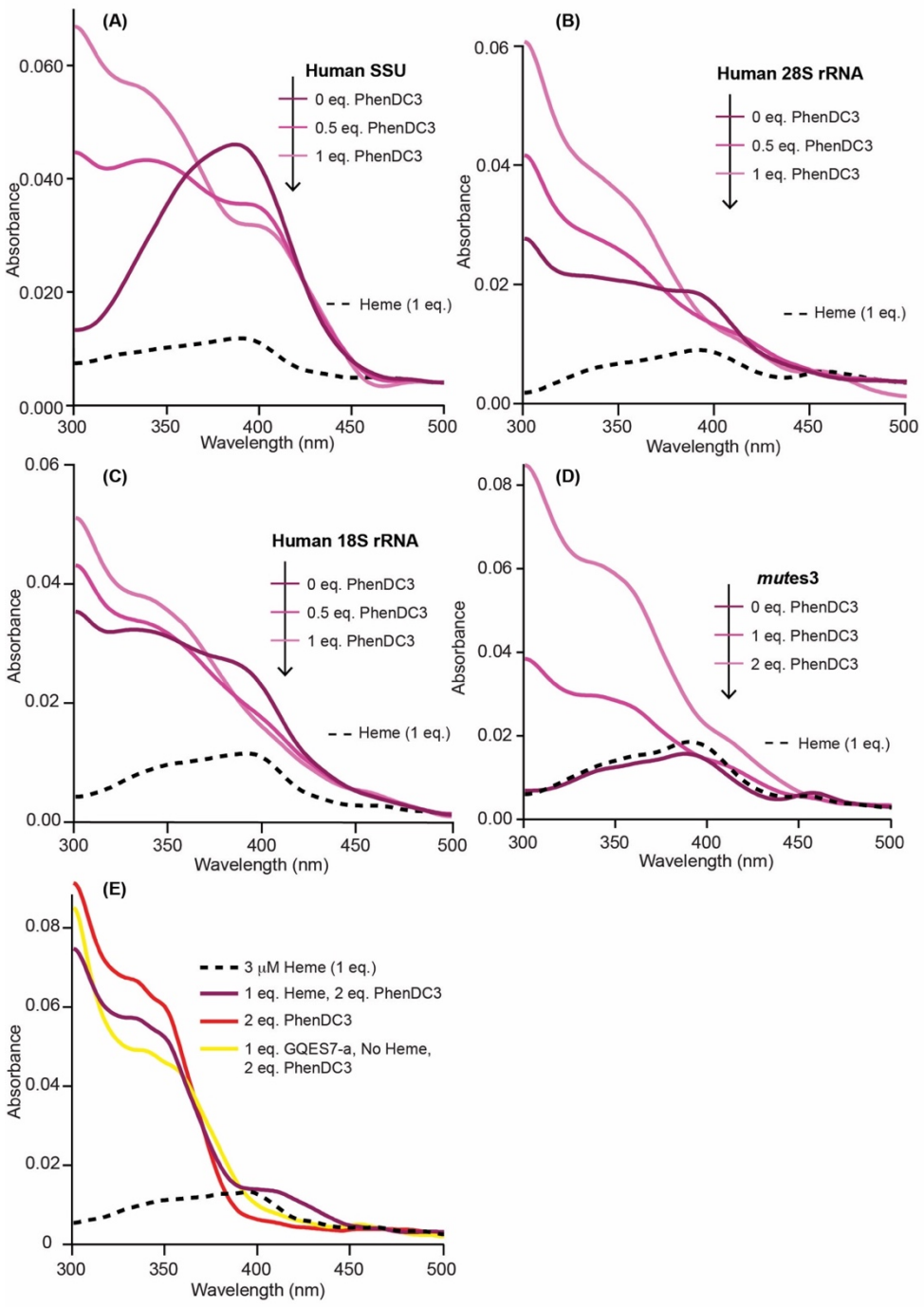


Figure C.2. UV-Vis spectra of titration of PhenDC3 to constant heme and (A) human assembled SSU, (B) intact human 28S rRNA, (C) intact human 18S rRNA, and (D) and *mut*es3. (E) UV-Vis spectra of PhenDC3 reveals peak at 350 nm is intrinsic of PhenDC3.

