STRUCTURE, FUNCTION AND DRUG TARGETING: RIBOSOME EXPANSION SEGMENTS

A Dissertation
Presented to
The Academic Faculty

by

Lizzette M. Gómez Ramos

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical & Biomolecular Engineering

Georgia Institute of Technology
May 2017

STRUCTURE, FUNCTION AND DRUG TARGETING: RIBOSOME EXPANSION SEGMENTS

Approved by:

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

Dr. Elsa Reichmanis, Advisor
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

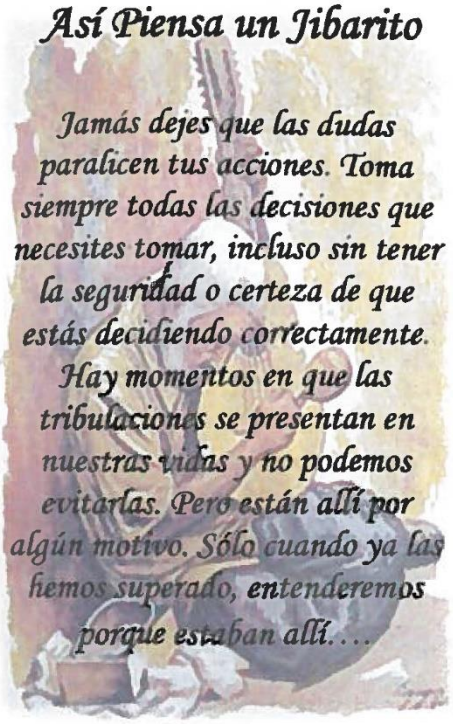
Dr. Mark R. Prausnitz
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Martha Grover
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

Dr. Mark P. Styczynski
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology

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

Date Approved: December 13, 2016



José Ramón Gómez Taboada
(1957–2013)

Para los que se fueron…
Papi, abuela Grace, abuelo Frank y José Miguel
(To those I have lost…)

Y en especial para los que siguen conmigo…
Mami, Samuel, Samuel Gabriel y Luis Gerardo
(And especially to those who are still with me…)

Con amor…
(With love…)

ACKNOWLEDGEMENTS

I would like to start by thanking my advisors, Dr. Loren D. Williams and Dr. Elsa Reichmanis, for stepping in at moments where my future looked uncertain and for providing me the opportunity to finish my doctoral degree in Chemical & Biomolecular Engineering. Your support, advice and guidance meant a lot to me. I would like to thank my committee members, Dr. Martha Grover, Dr. Mark R. Prausnitz, Dr. Mark P. Styczynski, and Dr. Nicholas V. Hud for your inquiries, recommendations and above all, your patience and understanding. I would like to extend a special "thank you note" to Elsa and Martha, for developing an atmosphere where young women in science and engineering could openly share ideas and concerns and foremost for being the best role-models for me at Georgia Tech.

I would like to thank the following organizations for financial support: Merck & Co. for the GEM Fellowship, the National Science Foundation for the Graduate Research Fellowship, the Georgia Tech University Center of Exemplary Mentoring in partnership with the Alfred P. Sloan Foundation, the Kinesis Foundation, the Gordon Research Conference for the Carl Storm Underrepresented Minority Fellowship and the School of Chemical & Biomolecular Engineering for the James P. Harris Fellowship.

I am very grateful to have been able to collaborate with many wonderful people. Jonathan, thank you for being a great mentor and turning a girl with no knowledge in biochemistry into a biochemical engineer. Paul, thanks for your friendship and the opportunity to collaborate with you in the VLP project. Johanna, Natalya and Stefany it was a pleasure to work with such friendly and dedicated ladies. To all the members in the

Bommarius and Williams lab (past and present) thank you for all the ideas and suggestions given to improve my projects; you helped me to produce a better dissertation. Sylvia, Dalar and Caleb thanks for your continuous friendship, and for all the help provided with numerical methods and transport assignments, MATLAB codes and preparation for the qualifying exams… so glad they are over! My deepest gratitude to Dr. Humberto Vega from Merck and my professors at the University of Puerto Rico at Mayagüez for inspiring me to attend graduate school.

I was fortunate to have been surrounded by many people that believed in me, especially in the many times that I did not believe in myself. Mariel, Nichole, Francisco, Ana, Adriana, Gisela, Teodoro, Dani, Nicole, Sarian, Carla, Natasha, Isa, Miner, Germán, Andrés Blanco, Titi Tanya and those I am afraid I may have forgotten to mention here, thanks for all the calls, texts, letters and most importantly for the trust you deposited in me. To all the Puerto Rican friends I made at Georgia Tech, thanks for being my home away from home.

I would like to thank my family, the most important thing in my life, for allowing me to be away for four and a half years, especially in situations in which they needed me the most. *Mami, Samuel, papi, abuela Grace, abuelo Frank, Samuel Gabriel y José Miguel* this dissertation is for all of you. This chapter of my life was completed because I wanted to make you proud. I will be forever grateful for your love, patience, understanding, and for making me the best version of myself. *Papi*, this is the completion of the book you left half written, for you, for me, for us…

Finally, I would like to thank my husband Luis Gerardo. Luis, you were with me in this journey since day 0 being my #1 cheerleader. Thanks for all your comprehension,

encouragement, trust, guidance, and especially for all your love. I could not have imagine my time at Georgia Tech without you by my side. The best of my PhD was you!

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

aaRS	aminoacyl tRNA synthetase
AspRS	aspartyl tRNA synthetase
CD	circular dichroism
CDC	Center for Disease Control and Prevention
cryo-EM	cryo electron microscopy
DAVID	Database for Annotation, Visualization and Integrated Discovery
EMSA	electrophoretic mobility shift assay
ER	endoplasmic reticulum
ES	expansion segment
ES7	expansion segment 7
ES27	expansion segment 27
$ES7_{CA}$	ES7 of *Candida albicans*
$ES7_{HS}$	ES7 of *Homo sapiens*
$ES7_{SC}$	ES7 of *Saccharomyces cerevisiae*
FD	fluorescence displacement
F-neo	fluorescein-neomycin
FMN	flavin mononucleotide
HTS	high throughput screening
IC_{50}	half maximum binding concentration
K_d	dissociation constant
LC-MS/MS	liquid chromatography tandem mass spectrometry

LSU	large ribosomal subunit
LysRS	lysil tRNA synthetase
MIC	minimum inhibitory concentration
mRNA	messenger RNA
mRBP	mRNA binding proteins
MS	mass spectrometry
NaOAc	sodium acetate
Neo	neomycin
nts	nucleotides
PA	peptidic aminosugar
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PTC	peptidyl transferase center
rDNA	ribosomal DNA
rProteins	ribosomal proteins
rRNA	ribosomal RNA
RT	reverse transcription
S	Svedbergs
SHAPE	Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension
SSU	small ribosomal subunit
SRP	signal recognition particles
T_m	melting temperature

tRNA	transfer RNA
TyrRS	tyrosyl tRNA synthetase
UBL	ubiquitin-like protein
UBC	ubiquitin-proteasome complex
UPS	ubiquitin-proteasome system

SUMMARY

The ribosome is essential to life as it transforms polynucleotides into polypeptides. In addition, ribosomes facilitate protein folding, protein degradation, stress response and mRNA surveillance. Studies on secondary and three-dimensional structures of rRNA evidenced that ribosomes of all species share a conserved common core of rRNA. Divergence in size and structure between prokaryotic and eukaryotic rRNAs and among eukaryotic rRNAs is focused on expansion segments (ESs). ESs are hypervariable eukaryotic rRNAs that emerge from the surface of the common core. In protists, ESs are important for ribosome biogenesis, rRNA processing and stabilization, translation initiation, and recruitment of non-ribosomal factors.

In this work, we investigate the functions of expansion segment 7 (ES7) of *Saccharomyces cerevisiae*, *Candida albicans* and human. ES7 is an extension of Helix 25 of the prokaryotic ribosome and is among the largest ESs of the ribosome. Thermal denaturation studies and footprinting experiments confirmed that isolated ES7 of multiple species are stable and maintain near-native secondary and tertiary conformations. *In vitro* RNA-protein pull-down experiments using ES7 as baits indicated that ES7 are binding hubs for a variety of non-ribosomal proteins that are essential to ribosomal function in eukaryotes. We hypothesize that larger rRNAs of complex organisms could confer increased functionality to the ribosome.

Pull-down experiments performed using ES7 of *S. cerevisiae* indicate that ES7-associated proteins cluster into four groups based on biological process: (i) response to abiotic stimulus, (ii) ribosomal large subunit biogenesis, (iii) protein transport and

localization, and (iv) transcription elongation. Seven amino acyl tRNA synthetases (Ala-, Arg-, Asp-, Asn-, Leu-, Lys- and Tyr-RS) appear to associate with ES7. Affinities of AspRS, TyrRS and LysRS for ES7 were confirmed by binding assays *in vitro*. The results suggest that ES7 of *S. cerevisiae* could play a role analogous to the multi-synthetase complex present in higher order organisms and that it could be important for the appropriate function of the ribosome.

Protein association to human ESs were also investigated using pull-down assays. Human ESs associate with a variety of critical proteins, including members of the ubiquitin-proteasome system. The results suggest that protein synthesis is physically coupled with protein degradation in humans. Proteins involved in metabolism, transcription and translation also associate with human ESs. In addition, human ESs contain sequences that are suggestive of potential G-quadruplex sequences. Single-stranded ES fragments show K^+-dependence of circular dichroism spectra which is generally characteristic of G-quadruplex formation.

At last, we investigated whether ES rRNAs contain useful targets for chemotherapeutics directed at eukaryotic pathogens. *C. albicans* and ES7 were used to test the general utility of ES rRNAs for small molecule targeting of eukaryotic pathogens. We demonstrated high affinity and selectivity of selected aminoglycoside-derived compounds for ES7. We also demonstrate that molecules that bind specifically and with high affinity to ES7 also induce mortality on *C. albicans,* but not in human cells. The results are consistent with our hypothesis that ESs in pathogenic eukaryotes represent useful targets for the development of new chemotherapeutics.

CHAPTER 1

INTRODUCTION

1.1 The Ribosome

The ribosome helps define life, transducing information between polynucleotides and polypeptides. Every protein in the cell is synthesized by ribosomes. Ribosomes can navigate through the cytosolic cavity or can be attached to the endoplasmic reticulum (ER); they are also components of mitochondria and chloroplasts in eukaryotic cells. Cytosolic ribosomes produce soluble proteins that remain inside the cell. ER-attached ribosomes produce membrane proteins and proteins destined for secretion (e.g. cell surface receptors and digestive enzymes).

The ribosome is composed of two subunits that assemble together, in combination with messenger RNA (mRNA), charged transfer RNA (tRNA) and protein factors, to achieve translation. The large ribosomal subunit (LSU) and the small ribosomal subunit (SSU) are composed of one or more ribosomal RNA (rRNA) chains and many ribosomal proteins (rProteins). Ribosomal subunits are commonly identified by their sedimentation coefficients, expressed in Svedbergs (S). Bacterial ribosomes are commonly composed of 50S (5S and 23S rRNAs, 33 rProteins) and 30S (16S rRNA and 21 rProteins) particles (1). Eukaryotes have larger and more diverse ribosomes in comparison to prokaryotes (2-4). *Saccharomyces cerevisiae* ribosomes are composed of 60S (5S, 5.8S and 28S rRNAs and 46 proteins) and 40S (18S rRNA and 32 proteins) particles (1). Mass ratios of rRNA to rProtein can vary significantly, from ~ 2:1 bacteria, and ~ 3:2 in chloroplast ribosomes, to ~ 1:2 in protozoa and ~ 1:3 in mammalian mitochondrial ribosomes (5).

The ribosome is catalogued as a ribozyme (6). No protein is present at the peptidyl transferase center (PTC), which is composed entirely of rRNA (7). Therefore, only rRNA catalyzes peptide bond formation in the LSU.

1.1.1 All ribosomes share a common core

The assembled *Escherichia coli* ribosome (70S particle) has been characterized by cryo-electron microscopy (cryo-EM) (8-12) and X-ray crystallography (13-14). Similarly, three-dimensional structures of bacterium *Thermus thermophilus* (15), archaean *Haloarcula marismortui* (16), fungus *S. cerevisiae* (17), common wheat *Triticum aestivum* (18) and higher order metazoans, like *Drosophila melanogaster* (19) and *Homo sapiens* (19-20), have also been determined. Compilation of structural information reinforce the idea that ribosomes share a conserved secondary- and three-dimensional structure of rRNA known as the 'common core' (Figure 1.1) (2, 4, 19, 21). Covariation allows conservation of structure at locations where sequence is totally divergent. The common core comprises regions of universal ribosome function such as the A and P sites for tRNA binding, the 5S RNA binding site, the PTC, and few additional RNA helices. Structural conservation of the ribosomal core across species allows identification, comparison and study of rRNA regions outside the common core in eukaryotic ribosomes.

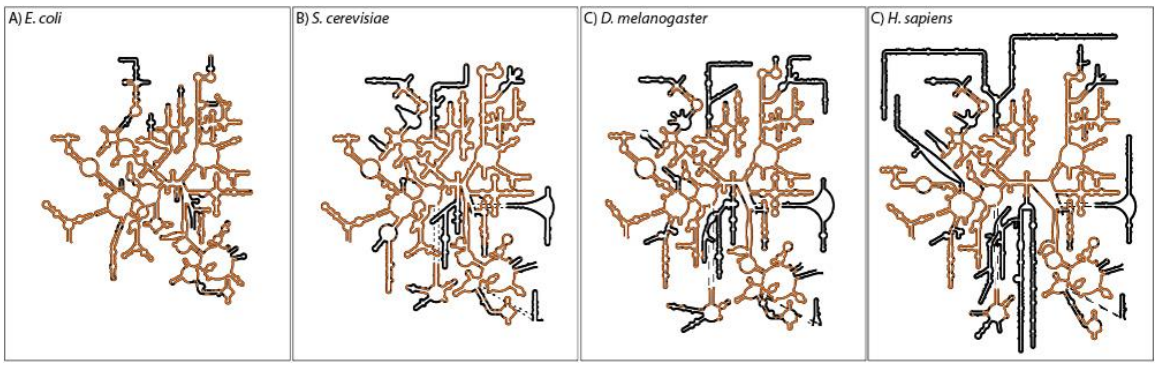


Figure 1.1 LSU rRNA secondary structures of A) *E. coli*, B) *S. cerevisiae*, C) *D. melanogaster* and D) *H. sapiens*. The common core is shown in orange. The regions outside the common core are shown in black. Figure provided by Chad R. Bernier.

1.1.2 The accretion model of ribosomal growth

Increase in rRNA size, protein quantity, and consequently, ribosomal mass, is observed over the phylogenetic tree. This increase is small for bacteria and archaea but can be substantial in eukaryotes, where ribosome sizes diverge widely (4). Comparison of *H. marismortui* and *T. thermophilus* 23S rRNAs confirmed the high degree of conservation in sequence, structure and size within the LSU of bacteria (22). However, eukaryotic rRNAs are bigger and more complex by up to ~ 40%, when compared with rRNAs of smaller organisms (comparison between *E. coli* and *H. sapiens* LSU rRNAs). Higher order species have larger LSU rRNAs: *S. cerevisiae* ~ 3554 nucleotides (nts), *D. melanogaster* ~ 4078 nts, *Gallus gallus* ~ 4725 nts and *H. sapiens* ~ 5227 nts. The rRNA of the LSU particle increased significantly due to the incorporation of hundreds to thousands of nucleotides to the 23S rRNA chain. The rRNA length of the SSU, in contrast, maintained a constant size (23).

Comparison of eukaryotic ribosomes suggest that ribosomal growth initiated in the PTC with recursive addition of sequence and structurally diverse RNA fragments. Insertion of these fragments within the ribosome occurred successively and accretion can be detected

at the atomic level (4, 24). Each accretion step enlarged the previous RNA template without perturbing the underlying structure (Figure 1.2). For this reason, many people describe the ribosome as ‘a mosaic of evolutionarily conserved structural elements interrupted by blocks or *expansions* of non-conserved sequences and structures’ (3, 25).

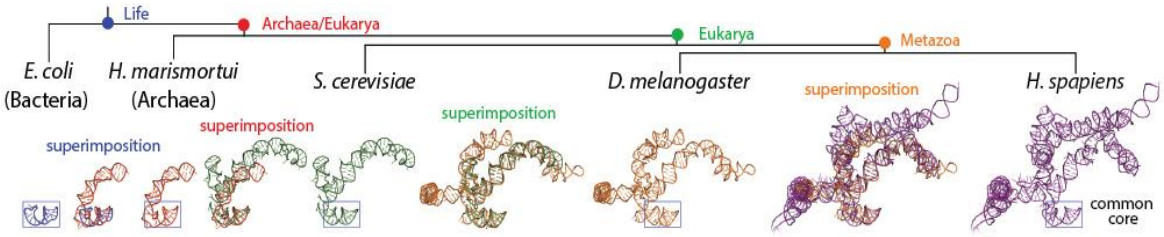


Figure 1.2 Helix 25/ES7 shows the serial accretion of rRNA onto a frozen core. This image illustrates at the atomic level how Helix 25 of the LSU rRNA grew from a small stem loop in the common core into a large rRNA domain in metazoans. Each accretion step adds to the previous rRNA core but leaves the core unaltered. Common ancestors are indicated. Pairs of structures are superimposed to illustrate the differences and to demonstrate how new rRNA accretes with preservation of the ancestral core rRNA. Each structure is experimentally determined by X-ray diffraction or cryo-EM. This figure is reproduced with permission from Petrov et al. 2014 (4).

1.1.3 rRNA expansion segments

The first complete sequences of non-mitochondrial eukaryotic rRNAs became available in the early 1980s. SSU sequences of *S. cerevisiae* (26) and *Xenopus laevis* (27) and the LSU sequence of *Saccharomyces carlsbergensis* (28) were the first to be determined. The obtention of these sequences provided the first insights on the divergence trends of eukaryotic ribosomes and how they differ from prokaryotic scaffolds, mainly composed by the common core and represented by *E. coli* or *T. thermophilus* secondary structures.

Veldman et al. noticed initially that *S. carlsbergensis* rRNA is 640 nts bigger than *E. coli* rRNA. These nucleotides are localized in 9 specific regions, rather than at random locations, described by Veldman as ‘insertions’. ‘Insertions’ could be reduced or increased without altering the protein synthesis function of the ribosome and fold as separate, individual stem-loop structures (2, 28). Evidence suggests that eukaryotic rRNA ‘inserts’ are located at regions of rRNA deletion in mammalian mitochondrial rRNAs (29-31). The absence of significant ‘insertions’ in domains V and VI of *S. carlsbergensis* in comparison to *E. coli* highlights the importance of the structural integrity of these domains for ribosomal function. Chan et al. re-named the ‘insertions’ as ‘variable’ or V-regions (23) whereas Hassouna et al. called them ‘divergent’ or D-domains (32). Clark et al. introduced the term ‘expansion segments’ (ESs) for these regions, which has been widely adopted and remains popular today (33).

At the beginning of these studies there was a lack of consistency in the nomenclature and definition of ESs. Many ESs were named and defined taking as a base the organism in which the study was being performed and following the nomenclature, if any, for that particular organism (34-35). The absence of a standardized naming system and the interchange between ‘variable regions’, ‘divergent domains’ and ‘expansion segments’ by different authors makes it hard to compare side-by-side functional or structural remarks between ESs of different organisms. One of the goals of the introductory review presented here is to adopt the terminology developed by Clark and Gerbi in 1996 and re-emphasize the importance of a universal nomenclature system that facilitates the study and deciphering of ESs within the upcoming years. However, to maintain the fidelity of the work done by our preceding colleagues the designation used in the original

manuscripts will be retained in parenthesis when their work is referenced to, using the table of comparisons made by Clark et al. (3).

1.1.4 Expansion segments: origins, size and sequence composition

As more rRNA sequences became available and comparative alignment analysis tools evolved, the heterogeneity of ESs over phylogeny became evident. GC content is more conserved between ESs of a given organism than between organisms for a given ES. Expansion segments 7 (D2) and 27 (D8) of the LSU possess GC contents ranging from 40% in slime mold *Dictyostelium discoideum* to 80-86% in vertebrates (2). In general, closely related species shared similar GC content in their ESs which suggest that nucleotide composition of ESs might be related throughout evolution (2). It is known that GC content of prokaryotic rRNAs and tRNAs is correlated with optimum growth temperature and secondary structure for these microorganisms (36-37). However, no evidence of this or a similar behavior was found in the literature for eukaryotic organisms or its ESs.

Close inspection revealed that ESs can be organized in regions of short tandem repeats (38-41). Hancock and Dover observed in 1990 that while the nucleotide composition of the common core in human LSU rRNA is mainly random, the base composition of ESs have bias towards specific trinucleotides and tetranucleotides sequences like CGG, GCG and GGC. Trinucleotide and tetranucleotide compositions vary along human ESs with some trinucleotides being more frequent in one segment of the rRNA and other trinucleotides being more frequent in another segment of the same ES. Tautz et al. suggested the existence of a mechanism able to create such sequence patterns throughout most of the genome, and proposed it to be slippage (42-43). It is therefore

hypothesized that some ESs evolved as the result of slippage-like mechanisms during replication of ribosomal DNA (rDNA). To a less extent it has been also proposed that ESs may have arisen by unequal homologous exchange between sisters chromatids (38-39).

Large shifts in base composition of ESs, in specific cases, can also occur over closely related species. ES9 (D3) of isopods possess a GC content of 32% for the species *Jaera albifrons*, while a variety of other isopods have a GC content of 54-64% for the same ES (44). Sequence composition of ESs can vary to such a degree that high frequency variation in rDNAs of human LSU rRNAs from different samples (e.g. ES15 (D6)) suggests heterogeneity of RNA transcripts (38). Dipterians *D. melanogaster* and *Chironomus thummi thummi* represent a small sub-set of organisms with predominant AT rich ESs. *C. thummi thummi* exhibits AT rich composition in at least six ESs and *D. melanogaster* exhibit it in all of its ESs (45-47).

Expansion segments typically project from apical regions of stems extending from the consensus eukaryotic core. This reference to a stem structure attached to the ribosomal core is what allows them to retain an invariant location. Boundaries are commonly delineated by their independent folding criteria (2).

The lack of size and sequence consensus between ESs has made it a challenge to determine their role in ribosome function (31). Both phenomena, the presence as well as the absence of ESs in the ribosome have shown to have no deleterious effect in protein synthesis in selected examples. Expansion segments are not involved in base-pair interactions with the rRNA of the common core and therefore have been allowed in mature rRNA during evolution (25, 32). Three different categories of rDNA ‘insertions’ have been proposed to affect the size of an rRNA gene and are based on the fate of the ‘insertion’:

introns, ESs and transcribed spacers (25, 33). The first category, the intron transcripts, cannot be tolerated by mature rRNA due to interruption of function and are spliced out of precursor rRNA. Expansion segments, the second group of 'insertions', are not removed from mature rRNA transcripts, enlarging the size of the eukaryotic ribosome. At last, internal transcribed spacers, are made up of sequences that divide distinct portions (or subunit rRNA chains) of immature rRNAs at the genetic level and are removed after transcription by processing events. Similarities such as the cryptic simplicity in sequence, the low degree of evolutionary constraint in both, ESs and transcribed spacers, and the hypothesis that ESs are degenerate transcribed spacers that cannot be excised out of rRNA, suggest that transcribed spacers and ESs may share common evolutionary origins (3, 40, 48-49).

As first pointed out by Michot et al., enlargement of LSU rRNA in eukaryotes mainly resides in ES7 (D2) and ES27 (D8) with a size variation of 600-800 nts in mammals (2). These ESs account for over 70% of the enlargement of the LSU observed in eukaryotic rRNA. Overall, ESs contain sequences responsible for more than 80% increase in the size of the human LSU rRNA over that of *E. coli* (39). Despite their large presence in the eukaryotic ribosome, many of the functions of ESs remain still unknown.

1.1.5 Expansion segments: locations

Studies performed with methidiumpropyl-EDTA·Fe (II) and EDTA·Fe (II) in the LSU and SSU rRNAs of *D. melanogaster* provided the first insights on the locations of ESs in the ribosome (50). The accessibility of ESs to cleaving reagents suggested that these

RNAs were located near or at the surface of the ribosome, at sites where they are less likely to interfere with the assembly and function of the ribosomal common core (50-51).