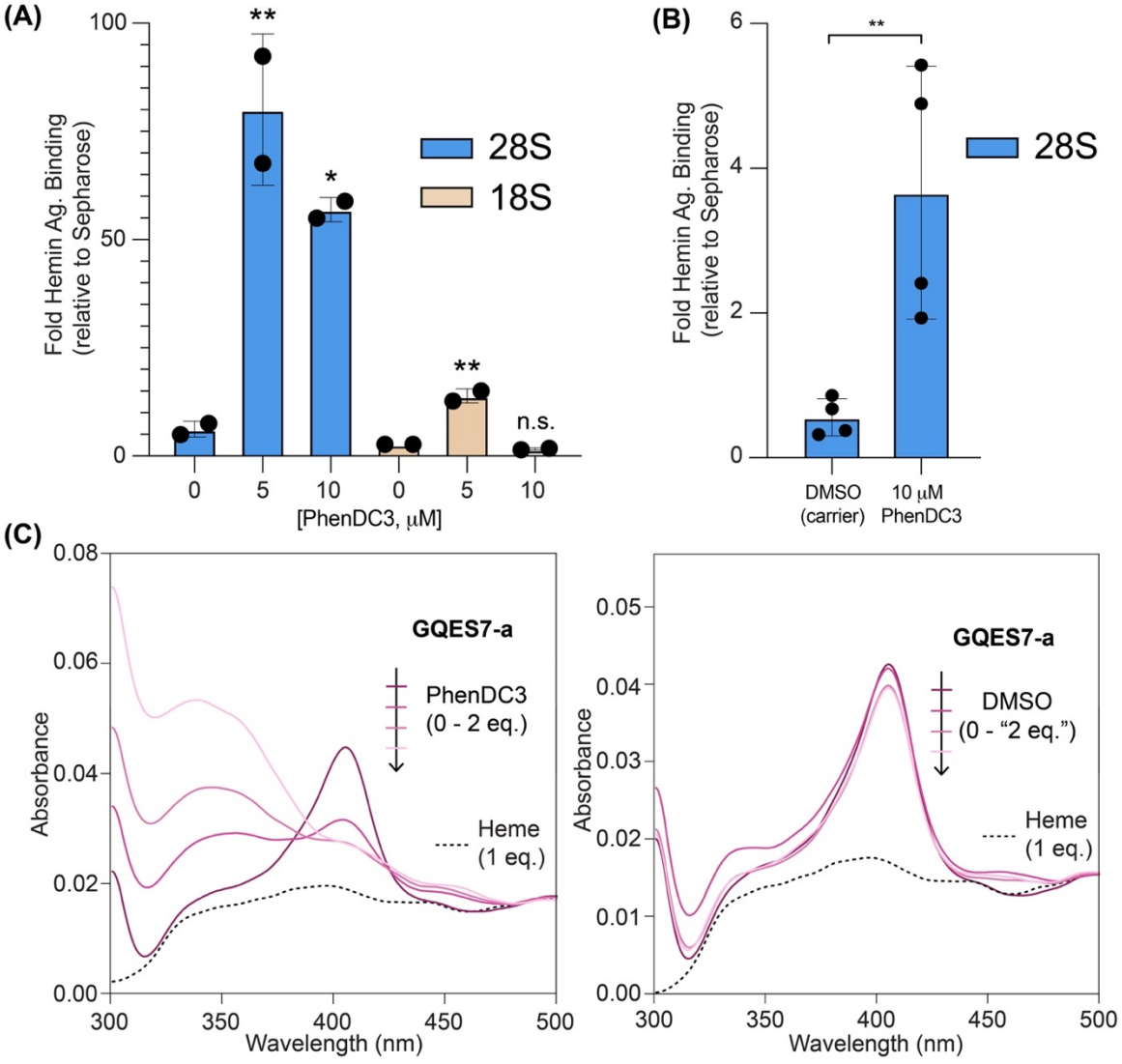


Figure C.3. (A) RT-qPCR analysis oif 28S and 18S rRNAs from PhenDC3-treated HEK293 cells. PhenDC3 treatment consisted of 48 hrs at 37 °C in the concentrations listed in the figure. Each dot represents a technical replicate coming from individual biological replicates. (B) RT-qPCR analysis of 28S rRNA from HEK293 cells treated with 10 mM PhenDC3 or carrier DMSO. Each dot in (B) represents a technical replicate coming from

2 biological replicates. (C) UV-Vis spectra of increasing concentrations of PhenDC3 or carrier DMSO to constant heme and GQES7-a show DMSO does not affect the binding of heme to ribosomal G4s. Data in (A) and (B) are represented as RNA enrichment in hemin agarose beads relative to control sepharose beads. Statistical significance is represented by asterisks using ordinary one-way ANOVA with Dunnett's post-hoc test. * $P < 0.05$; ** $P < 0.01$.

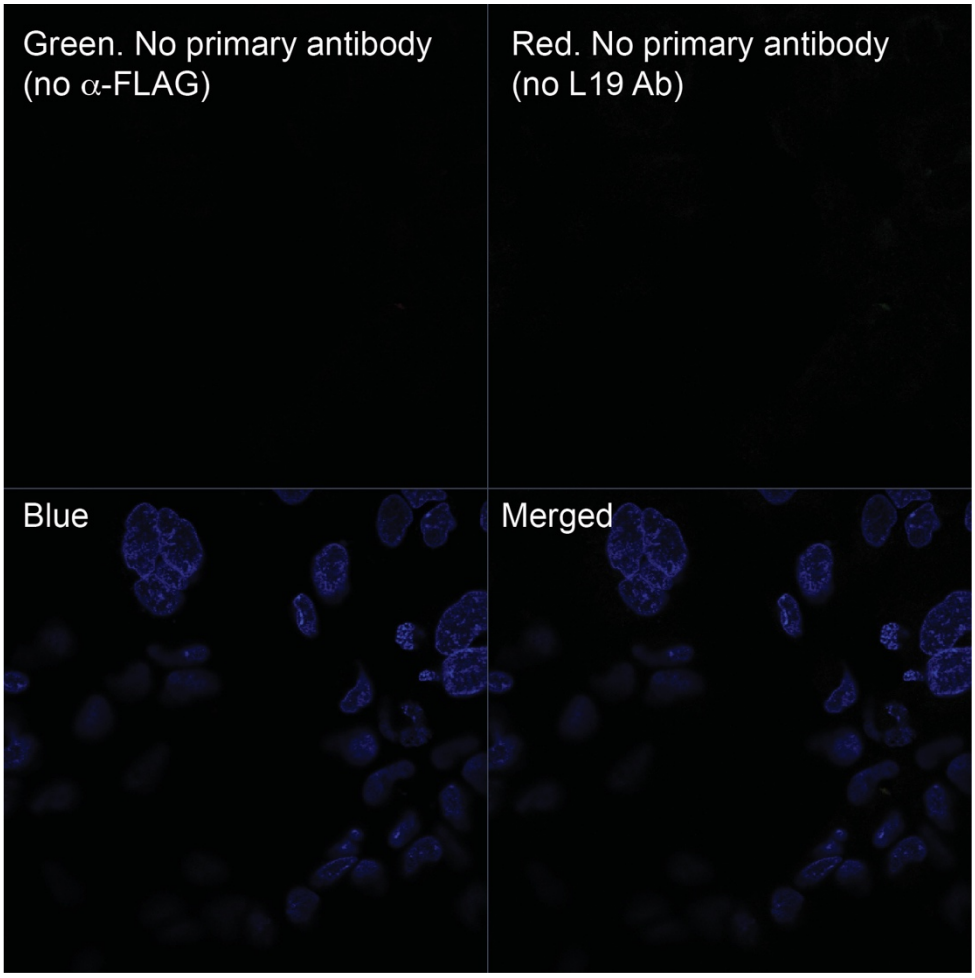


Figure C.4. Confocal microscopy images of fixed HEK293 cells without the primary antibodies α-FLAG (green channel) and without α-L19 antibody (red channel). Results demonstrate signals obtained in confocal microscopy images are coming from the primary antibodies.

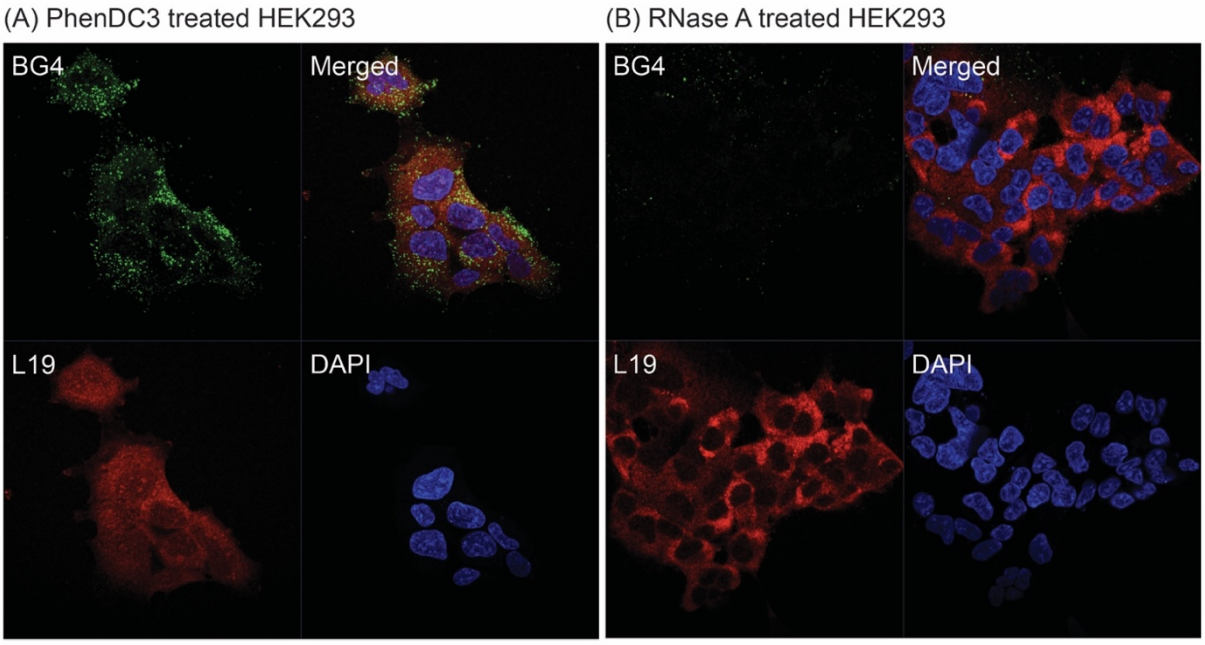


Figure C.5. Confocal microscopy images of fixed HEK293 cells treated with (A) PhenDC3 or (B) RNase A. Results demonstrate BG4 signal is coming from cellular RNA G-quadruplexes. Note that "merged" images presented here are not shown in terms of BG4-L19 colocalization.

Table C.1. DNA and RNA sequences encoding RNAs used.

Gene	Sequence (5' to 3')
GQES7-a	GAATTC*TAATACGACTCACTATAGGG*CGGA**GGGGG**C**GGG** CTCCGGC**GGG**TGC**GGGGGG**T**GGG**C**GGG**C**GGGGG**CC**GGG** **GG**T**GGGG**TCGGC**GGGGGG**ACCGAAGCTT
GQES7-b	GAATTC*TAATACGACTCACTATAGGG*CCTC**GGG**A**GGG**CGC GC**GGG**TC**GGGG**CGGCAAGCTT
GQes3	**GG**CCCC**GG**CC**GGGGGGG**C**GGG**CGCC**GG**
*mut*es3	**AA**CCCC**AA**CC**GAAAAG**C**GAA**GCC**AA**

- The T7 promoter region is in italics, restriction sites are underscored (EcoRI 5', HindIII 3'). G-tracts and G-tract mutations are underscored and in bold.

Table C.2. Primer sets used for RT-qPCR

Target RNA	**Forward**	**Reverse**
28S rRNA	AGAGGTAAACGGGTGGGGTC	GGGGTCGGGAGGAACGG
18S rRNA	GATGGTAGTCGCCGTGCC	GCCTGCTGCCTTCCTTGG

REFERENCES

1. Palade GE (1955) A Small Particulate Component of the Cytoplasm. *The Journal of Biophysical and Biochemical Cytology* 1(1):59.

2. Crick FH (1958) On Protein Synthesis. *Symp Soc Exp Biol*, p 8.

3. Noller HF, Hoffarth V, & Zimniak L (1992) Unusual Resistance of Peptidyl Transferase to Protein Extraction Procedures. *Science* 256(5062):1416-1419.

4. Garrett R & Grisham C (2009) *Biochemistry* (Belmont, Ca: Thomson Brooks/Cole).

5. Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanov VG, Harvey SC, Fox GE, & Wartell RM (2014) Evolution of the Ribosome at Atomic Resolution. *Proceedings of the National Academy of Sciences* 111(28):10251-10256.

6. Gerbi SA (1996) Expansion Segments: Regions of Variable Size That Interrupt the Universal Core Secondary Structure of Ribosomal RNA. *Ribosomal RNA—Structure, evolution, processing, and function in protein synthesis*:71-87.

7. Hassouna N, Mithot B, & Bachellerie J-P (1984) The Complete Nucleotide Sequence of Mouse 28S rRNA Gene. Implications for the Process of Size Increase of the Large Subunit rRNA in Higher Eukaryotes. *Nucleic Acids Research* 12(8):3563-3583.

8. Bachellerie J-P & Michot B (1989) Evolution of Large Subunit rRNA Structure. The 3′ Terminal Domain Contains Elements of Secondary Structure Specific to Major Phylogenetic Groups. *Biochimie* 71(6):701-709.

9. Michot B & Bachellerie J-P (1987) Comparisons of Large Subunit rRNAs Reveal Some Eukaryote-Specific Elements of Secondary Structure. *Biochimie* 69(1):11-23.

10. Milo R & Phillips R (2015) *Cell Biology by the Numbers* (Garland Science).

11. Romanova L, Korobova F, Noniashvilli E, Dyban A, & Zatsepina O (2006) High Resolution Mapping of Ribosomal DNA in Early Mouse Embryos by Fluorescence in Situ Hybridization. *Biology of Reproduction* 74(5):807-815.

12. Wai HH, Vu L, Oakes M, & Nomura M (2000) Complete Deletion of Yeast Chromosomal Rdna Repeats and Integration of a New rDNA Repeat: Use of rDNA Deletion Strains for Functional Analysis of rDNA Promoter Elements in Vivo. *Nucleic Acids Research* 28(18):3524-3534.

13. Fujii K, Susanto TT, Saurabh S, & Barna M (2018) Decoding the Function of Expansion Segments in Ribosomes. *Molecular Cell* 72(6):1013-1020. e1016.

14. Ramesh M & Woolford JL (2016) Eukaryote-Specific rRNA Expansion Segments Function in Ribosome Biogenesis. *RNA* 22(8):1153-1162.

15. Ramos LMG, Smeekens JM, Kovacs NA, Bowman JC, Wartell RM, Wu R, & Williams LD (2016) Yeast rRNA Expansion Segments: Folding and Function. *Journal of Molecular Biology* 428(20):4048-4059.

16. Kwok CK & Merrick CJ (2017) G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends in Biotechnology* 35(10):997-1013.

17. Mestre-Fos S, Penev PI, Richards JC, Dean WL, Gray RD, Chaires JB, & Williams LD (2019) Profusion of G-Quadruplexes on Both Subunits of Metazoan Ribosomes. *PLoS One* 14(12).

18. Lightfoot HL, Hagen T, Cléry A, Allain FH-T, & Hall J (2018) Control of the Polyamine Biosynthesis Pathway by G_2-Quadruplexes. *eLife* 7:e36362.

19. Jodoin R, Bauer L, Garant J-M, Laaref AM, Phaneuf F, & Perreault J-P (2014) The Folding of 5′-UTR Human G-Quadruplexes Possessing a Long Central Loop. *RNA* 20(7):1129-1141.

20. Martadinata H & Phan AT (2014) Formation of a Stacked Dimeric G-Quadruplex Containing Bulges by the 5′-Terminal Region of Human Telomerase RNA (hTERC). *Biochemistry* 53(10):1595-1600.

21. Pandey S, Agarwala P, & Maiti S (2013) Effect of Loops and G-Quartets on the Stability of RNA G-Quadruplexes. *The Journal of Physical Chemistry B* 117(23):6896-6905.

22. Sacca B, Lacroix L, & Mergny J-L (2005) The Effect of Chemical Modifications on the Thermal Stability of Different G-Quadruplex-Forming Oligonucleotides. *Nucleic Acids Research* 33(4):1182-1192.

23. Bhattacharyya D, Mirihana Arachchilage G, & Basu S (2016) Metal Cations in G-Quadruplex Folding and Stability. *Frontiers in Chemistry* 4:38.