Comparison of cryo-EM reconstructions of rabbit reticulocyte ribosomes against *E. coli* ribosomes was performed to determine the location of the three-dimensional structures of mammalian ESs (52). The largest ES of the LSU particle and the ribosome as a whole is ES7, an expansion of Helix 25. ES7 cross-links with proteins eL13 and eL21 that lie on the solvent accessible surface of the LSU. The results from this study confirmed that several ESs were located at the surface of eukaryotic ribosomes (52).

Researchers in Japan used the 'systematic selection of functional sequences by enforced replacement' method to systematically inserted short RNA fragments into rRNAs of *E. coli*, and demonstrated the tolerance of intrinsic rRNAs to short insertions (53). Most of the functional insertions were located in the periphery of the ribosome and solvent-exposed surfaces, ordinarily avoiding rProteins binding sites. In fact, many of the insertions overlapped with ES locations in both, LSU and SSU rRNAs of *E. coli*, reaffirming the tolerance of ESs to alterations without disruption in ribosomal function (53).

Recent studies on the structure of *S. cerevisiae* at 3 Å resolution showed that the majority of the ~1.35 MDa of eukaryote-specific parts (of which 350 kDa are ESs) are located on the surface of the ribosome, wrapping and protecting the common core (17, 54). *S. cerevisiae* and *Tetrahymena thermophila* three-dimensional structures demonstrated that most of the eukaryotic specific SSU rRNA is located where the two largest ESs of the SSU (es3 and es6) strongly interact (17, 55). The same trend is observed in human and *Drosophila* SSU particles, with es3 and es6 (19). ESs distribution in the LSU is quite different. Expansion segments from *S. cerevisiae* are uniformly distributed through the

whole LSU in a ring-like conformation. *T. thermophila* ESs distribution resemble that of *S. cerevisiae,* with the majority of the ESs exposed to the ribosomal surface, except for ES31 and ES41, which occupy positions at the subunit interface (56). ESs of human and *Drosophila* are located mainly in the back and the sides of the particle with agglomerates formed near the uL1 and P stalks.

Due to the strategic location of ESs along the ribosome surface it could be considered that perhaps one of its functions is to protect the common core from proteolytic cleavage and therefore, ensure protein synthesis and cell survival. Perhaps, another function could be to confer structural stability to the overall ribosome assembly, like rProteins. Few theories suggest that ESs are dispensable, make no contribution, have no relevance towards ribosome function and therefore could not be located near functional regions where protein synthesis and decoding happens. However, we hypothesized in this work that ESs at least could be accessible for functional interactions with other macromolecules given their proximity to the ribosome surface. The limited information available on the role of ESs is discussed below.

Table 1.1 Proteins from the LSU of *S. cerevisiae* that interact with ESs, along with the helices from which ESs are derived from.

Expansion Segment	Helix	Protein interactions (57)
ES3	9	uL23
ES4	10	eL8, eL15, uL23, uL29
ES7	25	uL4, eL6, uL13, eL14, eL20, uL22, uL30, eL32, eL33
ES8	28	uL4, eL13, eL15
ES9	31	eL13, eL18, eL29, eL44
ES12	38	uL18, eL21, eL29
ES19	52	uL2, eL8, uL23
ES20	54	eL27, eL34

ES27	63	eL19
ES31	79	uL2, eL8, eL15, uL23, eL27, eL30, eL34, eL43
ES39	98	uL3, eL6, uL6, uL13, eL14, eL20, uL22, eL33
ES41	101	uL3, eL24, eL31
ES43	18	uL24, uL29, eL37

Table 1.2 Proteins from the SSU of *S. cerevisiae* that interact with ESs, along with the helices from which ESs are derived from.

Expansion Segment	Helix	Protein interactions (57)
es3	9	eS4. eS8, uS17
es6	10	eS4, uS4, eS7, uS8, uS15, uS17, eS24
es7	26	eS1, uS15, eS26, eS27
es9	39	uS9, uS10, eS19
es10	41	uS3, uS7, uS9, uS10, uS13, eS14, es19, eS25
es12	44	eS6, eS8, uS12, eS30

Table 1.3 Proteins from the LSU of *H. sapiens* that interact with ESs, along with the helices from which ESs are derived from.

Expansion Segment	Helix	Protein interactions (57)
ES3	9	uL23
ES4	10	eL8, eL15, uL23, uL29
ES7	25	uL4, eL6, uL13, eL13, eL14, uL15, uL18, eL20, eL21, eL28, uL30, eL32, eL33
ES8	28	uL4, eL13, eL15, uL24
ES9	30, 31	eL13, uL15, eL18, eL29, eL36
ES10	38	uL30
ES12	38	uL18, eL29
ES15	45	uL4, eL6, eL18, eL28, uL30, eL32
ES19	52	eL8, uL23
ES20	54	eL27, eL34
ES27	63	eL19, eL27
ES30	78	uL1
ES31	79	uL2, eL8, eL15, eL27, eL34, eL43
ES39	98	uL3, eL6, uL6, uL13, eL14, eL20, eL33
ES41	101	uL3, eL24, eL31
ES43	18	uL24

Table 1.4 Proteins from the SSU of *H. sapiens* that interact with ESs, along with the helices from which ESs are derived from.

Expansion Segment	Helix	Protein interactions (57)
es3	9	eS4, eS8
es6	21	uS4, eS4, eS7, uS8, uS15
es7	26	eS1, uS8, uS15
es9	39	uS9, uS10
es10	41	uS3, uS7, uS9, uS10, uS13
es12	44	uS5, eS8, uS12, eS30

1.1.6 Expansion segments: functions

1.1.6.1 Expansion segments analysis of function by nucleotide alteration, addition or deletion

In the late 1980s, the first studies on genetic engineering of ESs were performed. As a result, initial insights on the functions of ES became available. The first system for *in vivo* mutational studies of eukaryotic rDNAs was developed by Yao and co-workers and consisted in transforming conjugated *T. thermophila* cells with mutated rDNA (58-59). Sweeney et al. performed experiments in which cloned micronuclear rDNA of *T. thermophila* was altered by the insertion of random DNA segments into coding and non-coding regions, including ESs (58). The most striking finding of their work was the ability to produce apparent functional ribosomes possessing normal doubling times in the presence of mutated 26S rRNA containing a 119 nucleotide linker in an ES near the 3' end (58, 60). These results provided the first indication that the primary structure of ESs does not have to be fixed to produce functional ribosomes.

Later on, an 18-mer tag was attached to ES43 of *S. cerevisiae* 26S rRNA to perform mutational studies on yeast rDNA (61). Observations from this study revealed that the concise structure of ES43 in *S. cerevisiae* is not essential for ribosome biogenesis or

function. However, the possibility of secondary effects on maturation kinetics could not be discarded. Similarly, an experiment designed to monitor the effects of rDNA mutations in the 40S subunit of *S. cerevisiae* was performed by the addition of a 19 nucleotide tag to es9 (V8) (62). Analogous to the 26S rRNA tag incorporation, addition of this tag to es9 in 17S rRNA did not produce signs of harmful effects in 40S subunit formation or function. Therefore, it was concluded that the structure of es9 in the SSU is not of critical importance to the ribosome. Still, later experiments using an alternative *in vivo* system suggested that incorporation of the tag into es9 has lethal effects in few essential ribosomal functions, being translation and termination the ones with most probability (63). Additional attempts to insert two different oligomer tags into es3 (V3) of 17S rRNA turned out to be deleterious to the formation of small ribosomal subunits, although the 37S pre-rRNA and the LSU production remained unaffected, indicating that not all ESs of mature rRNAs have structural flexibility (62).

Further studies were performed by Musters et al. to analyze the structure-function relationship of ES19 (V9) of yeast 26S rRNA. ES19 is located in domain III of the LSU and forms part of the conserved binding site for ribosomal protein uL23. Its counterpart in *E.coli* is composed of only 3 nucleotides. In these studies ES19 was exchanged with equivalent regions from other species or excised by *in vitro* loop out mutagenesis (64). The results indicated that both, the presence or absence of ES19, has no detectable deleterious effects on the function of *S. cerevisiae* LSU and therefore it is a dispensable region (64). However, later experiments observed a two-fold-reduction in the growth of cells expressing only mutated rRNA (65).

Similar to Musters et al., experiments were performed in which ES27 (D8) of *T. thermophila* was either removed or replaced by a variety of related sequences (60). Opposite to what was observed with ES19 of yeast, the results indicated that ES27 is essential and plays a role in either rRNA processing or stabilization of mature rRNA pieces of *T. thermophila*. Additionally, ES27 of *T. thermophila, S. cerevisiae, D. discoideum,* and *Caenorhabditis elegans* can be exchanged without any lethal effect for *T. thermophila,* being hypothesized that the tertiary structure of rRNA might be the reason for this behavior.

Further studies on ESs from yeast revealed that helix 6, which forms a considerable portion of *S. cerevisiae* V1 region (no equivalence in nomenclature could be established) in the SSU, is important in maintaining normal production of 18S rRNA as well as its correct assembly into the 40S subunit (66). Augmenting the helical or loop portions of helix 6 strongly diminished the production of 18S rRNA possibly due to steric inhibition with protein uS15, an essential rProtein (66). However, almost complete deletion of this helix has less severe effects since significant amounts of accurately processed mutant 18S rRNA still accumulate and cells containing this mutant continue to grow.

As it can be observed the majority of genetic studies on ESs mentioned so far target the addition or alteration of nucleotides into existing ES scaffolds. The disadvantage of this approach arises from the high rates of false positives that could relate ribosome impairment (caused by disturbance to rRNA structure by nucleotide addition) to the essentiality of these regions in the ribosome. For this reason, ES7 (V3) and ES27 (V13) of *S. cerevisiae* LSU were subjected to study by deletion analysis (67). Removal of a portion of Helix 63 from ES27 resulted to be dispensable for ribosome biogenesis and function (67). The results suggested that the lethal phenotype observed in *T. thermophila* by nucleotide deletion in

ES27 (60) is not due to the destruction of an essential functional element, but rather to disturbances in pre-rRNA processing or steric hindrance with rProteins (67). Almost complete elimination of *S. cerevisiae* ES27 to mimic its prokaryotic counterpart proved to be lethal. Deletion of ES7 in the same study reveals that partial removal of Helix 25 caused severe growth defects while its whole deletion, as ES27, was lethal. Metazoans contain ES30 and ES43, segments not present in yeast and *Tetrahymena* (19). This adds up additional complexity as to whether all or only few expansion segments are essential for cell survival.

1.1.6.2 Expansion segments analysis of function by studying translation mechanisms

Cryo-EM data of *Trypanosoma cruzi* reveals unusual structural components in the secondary structure of ESs (68). Fragments of es6 and es7, form part of a large helical structure termed “the turret” located in the most lateral side of the 40S subunit and present only in Trypanosomes. The turret structure is believed to be involved in the unique translation initiation mechanism of Trypanosomes along with the spliced leader sequence and cap 4-structure at the 5’ end of mRNA (68). Otherwise, ES19 (D7a) of *C. thummi thummi* its been known to be involved in the rRNA processing of insects (47). Both, the primary and secondary structures of ES19 in dipterans, seems to be essential for the cleavage of 26S rRNA (47).

Interaction studies between ES24 of yeast and the Sec61 translocon suggests that ES24 can play a role in recruiting and coordinating the access of non-ribosomal factors such as chaperones, modifying enzymes or signal recognition particles (SRP) (69). Further studies to investigate the degree of conservation of ES structures in fungi were performed

using *Thermomyces lanuginosus* as a model, and compared against data obtained for *S. cerevisiae* (70). Although numerous changes were detected in the primary structure of ESs, the most dramatic conformational differences were located in ES7, ES27 and ES39 of 25S rRNAs. Only the base of ES27 was visible in the cryo-EM structure of *T. lanuginosus* suggesting that the allosteric mechanism of *S. cerevisiae* for this helix is also shared by this thermophile (70). Moreover, conformational changes in ES7 and ES39 caused by ribosomal protein uL4 in *T. lanuginosus* are hypothesized to influence the binding of non-ribosomal factors, such as SRP, that are located close to these rRNA elements and therefore be of functional importance for the regulation of the peptide binding site (70). The tip of ES7 in *S. cerevisiae* is believed to interact with protein P0, a component of the P-stalk that recruits G proteins into the ribosome (17, 71). Since the same set of core rProteins exists in both fungi, deciphering of ESs functions in one species can be easily transferred to the other one.