24. Biffi G, Di Antonio M, Tannahill D, & RNA S (2014) Visualization and Selective Chemical Targeting of Rna G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nature Chemistry* 6(1):75.

25. Biffi G, Tannahill D, Mccafferty J, & Balasubramanian S (2013) Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nature Chemistry* 5(3):182.

26. Renard I, Grandmougin M, Roux A, Yang SY, Lejault P, Pirrotta M, Wong JM, & Monchaud D (2019) Small-Molecule Affinity Capture of DNA/RNA Quadruplexes and Their Identification in Vitro and in Vivo through the G4RP Protocol. *Nucleic Acids Research* 47(11):5502-5510.

27. Yang SY, Lejault P, Chevrier S, Boidot R, Robertson AG, Wong JM, & Monchaud D (2018) Transcriptome-Wide Identification of Transient RNA G-Quadruplexes in Human Cells. *Nature Communications* 9(1):1-11.

28. Chen S-B, Hu M-H, Liu G-C, Wang J, Ou T-M, Gu L-Q, Huang Z-S, & Tan J-H (2016) Visualization of NRAS RNA G-Quadruplex Structures in Cells with an Engineered Fluorogenic Hybridization Probe. *Journal of the American Chemical Society* 138(33):10382-10385.

29. Chen XC, Chen SB, Dai J, Yuan JH, Ou TM, Huang ZS, & Tan JH (2018) Tracking the Dynamic Folding and Unfolding of RNA G-Quadruplexes in Live Cells. *Angewandte Chemie International Edition* 57(17):4702-4706.

30. Bao H-L, Ishizuka T, Sakamoto T, Fujimoto K, Uechi T, Kenmochi N, & Xu Y (2017) Characterization of Human Telomere RNA G-Quadruplex Structures in Vitro and in Living Cells Using 19F NMR Spectroscopy. *Nucleic Acids Research* 45(9):5501-5511.

31. Guo JU & Bartel DP (2016) RNA G-Quadruplexes Are Globally Unfolded in Eukaryotic Cells and Depleted in Bacteria. *Science* 353(6306):aaf5371.

32. Fay MM, Lyons SM, & Ivanov P (2017) Rna G-Quadruplexes in Biology: Principles and Molecular Mechanisms. *Journal of Molecular Biology* 429(14):2127-2147.

33. Hänsel-Hertsch R, Spiegel J, Marsico G, Tannahill D, & Balasubramanian S (2018) Genome-Wide Mapping of Endogenous G-Quadruplex DNA Structures by Chromatin Immunoprecipitation and High-Throughput Sequencing. *Nature Protocols* 13(3):551.

34. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, & Balasubramanian S (2016) Rg4-Seq Reveals Widespread Formation of G-Quadruplex Structures in the Human Transcriptome. *Nature Methods* 13(10):841.

35. Kumari S, Bugaut A, Huppert JL, & Balasubramanian S (2007) An Rna G-Quadruplex in the 5′ Utr of the Nras Proto-Oncogene Modulates Translation. *Nature Chemical Biology* 3(4):218-221.

36. Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, & Kurreck J (2008) Inhibition of Translation in Living Eukaryotic Cells by an RNA G-Quadruplex Motif. *RNA* 14(7):1290-1296.

37. Morris MJ & Basu S (2009) An Unusually Stable G-Quadruplex within the 5′-UTR of the MT3 Matrix Metalloproteinase Mrna Represses Translation in Eukaryotic Cells. *Biochemistry* 48(23):5313-5319.

38. Gomez D, Guedin A, Mergny J-L, Salles B, Riou J-F, Teulade-Fichou M-P, & Calsou P (2010) A G-Quadruplex Structure within the 5′-UTR of TRF2 mRNA Represses Translation in Human Cells. *Nucleic Acids Research* 38(20):7187-7198.

39. Balkwill GD, Derecka K, Garner TP, Hodgman C, Flint AP, & Searle MS (2009) Repression of Translation of Human Estrogen Receptor A by G-Quadruplex Formation. *Biochemistry* 48(48):11487-11495.

40. Beaudoin J-D & Perreault J-P (2010) 5′-UTR G-Quadruplex Structures Acting as Translational Repressors. *Nucleic Acids Research* 38(20):7022-7036.

41. Kwok CK, Sahakyan AB, & Balasubramanian S (2016) Structural Analysis Using Shalipe to Reveal RNA G-Quadruplex Formation in Human Precursor microRNA. *Angewandte Chemie International Edition* 55(31):8958-8961.

42. Pandey S, Agarwala P, Jayaraj GG, Gargallo R, & Maiti S (2015) The RNA Stem–Loop to G-Quadruplex Equilibrium Controls Mature microRNA Production inside the Cell. *Biochemistry* 54(48):7067-7078.

43. Rouleau SG, Garant J-M, Bolduc F, Bisaillon M, & Perreault J-P (2018) G-Quadruplexes Influence pri-microRNA Processing. *RNA Biology* 15(2):198-206.

44. Martadinata H & Phan AT (2009) Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric Rna Sequences in K+ Solution. *Journal of the American Chemical Society* 131(7):2570-2578.

45. Xiao C-D, Shibata T, Yamamoto Y, & Xu Y (2018) An Intramolecular Antiparallel G-Quadruplex Formed by Human Telomere RNA. *Chemical Communications* 54(32):3944-3946.

46. Métifiot M, Amrane S, Litvak S, & Andreola M-L (2014) G-Quadruplexes in Viruses: Function and Potential Therapeutic Applications. *Nucleic Acids Research* 42(20):12352-12366.

47. Jaubert C, Bedrat A, Bartolucci L, Di Primo C, Ventura M, Mergny J-L, Amrane S, & Andreola M-L (2018) RNA Synthesis Is Modulated by G-Quadruplex Formation in Hepatitis C Virus Negative RNA Strand. *Scientific Reports* 8(1):1-9.

48. Zhang Y, Yang M, Duncan S, Yang X, Abdelhamid MA, Huang L, Zhang H, Benfey PN, Waller ZA, & Ding Y (2019) G-Quadruplex Structures Trigger RNA Phase Separation. *Nucleic Acids Research* 47(22):11746-11754.

49. Jain A & Vale RD (2017) RNA Phase Transitions in Repeat Expansion Disorders. *Nature* 546(7657):243-247.

50. Sen D & Poon LC (2011) RNA and DNA Complexes with Hemin [Fe (III) Heme] Are Efficient Peroxidases and Peroxygenases: How Do They Do It and What Does It Mean? *Critical Reviews in Biochemistry and Molecular Biology* 46(6):478-492.

51. Poon LC-H, Methot SP, Morabi-Pazooki W, Pio F, Bennet AJ, & Sen D (2011) Guanine-Rich RNAs and DNAs That Bind Heme Robustly Catalyze Oxygen Transfer Reactions. *Journal of the American Chemical Society* 133(6):1877-1884.

52. Li Y & Sen D (1996) A Catalytic DNA for Porphyrin Metallation. *Nature Structural Biology* 3(9):743-747.

53. Hooda J, Shah A, & Zhang L (2014) Heme, an Essential Nutrient from Dietary Proteins, Critically Impacts Diverse Physiological and Pathological Processes. *Nutrients* 6(3):1080-1102.

54. Severance S & Hamza I (2009) Trafficking of Heme and Porphyrins in Metazoa. *Chemical Reviews* 109(10):4596-4616.

55. Donegan RK, Moore CM, Hanna DA, & Reddi AR (2019) Handling Heme: The Mechanisms Underlying the Movement of Heme within and between Cells. *Free Radical Biol Med* 133:88-100.

56. Hanna DA, Martinez-Guzman O, & Reddi AR (2017) Heme Gazing: Illuminating Eukaryotic Heme Trafficking, Dynamics, and Signaling with Fluorescent Heme Sensors. *Biochemistry* 56(13):1815-1823.

57. Reddi AR & Hamza I (2016) Heme Mobilization in Animals: A Metallolipid's Journey. *Accounts of Chemical Research* 49(6):1104-1110.

58. Kumar S & Bandyopadhyay U (2005) Free Heme Toxicity and Its Detoxification Systems in Human. *Toxicology Letters* 157(3):175-188.

59. Atamna H & Frey WH (2004) A Role for Heme in Alzheimer's Disease: Heme Binds Amyloid B and Has Altered Metabolism. *Proceedings of the National Academy of Sciences* 101(30):11153-11158.

60. Hanna DA, Harvey RM, Martinez-Guzman O, Yuan X, Chandrasekharan B, Raju G, Outten FW, Hamza I, & Reddi AR (2016) Heme Dynamics and Trafficking Factors Revealed by Genetically Encoded Fluorescent Heme Sensors. *Proceedings of the National Academy of Sciences* 113(27):7539-7544.

61. Martinez-Guzman O, Dietz JV, Bohovych I, Medlock AW, Khalimonchuk O, & Reddi AR (2019) Mitochondrial-Nuclear Heme Trafficking Is Regulated by Gtpases That Control Mitochondrial Dynamics. *bioRxiv*:539254.

62. Martinez-Guzman O, Willoughby MM, Saini A, Dietz JV, Bohovych I, Medlock AE, Khalimonchuk O, & Reddi AR (2019) The Heme Biosynthetic Enzyme, 5-Aminolevulinic Acid Synthase (Alas), and Gtpases in Control of Mitochondrial Dynamics and Er Contact Sites, Regulate Heme Mobilization to the Nucleus. *bioRxiv*:734780.

63. Gray LT, Lombardi EP, Verga D, Nicolas A, Teulade-Fichou M-P, Londoño-Vallejo A, & Maizels N (2019) G-Quadruplexes Sequester Free Heme in Living Cells. *Cell Chemical Biology* 26(12):1681-1691. e1685.

64. Bernier CR, Petrov AS, Kovacs NA, Penev PI, & Williams LD (2018) Translation: The Universal Structural Core of Life. *Molecular Biology and Evolution* 35(8):2065-2076.

65. Ware VC, Tague BW, Clark CG, Gourse RL, Brand RC, & Gerbi SA (1983) Sequence Analysis of 28s Ribosomal DNA from the Amphibian *Xenopus Laevis*. *Nucleic Acids Research* 11(22):7795-7817.

66. Clark CG, Tague BW, Ware VC, & Gerbi SA (1984) *Xenopus Laevis* 28S Ribosomal RNA: A Secondary Structure Model and Its Evolutionary and Functional Implications. *Nucleic Acids Research* 12(15):6197-6220.

67. Hassouna N, Michot B, & Bachellerie JP (1984) The Complete Nucleotide Sequence of Mouse 28S rRNA Gene. Implications for the Process of Size Increase of the Large Subunit rRNA in Higher Eukaryotes. *Nucleic Acids Research* 12(8):3563-3583.

68. Leidig C, Bange G, Kopp J, Amlacher S, Aravind A, Wickles S, Witte G, Hurt E, Beckmann R, & Sinning I (2013) Structural Characterization of a Eukaryotic Chaperone--the Ribosome-Associated Complex. *Nature Structural Molecular Biology* 20(1):23-28.

69. Gumiero A, Conz C, Gesé GV, Zhang Y, Weyer FA, Lapouge K, Kappes J, Von Plehwe U, Schermann G, & Fitzke E (2016) Interaction of the Cotranslational Hsp70 Ssb with Ribosomal Proteins and rRNA Depends on Its Lid Domain. *Nature Communications* 7:13563.

70. Knorr AG, Schmidt C, Tesina P, Berninghausen O, Becker T, Beatrix B, & Beckmann R (2018) Ribosome-NatA Architecture Reveals That rRNA Expansion Segments Coordinate N-Terminal Acetylation. *Nature Structural Molecular Biology* 26:35-39.

71. Melnikov S, Ben-Shem A, Garreau De Loubresse N, Jenner L, Yusupova G, & Yusupov M (2012) One Core, Two Shells: Bacterial and Eukaryotic Ribosomes. *Nature Structural Molecular Biology* 19(6):560-567.

72. Kikin O, D'antonio L, & Bagga PS (2006) Qgrs Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Research* 34(Web Server issue):W676-682.

73. Lammich S, Kamp F, Wagner J, Nuscher B, Zilow S, Ludwig A-K, Willem M, & Haass C (2011) Translational Repression of the Disintegrin and Metalloprotease Adam10 by a Stable G-Quadruplex Secondary Structure in Its 5′-Untranslated Region. *Journal Biological Chemistry* 286(52):45063-45072.

74. Von Hacht A, Seifert O, Menger M, Schutze T, Arora A, Konthur Z, Neubauer P, Wagner A, Weise C, & Kurreck J (2014) Identification and Characterization of RNA Guanine-Quadruplex Binding Proteins. *Nucleic Acids Research* 42(10):6630-6644.

75. Fratta P, Mizielinska S, Nicoll AJ, Zloh M, Fisher EM, Parkinson G, & Isaacs AM (2012) *C9orf72* Hexanucleotide Repeat Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Forms RNA G-Quadruplexes. *Scientific Reports* 2:1016.

76. Halder K, Wieland M, & Hartig JS (2009) Predictable Suppression of Gene Expression by 5'-UTR-Based Rna Quadruplexes. *Nucleic Acids Research* 37(20):6811-6817.

77. Arora A, Dutkiewicz M, Scaria V, Hariharan M, Maiti S, & Kurreck J (2008) Inhibition of Translation in Living Eukaryotic Cells by an RNA G-Quadruplex Motif. *RNA* 14(7):1290-1296.

78. Kumari S, Bugaut A, Huppert JL, & Balasubramanian S (2007) An RNA G-Quadruplex in the 5' UTR of the NRAS Proto-Oncogene Modulates Translation. *Nature Chemical Biology* 3(4):218-221.

79. Choi KH & Choi BS (1994) Formation of a Hairpin Structure by Telomere 3' Overhang. *Biochim Biophys Acta-Gene Struct Expression* 1217(3):341-344.

80. Xu S, Li Q, Xiang J, Yang Q, Sun H, Guan A, Wang L, Liu Y, Yu L, & Shi Y (2016) Thioflavin T as an Efficient Fluorescence Sensor for Selective Recognition of RNA G-Quadruplexes. *Scientific Reports* 6.