Human and Drosophila ESs are also involved in the recruitment of external, non-ribosomal factors. es3 and es6 of the SSU contribute to the binding of eukaryotic initiation factors eIF3 and eIF4-G for eukaryotic translation (19, 72-74). Structural variations within these segments and the absence of this behavior in other eukaryotic organisms reflects the difference in eukaryotic translation mechanisms evidenced previously by the Trypanosomes' turret structure. Dynamic rearrangements between ES27 and ES31 in *D. melanogaster* in the presence of eukaryotic elongation factor eEF-2 seems to enable communication between the mRNA exit site in the SSU and the tunnel exit site in the LSU (19). This behavior is not observed in human ribosomes and evidence the variety of functions that could be acquired by ESs in phylogeny.

1.2 Antibiotic/Antimicrobial Resistance

1.2.1 Antibacterial targeting, resistance and development

The discovery of penicillin against gram-positive bacteria by Alexander Fleming marked a historical landmark towards the treatment of infectious diseases (75). The quick development of penicillin derivatives along with the discovery of additional classes of compounds possessing antibacterial activity (i.e. sulfa drugs, aminoglycosides, chloramphenicols, macrolides, tetracyclines, quinolones, etc.) led to the golden era of antibiotic discovery between the 1940s and 1960s. However, the rapid rate of success of antibiotic development was overshadowed by the emergence of resistance. Bacterial resistance against antibiotics emerged as early as the 1940s. Studies on microorganism susceptibility against penicillin demonstrated that *Staphylococcus* produced an enzyme with the ability to minimize the treatment efficiency provided by penicillin (76). As this, there are many other examples where antibiotic resistant bacterial strains have emerged, reaching alarming levels (i.e. tetracycline resistant *Shigella*, methicillin resistant *Staphylococcus*, penicillin resistant *Pneumococcus*, gentamicin resistant *Enterococcus*, vancomycin resistant *Enterococcus*, etc.).

The Center for Disease Control and Prevention (CDC) estimated in 2013 that approximately 2 million people each year become infected with antibiotic resistant bacteria and 23,000 died as a consequence of these infections (77-78). An assessment conducted by this entity concerning antimicrobial resistant threats revealed a list of 18 microorganisms, mostly bacteria, that represent either urgent, serious or concerning threats to the public health. Antibiotic resistance in these organisms at large scale emerged from multiple reasons: incorrect prescribed drugs to treat non-bacterial infections, incorrect dosing,

spread of resistant-strain bacteria from person to person, antibiotic use in food animals, among others (79).

Strictly speaking the simple use of antibiotics creates resistance. The introduction of drugs to susceptible bacterial environments increases the selective pressure of bacteria to acquire resistance against it. Bacteria will have to adapt to the new environment or otherwise die. It has been suggested that the simplest method to fight bacterial resistance is to stop taking antibiotics (80). Many drug resistance mechanisms confer a fitness cost to bacteria as essential functions in bacterial cells are altered or impaired (81). This cost is variable and influences the ability of bacteria to maintain or change the rate of resistance (82-84). Fitness costs can be overcome by compensatory mutations, true reversion and other mechanisms and may differ between *in vitro* and *in vivo* studies making it difficult to study the real fitness cost on resistance.

At the molecular level, bacterial resistance is caused by two different mechanisms: intrinsic or acquired (85). Intrinsic mechanisms are defined as those mechanisms that allow bacteria to resist the antibiotic action as a result of inherited structural or functional characteristics (e.g. inability to cross bacterial cell wall). Acquired mechanisms, otherwise, are developed by bacteria to by-pass the effective action of the antibiotic. Three big groups were established by Blair and co-workers for acquired bacterial resistance mechanisms: minimization of antibiotic intracellular concentration by reduced drug permeability or increased efflux, target site changes *via* mutation, overexpression or chemical protection and antibiotic inactivation by hydrolysis or steric hindrance. Multiple cases have been encountered for all groups and the challenges for all of them are evident.

Development of novel antibacterials represent a scientific and economic challenge. The lack of new platforms for antibacterial discovery and the increasing costs in the discovery pipeline for new compounds make the development process even harder. Few new classes of antibacterials have been approved after 1960 (86). Rational drug design of molecules with new mechanisms of action have not been successful, mostly by the inability to penetrate the bacterial envelope (87). The Lipinski Rule of 5 used as a standard to predict oral bioavailability, have no counterpart in antimicrobial discovery. Attempts to adhere to these rules most likely have precluded antibiotic discovery as most current approved compounds do not obey all the parameters of the Lipinski Rule of 5 (87-89).

Antibacterial drugs developed to date target the following areas (90):

i. Inhibition of cell wall biosynthesis – β lactams, polypeptides, glycopeptides, phosphonic acid derivatives, D-alanine analogs
ii. Interference of cell membrane integrity– lipopeptides, diarylquinolines, polymyxins
iii. Inhibition of protein synthesis – aminoglycosides, chloramphenicols, macrolides, ketolides, tetracyclines, fusidane, oxazolidinones, streptogramins, monoxycarbolic acid
iv. Inhibition of nucleic acid biosynthesis – rifamycins, quinolones, macrocyclic, nitrofurans, nitroimidazoles
v. Inhibition of metabolic pathways – sulfadrugs, trimethoprim

Different approaches has been suggested to tackle the discovery of new class of antibacterials and combat antibiotic resistance. Development of new compounds for old targets have been suggested (90). This approach will require an intense search for sources

where new compounds can be obtained. A different approach could involve chemical modifications to already existing compounds (91). It is expected that the risks associated with safety and toxicity in derivatives will be lower than the ones present in new classes of compounds. The same strategy was used for the development of penicillin derivatives back in the 1940s.

Combination therapies have been also proposed to fight antibiotic resistance (92). Currently many oncologic treatments use this approach. The treatment of infections with multiple, rather than individual drugs will make it harder for bacteria to develop resistance as the probability to develop resistance against both drugs combined will be less than the probability of developing resistance against each antibiotic individually.

Restoration of natural product drug discovery programs is probably one of the most promising platforms. Most of the currently used antibiotics were completely or partially derived from natural sources with very few exceptions (sulfonamides, quinolones, oxazolidinones, nitrofurans) (93). Marine natural products represent a good source for the development of new antibiotics as they are rich in secondary metabolite production (94). The same is true for plants, which are able to liberate toxins against grazing by animals. Marinomycins A-D, four different macrodiolides extracted from ocean microorganisms, showed significant antibacterial activities against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faceium* (95). The discovery of these compounds re-emphasize the potential of new sources of natural products for drug discovery.

At last, the development and screening of new targets could also lead to the discovery of new compounds. Recently, a group of scientists at Merck Research

Laboratories discovered ribocyl, a synthetic analog of flavin mononucleotide (FMN) that competes with the natural ligand to inhibit FMN riboswitch-mediated expression and inhibit bacterial growth (96). Bacteria, fungi and plants synthesize FMN or obtain it from their environment, while humans only acquire it through their diet. The target-specific effect obtained in this case could be mimicked to find others able to alter bacterial metabolism. Most importantly this also represent an example of how the space of antibacterial targets could be expanded to account for new candidates, as non-coding RNAs, to combat infections.

1.2.2 Antifungal targeting, resistance and development

Fungal pathogens are one of the leading causes of infectious diseases within the immunocompromised population. Unfortunately, analogous to antibacterials, resistance against antifungals have also been observed. The antibiotic resistance handbook published by the CDC in 2013, focused mainly on bacterial infections, also includes *Candida* as a serious threat due to the existence of numerous drug-resistant strains. Information on fungal infections, mechanisms of resistance and drug development is limited when compared with the amount of information available for bacterial infections. For this reason information on bacterial resistance, antibacterial targets, mechanisms of resistance and new antibacterial discovery schemes were covered first in this literature survey, as comparisons between bacterial and fungal antimicrobials have been performed, and are not uncommon due to the disparity of available information.

Similar to bacteria, anti-fungal development started in the 1930s with the introduction of griseofulvin, the first oral antifungal drug (97). Griseofulvin was isolated

from *Penicillium griseofulvin* and used to treat skin, hair and nail infections caused mainly by *Trichophytons*. Posterior to griseofulvin, several classes of antifungals were developed (i.e. polyenes, flucytosine, echinocandin, imidazole, triazoles, allylamines, etc.) leading to the golden era of antifungal discovery between the 1950s and 1980s. An increasing concern of fungal diseases to public health has also become evident. Drug-resistant strains continue to emerge and very few antifungals have made it to the market since the 1980s.

Analogous to bacteria, antifungal resistance can be intrinsic or acquired. However, a third class of antifungal resistance, described as clinical resistance, has also been observed (98-99). Clinical resistance has been identified in patients with latent or progressive infections, in which an isolate of the fungi demonstrates susceptibility towards anti-fungal drugs in laboratory tests *in vitro*. Successful responses to antimicrobial therapies are not exclusively pathogen dependent and the host immune system along with the drug pharmacodynamics and pharmacokinetic profiles, location of infection, duration of the treatment, among others may contribute to the resistance acquisition, especially with fungi (100). Similar to bacteria, acquired resistance can be obtained through increased efflux, target-site alterations, up-regulation of target enzymes or development of by-pass pathways (99-100). Still, decreased cell permeability, drug modifications or combinations of these mechanisms have not been identified as mechanisms of resistance against antifungals (101).

Since both, yeast and human, are eukaryotic organisms closely related in the phylogenetic tree of life it is a challenge to design compounds with the desired specificity and pharmacological profiles. Antifungals developed to date target the following areas (90, 102):

i. Disruption of cell membrane – polyenes

ii. Inhibition of cell membrane biosynthesis pathways – azoles, allylamines, thiocarbamates

iii. Inhibition of cell wall biosynthesis pathways – echinocandins

iv. Inhibition of nucleotides synthesis – pyrimidine analogues

v. Cells – hydroxypyridones, benzofurans, phenyl ethers, sesquiterpenes

Comparison between antibacterial and antifungal drug targets revealed that many antibacterials possess no antifungals analogues. Inhibition of protein synthesis and metabolic pathways in bacteria have no counterpart in fungi. The structural and sequence resemblance between fungal and human ribosomes have minimized the efforts of scientists to screen and develop drugs against these targets. Sordarins constitute the only known exception of anti-fungals developed to inhibit ribosomal translation by interacting with elongation factor 2 (EF-2) (103). Still, there is no evidence in the literature of drug approval.

Different strategies have been proposed to overcome antifungal resistance. Natural products and its derivatives represent a good source for antifungal discovery. Similar to antibacterials, the majority of antifungals have been derived from nature with very few exceptions (i.e. azoles, allylamines). Fitness tests profiling allowed the discovery of a variety of natural products with diverse targets: biosynthesis of sphingolipids, fatty acids and proteins, 3' mRNA processing, ribosome biogenesis, nuclear export, protein translocation, etc. that could be developed into potential antifungals (104). Identification of new, target-specific sites have also been suggested. Several groups have worked successfully to screen and identify genes specific to *C. albicans* growth and survival for

drug target development (105-108). Every living cell require amino acids in order to maintain its function and methionine and threonine biosynthesis have emerged as exploitable pathways (109). This approach most likely can be applied to the development of new antibacterials as well.

Increased anti-fungal dose intensity, development of new drug delivery systems to improve the therapeutic index of existing antifungals, combination therapies and surgery for sequestered lesions emerge as additional proposals (110). Detailed comparison of data between antibacterial and antifungals operating mechanisms will provide more insight in the development of novel drug discovery platforms and how to overcome the mechanisms of resistance.

1.3 Thesis overview

This thesis studies the stability, structure and function of rRNA expansion segments in eukaryotic organisms and its role as antifungal targets. Chapter 2 investigates the role of *S. cerevisiae* ES7 in the ribosome and identifies direct ES7-binding partners through RNA-protein pull-down experiments. Chapter 3 characterizes human ES7 and compares the functions of various human ESs across different cellular environments. This chapter also investigates the existence of G-quadruplexes in the human ribosome. Chapter 4 examines *C. albicans* ES7 as a target for antimicrobials. Chapter 5 provides the overall conclusions of this work and discusses recommendations for future research.

CHAPTER 2

YEAST rRNA EXPANSION SEGMENTS: FOLDING AND FUNCTION

2.1 Introduction

The core function of the ribosome is to read mRNA and synthesize coded protein. In addition, ribosomes facilitate protein folding, protein degradation, stress response and mRNA surveillance (111-113). The increased size of ribosomal RNAs (rRNAs) in complex organisms appears to confer increased functionality to the ribosome (4). Many auxiliary functions of large ribosomes are yet to be characterized. However, auxiliary functions correlate with rRNA size, which follows well-defined patterns:

(i) All cytoplasmic ribosomes contain a "common core" of rRNA, which is reasonably approximated in structure by prokaryotic rRNA (Figure 2.1).

(ii) Expansion of rRNA beyond the common core is focused at few specific regions on the ribosome called "expansion segments" (ESs) (3, 21, 25, 32-33).

(iii) ESs are located on the surface of the ribosome and are excluded from the ribosome interior and from core functional centers such as the peptidyl transferase center, the decoding center, the subunit interface, and tRNA binding sites (114).

(iv) Size variation is significantly greater in eukaryotes than prokaryotes and in large subunit (LSU) rRNAs than in small subunit (SSU) rRNAs (3).

(v) rRNA size increases from bacteria/archaea to protists to metazoan and is greatest in mammals (4, 21).

(vi) rRNA size correlates with organismal complexity (4).

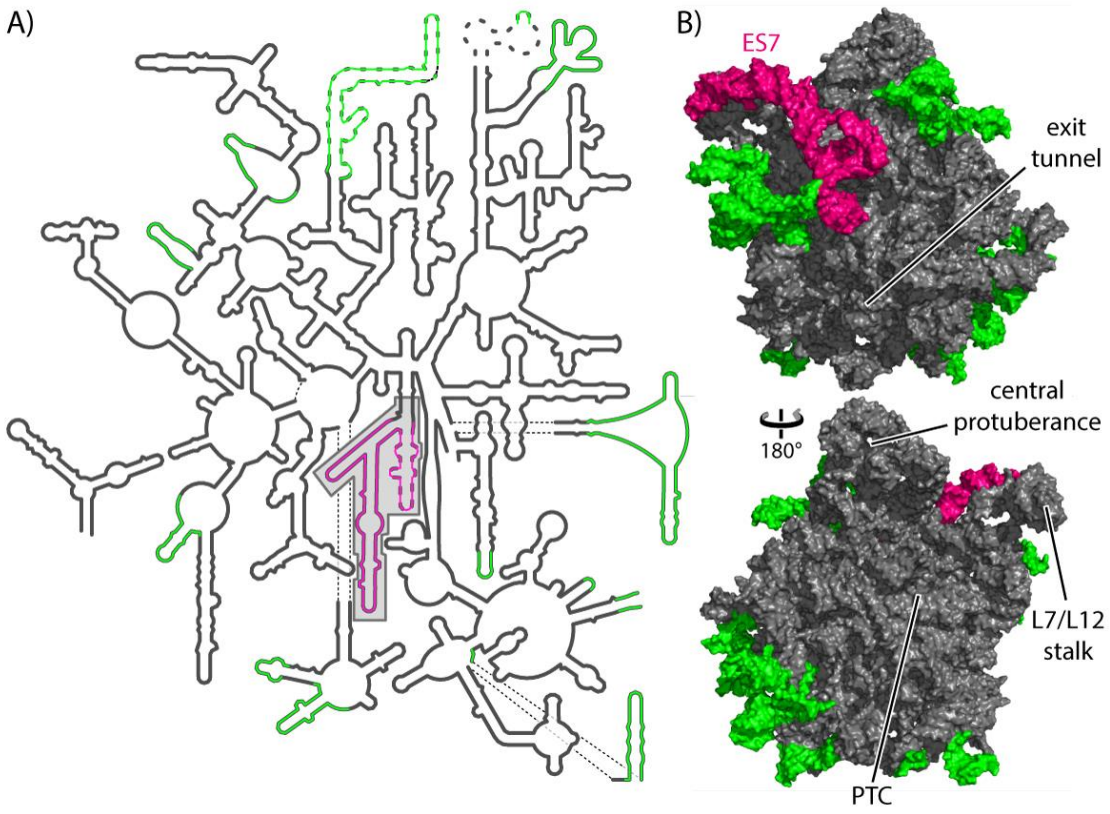


Figure 2.1 The LSU rRNA of *S. cerevisiae* with highlighted expansion segments. A) Secondary structure and B) three dimensional structure of LSU rRNA. Common core rRNA, which is common to all cytoplasmic ribosomes, is dark gray. Expansions of *S. cerevisiae* rRNA from the common core are green, except for expansion segment 7 (ES7) which is pink and enclosed in a shaded box in panel A. For panel B, the top shows the back view of the LSU, and the bottom shows the front (interfacial) view. Locations of structural hallmarks such as the peptidyl transferase center, the central protuberance, stalk and tunnel exit are indicated. Thick dashed lines in panel A represent RNA that is not observed in the three dimensional structure (114) and is absent in panel B.

Here we investigate the functions of expansion segment 7 (ES7, Figure 2.1) of *Saccharomyces cerevisiae*. ES7 is an extension of Helix 25 of the prokaryotic ribosome and is among the largest expansion segments of the LSU and the ribosome as a whole. ES7 is one of the most variable regions of eukaryotic rRNA. It is 210 nucleotides (nts) in *S.*

cerevisiae (Figure 2.2), 341 nts in *Drosophila melanogaster*, 800 nts in *Gallus gallus* and 876 nts in *Homo sapiens*. The locations of ESs on the surface of the ribosome were established by Dervan and coworkers (50) and confirmed by three-dimensional structure determination (reviewed in reference 21).

The ribosome has grown in size and complexity within certain branches of the eukaryotic tree. The LSU rRNA of *S. cerevisiae* is 3554 nts, while that of *D. melanogaster* is 4078 nts, *G. gallus* is 4725 nts and *H. sapiens* is 5227 nts. It has been proposed that ESs are involved in ribosomal regulation, packaging, turnover and biogenesis (60, 115-116). In trypanosomes, ESs of the small ribosomal subunit contribute to a 'turret-like' structure, linking the SSU and the LSU (68). The location of the turret suggests a role in translation initiation (117).

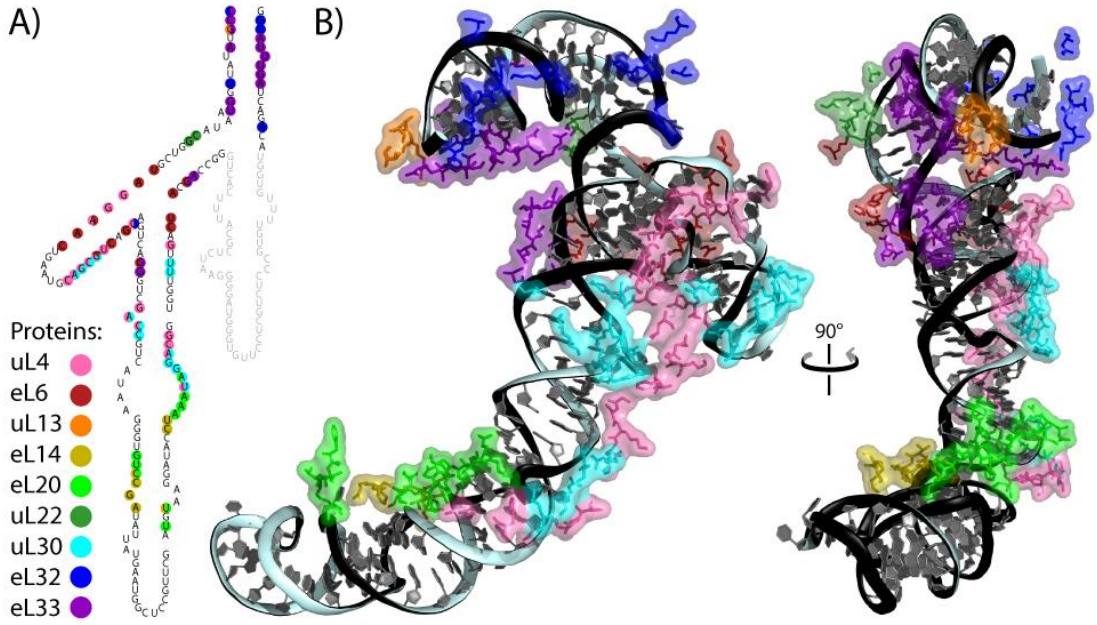


Figure 2.2 ES7 rRNA interacts with ribosomal proteins in the assembled ribosome. A) Secondary structure and B) three dimensional structure of *S. cerevisiae* ES7. Interactions of ribosomal proteins with ES7 are mapped on to the A) secondary and B) three-

dimensional structures. In panel A, nucleotides are colored by the protein with which they interact. Nucleotides not present in the three dimensional structure are shown in gray. In panel B, amino acids interacting with ES7 rRNA are colored by the proteins to which they belong. Coordinates are from PDB entry 4V88.

We characterized the secondary structure and protein binding capabilities of ES7. The results are consistent with a model in which expansion segments of eukaryotic ribosomes are docking sites for protein factors that are auxiliary to core ribosomal functions. These factors include aminoacyl tRNA synthetases (aaRSs), quality control proteins and chaperones.

2.2 Materials & Methods

2.2.1 *S. cerevisiae* ES7 DNA and RNA

The DNA encoding *S. cerevisiae* ES7 was amplified from yeast genomic DNA. The T7 promoter and restriction sites for EcoRI (5' end) and HindIII (3' end) were added *via* PCR. The DNA was inserted into a previously restricted, dephosphorylated pUC19 plasmid. Ligated plasmid was transformed into DH5α competent cells and colonies were selected based on X-gal Blue-White screening (Thermo Scientific). Primer sequences are shown in Table A.1.

The Weeks cassette (118) was incorporated onto the 3' end of ES7 to allow SHAPE mapping without losing information from ES7 termini using the Q5® Site-directed Mutagenesis Kit (New England BioLabs). Primer information is described in Table A.2. Primer synthesis and DNA sequencing were performed by Eurofins-MWG Operon.

2.2.1.1 *In vitro* RNA synthesis

S. cerevisiae ES7 RNAs were transcribed (HiScribe™ T7 High Yield RNA Synthesis Kit, New England BioLabs) with the following modifications to the manufacturer's protocol. pUC19 containing the *S. cerevisiae* ES7 gene was linearized at the 3' end with HindIII. In each 20 µL transcription reaction, 500 ng of DNA template were used. The DNA was transcribed overnight at 37 °C followed by incubation with TURBO DNase (Ambion) at 37 °C for 15 min. A 1:10 dilution of a 5 M ammonium acetate solution (pH 5.6) and 2.5 vol of 100% ethanol were added to the reaction mixture, which was then incubated at -20 °C for 1 h and centrifuged at 15,600*g* for 10 min at 4 °C. Pellets were washed three times with 80% ethanol before drying *via* Speedvac. RNA was resuspended in nuclease-free water and purified from unincorporated nucleotides with an illustra NAP 5 column (GE Healthcare). RNA yields were quantified by UV absorbance. RNA purity and integrity were verified by 6% acrylamide (29:1 acryl:bisacryl), 8 M urea gel with TBE buffer.

2.2.2 SHAPE reactions

Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) methods were adapted from published protocols (118-119). *In vitro*-transcribed RNA in nuclease-free water was added to folding buffer [final concentrations: 200 mM NaOAc, 50 mM Na-HEPES (pH 8.0) and 1 mM 1,2-trans- cyclohexane-N,N,N'N'-diaminetetraacetic acid or 5 mM of $MgCl_2$] to obtain 400 nM RNA in 80 µL. RNAs were annealed by cooling from 75 °C to 25 °C at 1 °C/min. Data were collected in 200 mM Na^+ and no Mg^{2+} to analyze secondary structure (120-122). RNA tertiary interactions were characterized in the presence of Mg^{2+} ions.

2.2.2.1 Benzoyl cyanide reaction

Chemical modification of ES7 RNA was performed with a 10X dilution of a 1000 mM benzoyl cyanide (Sigma) stock prepared in dimethyl sulfoxide (DMSO) into 36 µL of annealed RNA. Control reactions contained DMSO only. Reactions were carried out for ~1-2 min at room temperature (123).