81. Renaud De La Faverie A, Guédin A, Bedrat A, Yatsunyk LA, & Mergny J-L (2014) Thioflavin T as a Fluorescence Light-up Probe for G4 Formation. *Nucleic Acids Research* 42(8):e65-e65.

82. Zhang S, Sun H, Chen H, Li Q, Guan A, Wang L, Shi Y, Xu S, Liu M, & Tang Y (2018) Direct Visualization of Nucleolar G-Quadruplexes in Live Cells by Using a Fluorescent Light-up Probe. *Biochim Biophys Acta* 1862(5):1101-1106.

83. Mergny J-L, Phan A-T, & Lacroix L (1998) Following G-Quartet Formation by UV-Spectroscopy. *FEBS Lett* 435(1):74-78.

84. Zheng K-W, Chen Z, Hao Y-H, & Tan Z (2009) Molecular Crowding Creates an Essential Environment for the Formation of Stable G-Quadruplexes in Long Double-Stranded DNA. *Nucleic Acids Research* 38(1):327-338.

85. Cang X, Šponer J, & Cheatham Iii TE (2011) Explaining the Varied Glycosidic Conformational, G-Tract Length and Sequence Preferences for Anti-Parallel G-Quadruplexes. *Nucleic Acids Research* 39(10):4499-4512.

86. Li X-M, Zheng K-W, Zhang J-Y, Liu H-H, Yuan B-F, Hao Y-H, & Tan Z (2015) Guanine-Vacancy–Bearing G-Quadruplexes Responsive to Guanine Derivatives. *Proceedings of the National Academy of Sciences* 112(47):14581-14586.

87. Benhalevy D, Gupta SK, Danan CH, Ghosal S, Sun H-W, Kazemier HG, Paeschke K, Hafner M, & Juranek SA (2017) The Human CCHC-Type Zinc Finger Nucleic Acid-Binding Protein Binds G-Rich Elements in Target mRNA Coding Sequences and Promotes Translation. *Cell Reports* 18(12):2979-2990.

88. Khateb S, Weisman-Shomer P, Hershco-Shani I, Ludwig AL, & Fry M (2007) The Tetraplex $(CGG)_N$ Destabilizing Proteins hnRNP A2 and Cbf-a Enhance the in Vivo Translation of Fragile X Premutation mRNA. *Nucleic Acids Research* 35(17):5775-5788.

89. Matunis MJ, Xing J, & Dreyfuss G (1994) The hnRNP F Protein: Unique Primary Structure, Nucleic Acid-Binding Properties, and Subcellular Localization. *Nucleic Acids Research* 22(6):1059-1067.

90. Caputi M & Zahler AM (2001) Determination of the Rna Binding Specificity of the Heterogeneous Nuclear Ribonucleoprotein (HNRNP) H/H′/F/2H9 Family. *Journal of Biological Chemistry* 276(47):43850-43859.

91. Mcrae EK, Booy EP, Padilla-Meier GP, & Mckenna SA (2017) On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids. *Journal of Nucleic Acids* 2017.

92. Dardenne E, Espinoza MP, Fattet L, Germann S, Lambert M-P, Neil H, Zonta E, Mortada H, Gratadou L, & Deygas M (2014) Rna Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. *Cell Reports* 7(6):1900-1913.

93. Herdy B, Mayer C, Varshney D, Marsico G, Murat P, Taylor C, D'santos C, Tannahill D, & Balasubramanian S (2018) Analysis of NRAS RNA G-Quadruplex Binding Proteins Reveals DDX3X as a Novel Interactor of Cellular G-Quadruplex Containing Transcripts. *Nucleic Acids Research*.

94. Chaudhury A, Chander P, & Howe PH (2010) Heterogeneous Nuclear Ribonucleoproteins (hnRNPs) in Cellular Processes: Focus on hnRNP E1's Multifunctional Regulatory Roles. *RNA*.

95. Simsek D, Tiu GC, Flynn RA, Byeon GW, Leppek K, Xu AF, Chang HY, & Barna M (2017) The Mammalian Ribo-Interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* 169(6):1051-1065. e1018.

96. Huppert JL, Bugaut A, Kumari S, & Balasubramanian S (2008) G-Quadruplexes: The Beginning and End of UTRs. *Nucleic Acids Research* 36(19):6260-6268.

97. Eddy J & Maizels N (2008) Conserved Elements with Potential to Form Polymorphic G-Quadruplex Structures in the First Intron of Human Genes. *Nucleic Acids Reserach* 36(4):1321-1333.

98. Drygin D, Siddiqui-Jain A, O'brien S, Schwaebe M, Lin A, Bliesath J, Ho CB, Proffitt C, Trent K, & Whitten JP (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Research* 69(19):7653-7661.

99. Chiarella S, De Cola A, Scaglione GL, Carletti E, Graziano V, Barcaroli D, Lo Sterzo C, Di Matteo A, Di Ilio C, & Falini B (2013) Nucleophosmin Mutations Alter Its Nucleolar Localization by Impairing G-Quadruplex Binding at Ribosomal DNA. *Nucleic Acids Research* 41(5):3228-3239.

100. Kim J, Cheong C, & Moore PB (1991) Tetramerization of an RNA Oligonucleotide Containing a GGGG Sequence. *Nature* 351(6324):331.

101. Del Villar-Guerra R, Gray RD, & Chaires JB (2017) Characterization of Quadruplex DNA Structure by Circular Dichroism. *Curr Protoc Nucleic Acid Chem*:17.18. 11-17.18. 16.

102. Shcherbakov D & Piendl W (2007) A Novel View of Gel-Shifts: Analysis of RNA-Protein Complexes Using a Two-Color Fluorescence Dye Procedure. *Electrophoresis* 28(5):749-755.

103. Bernier C, Petrov AS, Waterbury C, Jett J, Li F, Freil LE, Xiong B, Wang L, Le A, Milhouse BL, Hershkovitz E, Grover M, Xue Y, Hsiao C, Bowman JC, Harvey SC, Wartel JZ, & Williams LD (2014) Ribovision: Visualization and Analysis of Ribosomes. *Faraday Discussions* 169(1):195-207.

104. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, & Lipman DJ (1997) Gapped Blast and Psi-Blast: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* 25(17):3389-3402.

105. Anonymous (2017) Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 45(D1):D12-d17.

106. Katoh K & Standley DM (2016) A Simple Method to Control over-Alignment in the Mafft Multiple Sequence Alignment Program. *Bioinformatics* 32(13):1933-1942.

107. Hall TA (1999) Bioedit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/Nt. *Nucleic Acids Symp Ser*, ([London]: Information Retrieval Ltd., c1979-c2000.), pp 95-98.

108. Waterhouse AM, Procter JB, Martin DM, Clamp M, & Barton GJ (2009) Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* 25(9):1189-1191.

109. Hedges SB, Dudley J, & Kumar S (2006) Timetree: A Public Knowledge-Base of Divergence Times among Organisms. *Bioinformatics* 22(23):2971-2972.

110. Eng JK, Mccormack AL, & Yates JR (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* 5(11):976-989.

111. Mestre-Fos S, Penev PI, Suttapitugsakul S, Hu M, Ito C, Petrov AS, Wartell RM, Wu R, & Williams LD (2019) G-Quadruplexes in Human Ribosomal RNA. *Journal of Molecular Biology* 431(10):1940-1955.

112. Millevoi S, Moine H, & Vagner S (2012) G-Quadruplexes in RNA Biology. *Wiley Interdisciplinary Reviews: RNA* 3(4):495-507.

113. Qin M, Chen Z, Luo Q, Wen Y, Zhang N, Jiang H, & Yang H (2015) Two-Quartet G-Quadruplexes Formed by DNA Sequences Containing Four Contiguous GG Runs. *Journal of Physical Chemistry B* 119(9):3706-3713.

114. Rudra D, Zhao Y, & Warner JR (2005) Central Role of Ifh1p–Fhl1p Interaction in the Synthesis of Yeast Ribosomal Proteins. *EMBO Journal* 24(3):533-542.

115. Khatter H, Myasnikov AG, Natchiar SK, & Klaholz BP (2015) Structure of the Human 80S Ribosome. *Nature* 520:640.

116. Singh YH, Andrabi M, Kahali B, Ghosh TC, Mizuguchi K, Kochetov AV, & Ahmad S (2010) On Nucleotide Solvent Accessibility in RNA Structure. *Gene* 463(1-2):41-48.

117. Kikin O, D'antonio L, & Bagga PS (2006) Qgrs Mapper: A Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Research* 34(suppl_2):W676-W682.

118. Del Villar-Guerra R, Trent JO, & Chaires JB (2018) G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angewandte Chemistry International Edition* 57(24):7171-7175.

119. Kypr J, Kejnovská I, Renčiuk D, & Vorlíčková M (2009) Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Research* 37(6):1713-1725.

120. Ratliff RL, Liu JJ, Vaughan MR, & Gray DM (1986) Sequence-Dependence of the Cd of Synthetic Double-Stranded Rnas Containing Inosinate Instead of Guanylate Subunits. *Biopolymers* 25(9):1735-1750.

121. Schuck P (2000) Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal* 78(3):1606-1619.

122. Renaud De La Faverie A, Guedin A, Bedrat A, Yatsunyk LA, & Mergny J-L (2014) Thioflavin T as a Fluorescence Light-up Probe for G4 Formation. *Nucleic Acids Research* 42(8):e65-e65.

123. Zhang S, Sun H, Chen H, Li Q, Guan A, Wang L, Shi Y, Xu S, Liu M, & Tang Y (2018) Direct Visualization of Nucleolar G-Quadruplexes in Live Cells by Using a Fluorescent Light-up Probe. *Biochim Biophys Acta* 1862(5):1101-1106.

124. Dean WL, Gray RD, Monsen RC, Deleeuw L, & Chaires JB (2018) Putting a New Spin of G4 Structure and Binding by Analytical Ultracentrifugation. *bioRxiv*:359356.

125. Maharana S, Wang J, Papadopoulos DK, Richter D, Pozniakovsky A, Poser I, Bickle M, Rizk S, Guillén-Boixet J, Franzmann TM, Jahnel M, Marrone L, Chang Y-T, Sterneckert J, Tomancak P, Hyman AA, & Alberti S (2018) RNA Buffers the Phase Separation Behavior of Prion-Like RNA Binding Proteins. *Science* 360(6391):918-921.

126. Dormann D & Haass C (2013) Fused in Sarcoma (Fus): An Oncogene Goes Awry in Neurodegeneration. *Mol Cell Neurosci* 56:475-486.

127. Knorr AG, Schmidt C, Tesina P, Berninghausen O, Becker T, Beatrix B, & Beckmann R (2019) Ribosome-NatA Architecture Reveals That rRNA Expansion Segments Coordinate N-Terminal Acetylation. *Nature Structural Molecular Biology* 26:35-39.

128. Zhang Z, Schwartz S, Wagner L, & Miller W (2000) A Greedy Algorithm for Aligning DNA Sequences. *J Comput Biol* 7(1-2):203-214.

129. Anonymous (2018) Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 46(D1):D8-d13.

130. Katoh K & Standley DM (2013) Mafft Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30(4):772-780.

131. Waterhouse AM, Procter JB, Martin DMA, Clamp M, & Barton GJ (2009) Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* 25(9):1189-1191.

132. Kumar S, Stecher G, Suleski M, & Hedges SB (2017) Timetree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 34(7):1812-1819.

133. Paeschke K, Capra JA, & Zakian VA (2011) DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces Cerevisiae Pif1 DNA Helicase. *Cell* 145(5):678-691.

134. Siddiqui-Jain A, Grand CL, Bearss DJ, & Hurley LH (2002) Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress C-Myc Transcription. *Proceedings of the National Academy of Sciences* 99(18):11593-11598.

135. Ribeyre C, Lopes J, Boulé J-B, Piazza A, Guédin A, Zakian VA, Mergny J-L, & Nicolas A (2009) The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming Ceb1 Sequences in Vivo. *PLoS Genetics* 5(5).

136. Moye AL, Porter KC, Cohen SB, Phan T, Zyner KG, Sasaki N, Lovrecz GO, Beck JL, & Bryan TM (2015) Telomeric G-Quadruplexes Are a Substrate and Site of Localization for Human Telomerase. *Nature Communications* 6:7643.

137. Conlon EG, Lu L, Sharma A, Yamazaki T, Tang T, Shneider NA, & Manley JL (2016) The C9orf72 Ggggcc Expansion Forms RNA G-Quadruplex Inclusions and Sequesters hnRNP H to Disrupt Splicing in Als Brains. *eLife* 5:e17820.

138. Hänsel-Hertsch R, Beraldi D, Lensing SV, Marsico G, Zyner K, Parry A, Di Antonio M, Pike J, Kimura H, & Narita M (2016) G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nature Genetics* 48(10):1267.

139. Laguerre A, Wong JM, & Monchaud D (2016) Direct Visualization of Both DNA and RNA Quadruplexes in Human Cells Via an Uncommon Spectroscopic Method. *Scientific Reports* 6:32141.

140. Monchaud D, Allain C, Bertrand H, Smargiasso N, Rosu F, Gabelica V, De Cian A, Mergny J-L, & Teulade-Fichou M-P (2008) Ligands Playing Musical Chairs with G-Quadruplex DNA: A Rapid and Simple Displacement Assay for Identifying Selective G-Quadruplex Binders. *Biochimie* 90(8):1207-1223.

141. De Cian A, Delemos E, Mergny J-L, Teulade-Fichou M-P, & Monchaud D (2007) Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *Journal of the American Chemical Society* 129(7):1856-1857.

142. Reid DW & Nicchitta CV (2012) Primary Role for Endoplasmic Reticulum-Bound Ribosomes in Cellular Translation Identified by Ribosome Profiling. *Journal of Biological Chemistry* 287(8):5518-5527.

143. Shimizu H, Tai H, Saito K, Shibata T, Kinoshita M, & Yamamoto Y (2015) Characterization of the Interaction between Heme and a Parallel G-Quadruplex DNA Formed from D(TTAGGGT). *Bull Chem Soc Jpn* 88(5):644-652.