Benzoyl cyanide-modified RNA was purified from reaction mixtures by precipitation in ammonium acetate [1:10 dilution, 3 M (pH 5.2)] and ethanol (2.5 vol) at -20 °C for ~15 min. Reaction mixtures were centrifuged at 15,600*g* at 4 °C for 10 min and the pellet washed once with 80% ethanol before drying *via* Speedvac. Pellets were re-suspended in 22 µL 1X TE buffer [10 mM Tris-Cl, 1mM EDTA (pH 8.0)]. Recovery after purification was > 40%.

2.2.2.2 Reverse transcription of benzoyl cyanide-modified *S. cerevisiae* ES7 RNA

A single [6-FAM]-labeled primer (Table A.3) provided complete SHAPE reads on ES7 of *S. cerevisiae* (124). Modified RNA (20 µL) was added to 40 pmol primer in reverse transcription (RT) buffer to yield 1X First-Strand Buffer (Invitrogen), 2 mM DTT and 625 µM of each deoxynucleotide (dNTP) in 50 µL. Primer annealing was performed stepwise: 95 °C for 30 sec, 65 °C for 3 min, 60 °C for 3 min and 4 °C for 10 min. SuperScript III Reverse Transcriptase (Invitrogen) was used in RT reactions. RT mixtures were pre-incubated at 52 °C for 2 min before addition of 1 µL (200 U) of RT enzyme and reaction was allowed to proceed for 2 h at 52 °C. RT enzyme was inactivated at 70 °C for 15 min. Reverse transcription reactions were also performed on un-modified RNAs (DMSO only) at the conditions used for modified RNA samples.

RNA was sequenced by RT/chain termination using dideoxynucleotides (ddNTPs) at a ratio of 1:1 ddNTP to dNTP. Ten pmol of ES7 RNA in TE buffer and 40 pmol of primer were used for sequencing reactions. A reaction without ddNTPs identified natural polymerase fall-off peaks.

2.2.2.3 Capillary electrophoresis of RT reaction products

Analysis of RT reaction productions was performed as described previously (124).

2.2.2.4 SHAPE data processing

SHAPE data were processed as described (125). SHAPE data were mapped onto secondary structures with the program RiboVision (126).

2.2.3 RNA folding/unfolding

Melting curves for ES7 RNA were determined by absorbance at 260 nm as a function of temperature with a thermostated Varian Cary-1E UV spectrophotometer. RNA was concentrated by Speedvac to at least 1 µg/µL and added to melting buffer [180 mM NaCl, 20 mM Tris-HEPES (pH 8.0)] to reach an initial absorbance of 0.30-0.33 at 260 nm. Samples were heated from 15 °C to 95 °C at 3 °C/min in the spectrophotometer to unfold the RNA. Melting transitions were obtained by repeated heating and cooling at 0.5 °C/min. Due to the RNA size the temperature ramp was less than 1 °C/min, which is commonly used for melting short duplexes (127). Superimposition of heating and cooling curves was used to confirm equilibrium during melting. Thermodynamic model information as well as information on fitting and parameter estimation are described in Appendix A.

2.2.4 Pull-down assays and LC-MS/MS analysis

The DNA anchor was the same sequence as the SHAPE primer, modified with a biotin label on the 5' end (Table A.3). ES7 RNA (1.3 µM) was annealed to 2 µM of the DNA anchor in melting buffer with the annealing conditions described for SHAPE experiments. SoftLink™ Soft Release Avidin Resin (Promega) was washed three times with lysis buffer [0.1% (w/v) sodium deoxycholate, 100 mM NaCl, 1 mM EDTA, 50 mM Tris-Cl (pH 8.0) and 1X protease inhibitor cocktail containing 1 mM AEBSF and 1.54 mM Aprotonin] also containing 10 U of RNAse inhibitor, 5 mg/mL of heparin (Fisher Scientific) and 0.5 mg/mL of yeast tRNA (Roche) at 4 °C. Posteriorly, the RNA-biotin assembly was shaken gently at room temperature with 10 µL of equilibrated Avidin resin to promote the attachment of ES7 to the resin.

Pull-down assays were performed with *S. cerevisiae* cell lysates. One feature of these experiments that distinguishes them from conventional pull down assays is the high abundance of the competing rRNA within the lysate. It seems likely that intrinsic competition from endogenous rRNA might attenuate non-specific binding.

A starter culture of *S. cerevisiae* was grown in yeast extract-peptone-dextrose (YPD) media overnight at 30 °C and used to inoculate a 50 mL culture. Cells were harvested when they reached an optical density at 600 nm (OD600) of ~1.0 by pelleting at 770*g* and 4° C for 15 min. Pellets were stored at -80 °C until use. Cells were resuspended in lysis buffer and 0.5 mm Zirconia/Silica beads (BioSpec) were added. Cell lysis was performed on a Mini BeadBeater 16 (BioSpec) with three 30 sec cycles and interleaved by 2 min of incubation on ice. Lysates were separated from beads by centrifugation at 1000*g* for 3 min, vortexed and centrifuged at 15,000*g* for 10 min. The supernatant was collected and used for pull-down experiments.

To assay for non-specific protein binding to Avidin or the biotinylated DNA anchor, three parallel experiments were performed, that is, three samples were incubated with the cell lysate: 1) Avidin-biotin primer-RNA, 2) Avidin only, and 3) Avidin-biotin primer complex. One hundred twenty µL of cell lysate were added to each reaction and incubated for one hour at room temperature with end-over-end rotation. Following incubation, samples were transferred to 30 µm polyethylene spin columns and washed three times with 200 µL of lysis buffer. Digestion buffer containing 50 mM HEPES (pH 7.5), 0.1 M urea and 5% acetonitrile was added to the samples and digestion was carried out overnight at 31 °C with Lys-C at a ~1:200 Lys-C:protein ratio. Flow-through digests were collected. Avidin beads on spin columns were washed twice with a solution containing 75% acetonitrile and 0.5% acetic acid, and all flow-through was collected and combined for each sample. Acetonitrile was removed by evaporation and the peptides were purified with the stage tip method (128) and dried again. Purified samples were dissolved in 5% acetonitrile and 1% formic acid for liquid chromatography tandem mass spectrometry (LC-MS/MS). Further details on LC-MS/MS runs and data analysis are found in Appendix A.

2.2.5 EMSA

2.2.5.1 Cloning of aaRSs

The cDNA for *S. cerevisiae* aspartyl-tRNA synthetase (AspRS), lysil-tRNA synthetase (LysRS) and tyrosyl-tRNA synthetase (TysRS) was obtained from DNASU plasmid repository (129). Synthetase DNA was amplified by PCR to add NdeI and XhoI restriction sites at the 5’ and 3’ ends of the corresponding genes and cloned into pET21b

(Novagen) with a C-terminal 6x histidine tag. Primer sequences are described in Table A.4. *E. coli* BL21 (DE3) transformants were selected based on ampicillin resistance (50 μg/mL) and screened for insert by colony PCR.

2.2.5.2 aaRSs expression and purification

Three mL starter cultures of *E.coli* BL21 (DE3) transformed cells were incubated overnight at 37 °C with 50 μg/mL ampicillin. These cultures were used to inoculate (1% v/v) 100 mL cultures in a shaking incubator at 37 °C until OD600 reached 0.5 – 0.7. Expression was induced by addition of 0.5 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were incubated for 4 h at 37 °C. Cells were harvested by spinning at 4 °C and 3900*g* for 15 min. Pellets were stored at -80 °C until use.

Protein purification was carried out by standard immobilized metal affinity chromatography (IMAC). Cell pellets were resuspended in 5 mL of 50 mM sodium phosphate, 300 mM NaCl, and 20 mM imidazole (pH 8.0). Cell lysis was performed in ice water by sonication with QSonica Q125 at 50% amplitude for 30 sec nine times with 30 sec intervals followed by centrifugation in a Sorvall RC 5Bplus centrifuge at 4 °C, 23,426*g* for 30 min. Clarified cell lysate was bound to 1 mL of Ni-NTA resin with gentle rocking at 4 °C for 45 min, and then purified with column chromatography. The resin was washed twice with five column volumes (CVs) of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 50 mM imidazole. Protein was eluted with 0.5 CVs of 50 mM sodium phosphate (pH 8.0), 300 mM NaCl and 250 mM imidazole. Protein concentration was measured by BCA (130). Active fractions were concentrated and stored in 50% (v/v) glycerol. Protein purity and integrity was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.5.3 ES7 rRNA/aaRSs interactions

RNA-protein interactions were characterized by EMSA as described previously (119) with the following modifications. A 10 µM solution of ES7 RNA was prepared in 20 mM Tris-HEPES (pH 8.0) and annealed by cooling from 90°C to 25 °C at 2°C/min. aaRSs were introduced at varying concentrations (0.25 µM-10 µM final) and incubated with 1 µM of ES7 RNA (final concentration) for 20 min at room temperature. Multi-protein EMSA was performed with fixed RNA (4 µM) and fixed protein concentration (10 µM LysRS and 10 µM TyrRS) in 20 mM Tris-HEPES (pH 8.0). RNA-protein incubation was also performed for 20 min at room temperature. RNA-protein interactions were analyzed on 5% native-PAGE with 3% glycerol. Gels were visualized using a two-color fluorescent dye protocol in a Typhoon FLA 9500 (GE Healthcare) (131). In a two-color fluorescent EMSA, uncomplexed nucleic acid is green, free protein is red and the nucleic acid protein complexes are yellow, although slight leaking between channels of the Typhoon imager was observed. Control reactions used a fragment of Domain III rRNA from *T. thermophilus*.

2.3 Results

2.3.1 Protein binding to *S. cerevisiae* ES7

S. cerevisiae cell lysates were screened for proteins with affinity for ES7 using pull-down assays. Mass spectrometry (MS) was used to identify ES7-associated proteins extracted from the lysates. During exponential growth of *S. cerevisiae*, 36 proteins (Table A.6) were seen to interact with ES7. Our assay identifies primary proteins that interact directly with ES7 in addition to secondary proteins that bind to the primary proteins.

As an illustration of the process used here for identifying ES7-associated proteins, a tandem mass spectrum corresponding to a peptide from lysyl tRNA synthetase (LysRS) is shown in Figure A.1. We attached isolated ES7 to beads, incubated the beads with cell lysates, washed the beads to remove non-specific and weakly associated proteins, and proteolytically digested the ES7-associated proteins. The peptide ELELNFSRPWK was identified with an XCorr of 3.0, and a mass accuracy of -0.69 ppm. XCorr values represent the correlation between tandem mass spectra recorded in our experiment and the corresponding theoretic spectra of peptides in the database. The higher the XCorr, the greater the confidence in peptide identification. Similarly, mass accuracies determine the extent to which theoretical masses match observed masses of peptides in our experiments. Mass accuracies of less than 3 ppm indicate high confidence in peptide identification.

Many of the proteins seen here to associate with ES7 have been shown previously to bind to the ribosome. Of the 36 ES7-associated proteins identified here, 25 were identified by Link as components of intact translation complexes (132). These proteins include Aspartyl tRNA synthetase (AspRS), Translation Machinery-Associated Protein 10 (TMA10), Ribosome biogenesis protein ERB1 and ATP-dependent RNA helicase MAK5. Of these, 12 of them were identified as mRNA-binding proteins (mRBPs) by Hentze (133). Nine proteins that associate with ES7 in this work were found previously to associate with both intact ribosomes and with mRNA (132-133). Those nine are (i) protein transport factor SEC1, (ii) MAK21, which is involved in maturation and export of pre-LSU particles, (iii) fimbrin, which is an actin-bundling protein involved in cytoskeleton organization and maintenance, (iv) HXK1, which catalyzes hexose phosphorylation, (v) NAD(+) salvage pathway component, nicotinamidase, (vi) transcription elongation factor SPT6, (vii)

AspRS, (viii) 40S ribosomal protein S30, and (iv) protein PBI2, inhibitor of vacuolar protein B.

Seven aminoacyl tRNA synthetases (aaRSs) were identified here in association with ES7. AspRS and tyrosyl tRNA synthetase (TyrRS) were found in each of the replicate experiments. Cumulatively Ala-, Arg-, Asn-, Asp-, Leu-, Lys- and TyrRS were observed in at least one experiment. Our observations are consistent with Link et al., (132) who observed Ala-, Asn-, Asp-, Leu- and LysRS bound to intact ribosomes. Asp-, Leu- and LysRS associate more tightly with the intact ribosome than AlaRS and AsnRS. The binding of AspRS, TyrRS and LysRS to ES7 is confirmed here by *in vitro* binding assays (below).