144. Saito K, Tai H, Hemmi H, Kobayashi N, & Yamamoto Y (2012) Interaction between the Heme and a G-Quartet in a Heme–DNA Complex. *Inorganic Chemistry* 51(15):8168-8176.

145. Shinomiya R, Katahira Y, Araki H, Shibata T, Momotake A, Yanagisawa S, Ogura T, Suzuki A, Neya S, & Yamamoto Y (2018) Characterization of Catalytic

Activities and Heme Coordination Structures of Heme–DNA Complexes Composed of Some Chemically Modified Hemes and an All Parallel-Stranded Tetrameric G-Quadruplex DNA Formed from D (TTAGGG). *Biochemistry* 57(41):5930-5937.

146. Ebert PS, Hess RA, Frykholm BC, & Tschudy DP (1979) Succinylacetone, a Potent Inhibitor of Heme Biosynthesis: Effect on Cell Growth, Heme Content and Δ-Aminolevulinic Acid Dehydratase Activity of Malignant Murine Erythroleukemia Cells. *Biochem Biophys Res Commun* 88(4):1382-1390.

147. Hanna DA, Hu R, Kim H, Martinez-Guzman O, Torres MP, & Reddi AR (2018) Heme Bioavailability and Signaling in Response to Stress in Yeast Cells. *Journal of Biological Chemistry* 293(32):12378-12393.

148. Sweeny EA, Singh AB, Chakravarti R, Martinez-Guzman O, Saini A, Haque MM, Garee G, Dans PD, Hannibal L, & Reddi AR (2018) Glyceraldehyde-3-Phosphate Dehydrogenase Is a Chaperone That Allocates Labile Heme in Cells. *Journal of Biological Chemistry* 293(37):14557-14568.

149. Lat PK, Liu K, Kumar DN, Wong KKL, Verheyen EM, & Sen D (2020) High Specificity and Tight Spatial Restriction of Self-Biotinylation by DNA and Rna G-Quadruplexes Complexed in Vitro and in Vivo with Heme. *Nucleic Acids Research*:in press, 10.1093/nar/gkaa1281.

150. Hoog TG, Pawlak MR, Aufdembrink LM, Bachan BR, Galles MB, Bense NB, Adamala KP, & Engelhart AE (2019) DNA Transistors Switched by the Hofmeister Effect. *bioRxiv*:784561.

151. Swenson SA, Moore CM, Marcero JR, Medlock AE, Reddi AR, & Khalimonchuk O (2020) From Synthesis to Utilization: The Ins and Outs of Mitochondrial Heme. *Cells* 9(3):579.

152. Wilson DN (2014) Ribosome-Targeting Antibiotics and Mechanisms of Bacterial Resistance. *Nat Rev Microbiol* 12(1):35-48.

153. Hostetter AA, Osborn MF, & Derose VJ (2012) RNA-Pt Adducts Following Cisplatin Treatment of Saccharomyces Cerevisiae. *ACS Chemical Biology* 7(1):218-225.

154. Plakos K & Derose VJ (2017) Mapping Platinum Adducts on Yeast Ribosomal Rna Using High-Throughput Sequencing. *Chemical Communications* 53(95):12746-12749.

155. Rijal K & Chow CS (2008) A New Role for Cisplatin: Probing Ribosomal RNA Structure. *Chemical Communications* (1).

156. Del Valle AH, Seip B, Cervera-Marzal I, Sacheau G, Seefeldt AC, & Innis CA (2020) Ornithine Capture by a Translating Ribosome Controls Bacterial Polyamine Synthesis. *Nature Microbiology*:1-8.

157. Gesteland RF (1966) Unfolding of Escherichia Coli Ribosomes by Removal of Magnesium. *Journal of Molecular Biology* 18(2):356-IN314.

158. Stahli C & Noll H (1977) Structural Dynamics of Bacterial Ribosomes. *Molecular and General Genetics MGG* 153(2):159-168.

159. Weiss RL, Kimes BW, & Morris DR (1973) Cations and Ribosome Structure. Iii. Effects on the 30S and 50S Subunits of Replacing Bound Mg^{2+} by Inorganic Cations. *Biochemistry* 12(3):450-456.

160. Gordon J & Lipmann F (1967) Role of Divalent Ions in Poly U-Directed Phenylalanine Polymerization. *Journal of Molecular Biology* 23(1):23-33.

161. Nierhaus KH (2014) Mg^{2+}, K^{+}, and the Ribosome. *J Bacteriol* 196(22):3817-3819.

162. Rozov A, Khusainov I, El Omari K, Duman R, Mykhaylyk V, Yusupov M, Westhof E, Wagner A, & Yusupova G (2019) Importance of Potassium Ions for Ribosome Structure and Function Revealed by Long-Wavelength X-Ray Diffraction. *Nature Communications* 10(1):1-12.

163. Bray MS, Lenz TK, Haynes JW, Bowman JC, Petrov AS, Reddi AR, Hud NV, Williams LD, & Glass JB (2018) Multiple Prebiotic Metals Mediate Translation. *Proceedings of the National Academy of Sciences* 115(48):12164-12169.

164. Sinclair PR, Gorman N, & Jacobs JM (1999) Measurement of Heme Concentration. *Curr Protoc Toxicol* (1):8.3. 1-8.3. 7.

165. Zhu Y, Hon T, Ye W, & Zhang L (2002) Heme Deficiency Interferes with the Ras-Mitogen-Activated Protein Kinase Signaling Pathway and Expression of a Subset of Neuronal Genes. *Cell Growth Differ* 13(9):431-439.

166. Sankar SB, Donegan RK, Shah KJ, Reddi AR, & Wood LB (2018) Heme and Hemoglobin Suppress Amyloid B–Mediated Inflammatory Activation of Mouse Astrocytes. *Journal of Biological Chemistry* 293(29):11358-11373.

167. Wilkinson KA, Merino EJ, & Weeks KM (2006) Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE): Quantitative Rna Structure Analysis at Single Nucleotide Resolution. *Nature Protocols* 1(3):1610-1616.

VITA

Santi Mestre Fos was born in Vilafranca del Penedès, a wine region west of Barcelona, in Spain. At the age of 8, Santi and his family moved to la Garriga, also near Barcelona, and lived there for 11 years. After starting college at the University of Barcelona, he, his parents and brother moved to the United States. He continued his studies in chemistry at Kennesaw State University, near Atlanta, Georgia. There, he did organic synthesis research with Dr. John Haseltine and decided to apply for graduate school. In 2015, he was admitted in the School of Chemistry and Biochemistry at the Georgia Institute of Technology for a Ph.D. and began studying the structure and function of the enigmatic human rRNA expansion segments with professor Dr. Loren Dean Williams. During his time at Georgia Tech, he attended several scientific conferences across the US. He loves science but his true passion is cycling. Specifically, climbing mountain passes (or *cols*, in French). He has ridden in the Spanish and French Pyrenees and in the Alps. He has also been mountain biking in Utah and Colorado and has spent a lot of time riding his bike in the Appalachian Mountains.