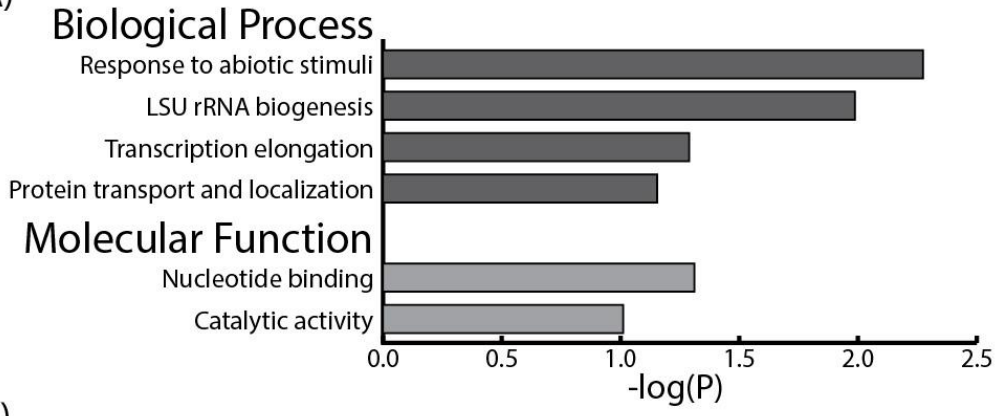
A total of six proteins known to localize in the nucleolus (134) are observed here to associate with ES7. Nucleolar proteins that associate with ES7 include (i) MAK5, (ii) MAK21, (iii) ERB1, (iv) Ribosome biogenesis ATPase RIX7, (v) DNA-directed RNA polymerase I subunit RPA49 and (vi) RNA 3'-terminal phosphate cyclase-like protein RCL1.

Clustering of ES7-associated proteins by biological process was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (135-136). Four clusters are highly enriched (Figure 2.3), including (i) response to abiotic stimulus such as temperature, (ii) ribosomal LSU biogenesis, (iii) protein transport and localization, and (iv) transcription elongation. For this analysis, if multiple clusters contained identical proteins, only the cluster with the lowest *P*-value was retained. The two most intense clusters correspond to response to abiotic stimulus ($P = 5.3\text{x}10^{-3}$) and ribosomal LSU biogenesis ($P = 1.03\text{x}10^{-2}$). Clusters with *P*-values lower than 0.05 are considered to be highly enriched and thus specifically associated with the corresponding biological process.

With DAVID we were unable to cluster all 36 ES7-associated proteins because the functional annotation for the *S. cerevisiae* proteome is not fully finalized in the database. Of the 36 ES7-associated proteins, 12 were not clustered. Six mitochondrial proteins observed were assumed to be artifacts and were excluded from the analysis.

For comparison, we clustered Hentze's mRBPs with the same methods used here for ES7-associated proteins, and observed some of the same clusters. Clusters common to mRBPs and ES7-associated proteins include (i) response to abiotic stimulus and (ii) protein transport and localization. ES7-associated proteins were also clustered based on molecular function with DAVID (Figure 2.3). The results demonstrate that ES7-associated proteins are important in (i) nucleotide binding and (ii) catalytic activity. Both of these clusters are also seen in mRBPs.

A)



B)

Biological Process	Genes
Response to abiotic stimuli	SEC1, PBS2, PBI2, SAC6, HXK1, TMA10, BDH1, PNC1
LSU rRNA biogenesis	MAK5, RIX7, ERB1, MAK21
Transcription elongation	SPT6, CHD1, CTR9
Protein transport and localization	NUM1, SEC1, PBS2, RIX7, YHB1, HXK1, YRB1, RTS1, YGR130C, VMA5, PEP8, SAC6, PBI2, SEC17
Molecular Function	**Genes**
Nucleotide binding	VMA5, MAK5, RIX7, PBS2, CHD1, HXK1, TYS1, DPS1, MET5, ERG1
Catalytic activity	MAL32, MAK5, PBS2, RIX7, SPT6, YHB1, HXK1, TYS1, PNC1, DPS1, ERG1, RTS1, RCL1, RPA49, VMA5, CHD1, GPT2, BDH1, MET5

Figure 2.3 Clustering of *S. cerevisiae* ES7 RNA-proteins identified in this work: based on biological process and molecular function. A) Clusters were organized based on their *P*-values. A greater –log(*P*) corresponds to higher statistical significance of a particular group of proteins. B) List of genes included in each cluster.

2.3.2 *S. cerevisiae* ES7 forming complexes with tRNA synthetases *in vitro*

To assess the validity of the results of the pull-down experiments, and to determine if some of the associations are direct or are mediated by other proteins, we selected three ES7-associated proteins and assayed their affinities for ES7 *in vitro*. The affinities of Asp-, Lys- and TyrRS for ES7 were assayed by Electrophoretic Mobility Shift Assay (EMSA). The results (Figure 2.4) demonstrate that Asp-, Lys- and TyrRS bind tightly to ES7, and suggest that the interaction we detect is direct and not mediated by other proteins. One μM of rRNA and 1 μM of protein produce a single shifted rRNA band. It is probable that aaRS proteins bind to ES7 rRNA with 1:1 stoichiometry. Control experiments show that the affinities for ES7 rRNA are 3 to 10 times greater than for a different fragment of rRNA (Figure A.2).

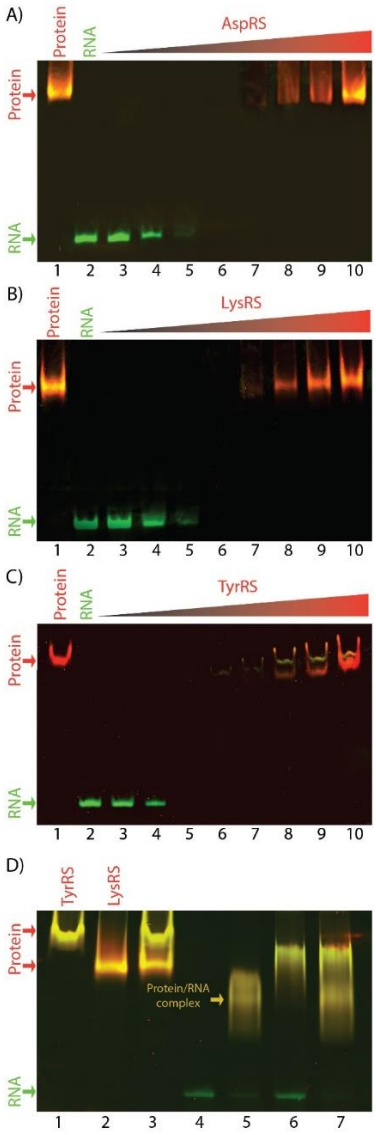


Figure 2.4 EMSA of *S. cerevisiae* ES7 RNA with aaRSs. Binding of ES7 to A) AspRS, B) LysRS, C) TyrRS and D) LysRS and TyrRS is shown *in vitro*. RNA, protein and RNA-protein complexes were visualized by two color EMSA in a native gel. Binding reactions for panels A-C were performed with 1 μM RNA and various concentrations of protein. Lane 1 contains 10 μM of protein and no RNA. Lane 2 consists of 1 μM RNA only. Lanes 3-10 contain 1 μM RNA and increasing protein concentration: lane 3 – 0.25 μM protein, lane 4 – 0.50 μM, lane 5 – 1 μM, lane 6 – 2 μM, lane 7 – 4 μM, lane 8 – 6 μM, lane 9 – 8 μM and lane 10 – 10 μM. Binding reactions for panel D were performed with 4 μM RNA and 10 μM of each protein. Lane 1 contains 10 μM of TyrRS, Lane 2 contains 10 μM of LysRS and Lane 3 contains 10 μM of TyrRS and 10 μM of LysRS, as controls. Lanes 4 – 7 contain 4 μM RNA and various aaRSs: lane 4 – no protein, lane 5 – 10 μM TyrRS, lane 6 – 10 μM LysRS and lane 7 – 10 μM TyrRS and 10 μM LysRS.

For these 1:1 complexes, gels were quantitated and dissociation constants, K_d's, were estimated to be approximately 100 nM for LysRS and 1 µM for Asp- and TyrRS. The affinities of these aaRS proteins for ES7 are roughly equivalent to the affinities observed elsewhere for ribosomal proteins for cognate rRNA (137-138). K_d's in the low micromolar to nanomolar ranges have been observed for ribosomal proteins uL1, uL2, uL3 and uL6 with rRNA.

Multi-protein EMSA experiments do not provide evidence of simultaneous binding of multiple aaRSs to ES7 (Figure 2.4D, Figure A.3). LysRS and TyrRS were selected for this assay based on the differential migration of the free proteins in native-PAGE. A fixed amount of ES7 RNA was incubated with TyrRS, LysRS and a combination of TyrRS and LysRS. The mobilities of LysRS and TyrRS alone differ from the mobility of the ES7-LysRS and ES7-TyrRS complexes. Incubation of ES7 simultaneously with TyrRS and LysRS showed at least two distinct bands that correspond to ES7-LysRS and ES7-TyrRS complexes. Bands suggesting the formation of additional ES-aaRS complexes could not be observed. The data does not support formation of the ES7-LysRS-TyrRS complex.

2.3.3 Secondary structure of ES7

SHAPE footprinting (139-140), computational folding (141) and phylogenetic data were used to determine the secondary structure of isolated *S. cerevisiae* ES7, which is seen to be the same as the ES7 secondary structure in the Yusupov three-dimensional structure (17) and the secondary structure of purified ribosomes determined by Dinman (142). SHAPE monitors relative reactivity of 2'-oxygens to an electrophile, indicating extent of local flexibility. Nucleotides in unpaired loops and bulges and in non-canonical base pairs

show elevated SHAPE reactivity while canonical helical regions show depressed SHAPE reactivity. A direct comparison of the SHAPE reactivity of isolated ES7 to that of Dinman's ES7 secondary structure demonstrates that the secondary structure of isolated ES7 *in vitro* is the same as in intact ribosomes.

SHAPE reactivities were determined for all nucleotides of isolated *S. cerevisiae* ES7 (Figure 2.5A). Nucleotides were clustered into three groups based on their SHAPE reactivity, after background subtraction and normalization. Positions of low SHAPE reactivity (<0.4) are located in helical regions in the secondary structure, whereas moderate and highly reactive SHAPE positions (>0.4) are concentrated in bulges, loops and RNA mismatches. The observed SHAPE data is fully consistent with the secondary model shown.

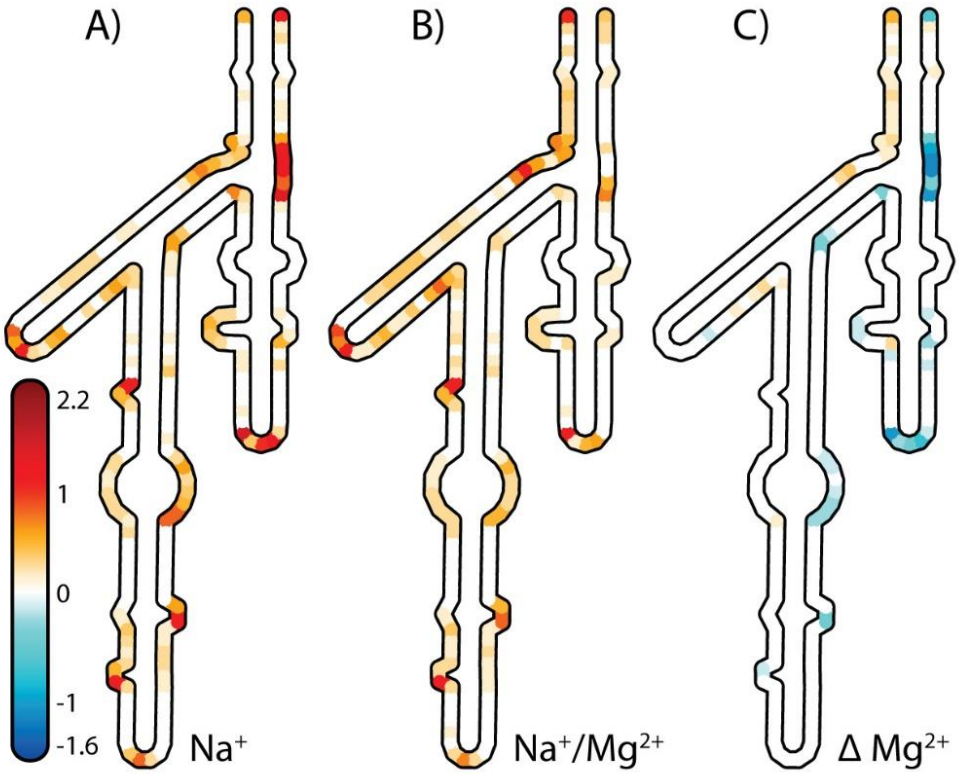


Figure 2.5 SHAPE reactivities of isolated *S. cerevisiae* ES7 rRNA. SHAPE reactivities were determined in presence of A) Na^+ or B) Na^+ and Mg^{2+}. Orange or red nucleotides show moderate to high SHAPE reactivity. Ivory or white nucleotides show low or no SHAPE reactivity. C) Mg^{2+} induces changes in SHAPE reactivity. Nucleotides in red show increased reactivity in the presence of Mg^{2+}. Nucleotides in blue show decreased reactivity. These figures were generated with RiboVision. All samples contained 200 mM NaOAc, 50 mM Na-HEPES (pH 8.0).

2.3.4 Tertiary structure of ES7

Changes in SHAPE reactivity suggest that Mg^{2+} induces formation of tertiary interactions in isolated ES7. Previously, the Mg^{2+} dependence of SHAPE reactivity was determined for a variety of RNAs including tRNA (124), RNase P (123), Domain III (125) and the intact LSU rRNA of *Thermus thermophilus* (Timothy K. Lenz, Nicholas Hud and Loren Williams, unpublished observations). The data were interpreted to indicate formation of tertiary interactions upon addition of Mg^{2+}. We have used this comparative method to demonstrate specific inter-domain, intra-domain and subunit folding of LSU rRNA caused by addition of Mg^{2+} (119, 125).

SHAPE data obtained in the presence of Na^+ alone (Figure 2.5A) was subtracted from that obtained in the presence of Na^+ and Mg^{2+} (Figure 2.5B). The heat map obtained by this method (Figure 2.5C) identifies nucleotides that increase or decrease in reactivity upon addition of Mg^{2+}. It can be seen that canonical helical regions are invariant to the addition of Mg^{2+}, consistent with the expectation that secondary structure is maintained upon formation of tertiary structure. Changes in reactivity are focused on loops and non-helical regions, which are expected to be involved in tertiary interactions. The results suggest that isolated ES7 of *S. cerevisiae* forms tertiary interactions upon addition of Mg^{2+}.

2.3.5 Thermal folding/unfolding of *S. cerevisiae* ES7

Isolated ES7 is well-behaved in solution. The number of transitions in the unfolding process provides information on the architecture of ES7. Thermal melts show that the folded structure is stable at room temperature and that folding and unfolding are reversible. Thermodynamic parameters were obtained from the melting curves by non-linear fitting to a multi-state model with independent transitions (Figure A.4) (143). An estimate of the number of melting transitions and the thermodynamic parameters associated with these transitions is obtained from the melting profile, which is the derivative of the fraction unfolded (dϴ/dT) as a function of temperature (Figure 2.6). The best fit to the melting profile of ES7 was obtained with a four transition model (Figure 2.6). A three-transition unfolding model gives significantly worse statistics of fit. However, unfolding models with additional transitions cannot be excluded by the methods used here and, in fact, are probable for such a large RNA.

Melting profiles were fit and deconvoluted to obtain estimates of thermodynamic parameters for each of the four transitions. The melting temperatures ($T_{m,n}$) of the transitions are $T_{m,1}$ = 51°C, $T_{m,2}$ = 64°C, $T_{m,3}$ = 75°C and $T_{m,4}$ = 83°C in 180 mM NaCl. The highest temperature transition has a relatively small hyperchromicity but was retained in the model because it is reproducible and observed in the thermal melting of isolated ES7 rRNA from a closely related fungi (unpublished observations). In this method, uncertainty arises because enthalpies can correlate with hyperchromicities (143). The T_m's are constant with varying RNA concentrations (unpublished observations) indicating that the transitions are unimolecular. Since ES7 rRNA is large and is subject to degradation, Mg^{2+} was

excluded from melting buffers to prevent in-line cleavage (144). Additional information on thermodynamic models and parameter estimation are provided in Appendix A.

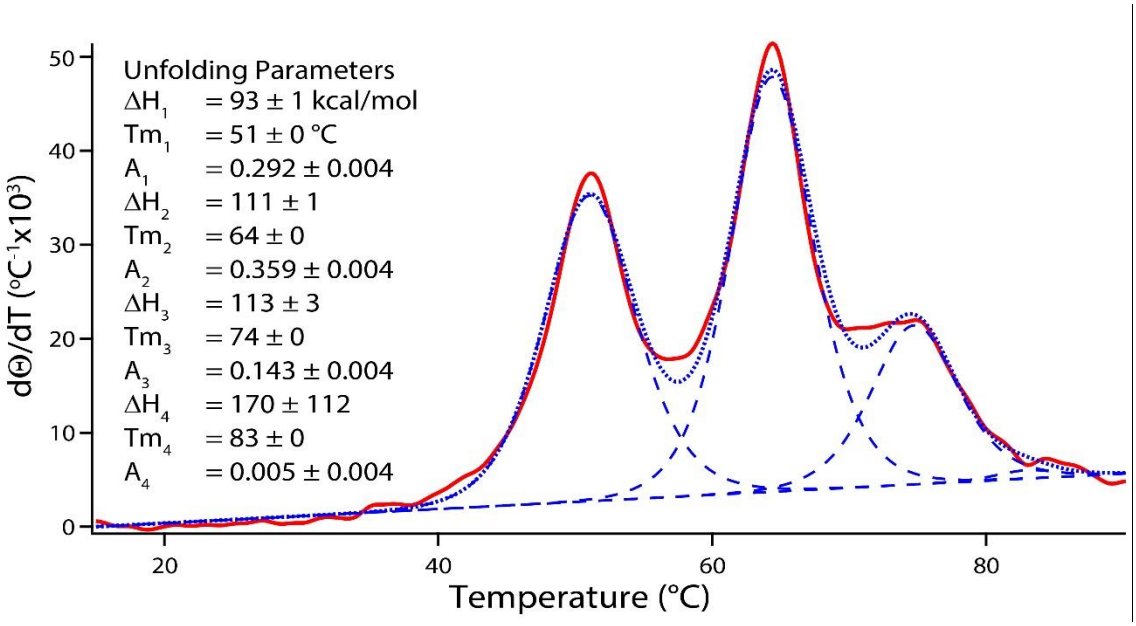


Figure 2.6 Melting profile of *S. cerevisiae* ES7 rRNA. The derivative of fraction unfolded at 260 nm with respect to temperature was computed. The derivative was fit to a non-sequential, independent transition model with IGOR Pro. The enthalpy (ΔH_n), melting temperature ($T_{m,n}$) and the relative hyperchromicity (A_n) were estimated for each transition, n. The observed derivative is red, the fit is dotted blue and the deconvolution of the fit is dashed blue.

2.4 Discussion

Mammals have the largest known rRNAs. Metazoans have larger rRNAs than protists, which have larger rRNAs that prokaryotes. We suggest that the breadth of ribosomal function is related to the size of the rRNA (4) and that eukaryotic ribosomes facilitate functions auxiliary to translation. One of the foci of rRNA expansion is ES7. By determining which proteins associate with ES7 we can deduce at least some of its contributions to function.

At least 36 proteins in cell lysates bind here to isolated ES7 rRNA of *S. cerevisiae* directly or through intermediates. Of these ES7-associated proteins, 25 were previously

identified by Link in association with intact ribosomes (132). It appears we have identified the subset of Link's proteins that bind to ES7 rRNA as opposed to other regions of the ribosome. Of these ES7-associated proteins observed here, 12 were also identified as mRBPs by Hentze (133). Some proteins identified here may bind at multiple sites on the ribosome. Six of the ES7-associated proteins reside in the nucleolus consistent with roles of ES7 in ribosome biogenesis (145).

ES7-associated proteins were clustered with DAVID (135-136) into four groups based on biological process, (i) response to abiotic stimulus such as temperature, (ii) LSU biogenesis, (iii) protein transport and localization, and (iv) transcription elongation (Figure 2.3). Clustering based on molecular function shows ES7-associated proteins are important in nucleotide binding and catalytic activity.

aaRSs. aaRSs couple amino acids with their cognate tRNAs, establishing the relationship of nucleotide triplets with amino acids (i.e., the genetic code). Cumulatively seven synthetases, Ala-, Arg-, Asp-, Asn-, Leu-, Lys- and TyrRS were observed in at least one experiment. ES7 may be a focus of aaRS association on the ribosome. Previously, 16 aaRSs were observed to associate with intact ribosomes in *S. cerevisiae* (132). ES7 may be an ancestor or proxy of large multi-aaRS complexes in other systems, maintaining close association of synthetases with each other and with the translational machinery. In *S. cerevisiae*, only two aaRSs are found in multi-synthetase complex (146). These two aaRSs are not observed here to associate with ES7. Half of the ES7-associated aaRSs in *S. cerevisiae* correspond with those observed previously in the human multi-aaRS complex (147). Yeast two-hybrid studies in archaea suggest that EF-1α associates with LeuRS and contributes to the overall translation efficiency by promoting aminoacylation (148). Our

work complements these findings. ES7 could act as a hub for tRNA aminoacylation and for improving translation efficiency.

Affinity and specificity of AspRS, TyrRS and LysRS for ES7 were confirmed by *in vitro* binding assays. AspRS and TyrRS were observed in each of the replica experiments and appear to be the most robust ES7 binders. The affinities and specificities of Asp-, Lys- and TyrRS for ES7 were assayed by EMSA. The data suggest that these aaRS proteins bind to ES7 rRNA with 1:1 stoichiometry. Their affinities for ES7 are greater than for a different rRNA fragment (Figure A.2). The results also suggest that these three aaRSs bind directly to ES7 supporting the hypothesis that ES7 in *S. cerevisiae* might promote close association of aaRSs with each other and with the ribosome. Based on the direct association of ES7 to Asp-, Lys- and TyrRS it seems likely that additional aaRSs might bind to ES7 *in vivo* and may form stable RNA-protein complexes. Multi-protein EMSA does not provide evidence for simultaneous association of ES7 with multiple aaRSs.

Observations of RNA-protein association in pull down experiments must account for specificity. A complex observed experimentally might not be relevant *in vivo* because of competition from other factors. Several measures were taken to control non-specific RNA-protein binding. For a subset of the ES7-associated proteins we have directly confirmed high affinities *in vitro* (Figures 2.4, A.2 and A.3). ES7 has 26,000 $Å^2$ of accessible surface area when stripped of ribosomal proteins and 20,000 $Å^2$ of accessible surface area in the assembled ribosome. Therefore ES7 provides ample surface area for association with non-ribosomal proteins. Considering that rRNA accounts for over 80% of cellular RNA (149), competition from other RNAs seems unlikely, and the proteins detected here must be considered highly probable ES7 binding partners *in vivo*. However,

binding of proteins could be transient and might depend upon the stage cell cycle and on external cell environment. We expect that the majority of the proteins identified in this study are representative of the proteins that bind to ES7 *in vivo* during log phase, although we cannot exclude additional binding partners.

Ribosomal proteins. Although some nucleolar proteins were identified, it seems likely that nucleoli and intact nuclei were depleted prior to the assay. None of the ES7-associated proteins found here are ribosomal proteins. ES7 rRNA does contact multiple ribosomal proteins in the assembled ribosome (Figure 2.2). rProtein uL4 makes the most extensive contacts with ES7 *in vivo*, yet is not observed in our assay. It is possible that uL4 in the cytoplasm is tied up in association with ribosomes, and that the amount of free uL4, not associated with assembled LSU, is sufficiently small that it cannot be detected in our assay.

ES7 folding. Our footprinting and folding/unfolding experiments along with computational studies indicate that isolated ES7 folds to a native-like secondary structure, consistent with the secondary structure inferred for intact *S. cerevisiae* ribosomes (17, 142). In the presence of Mg^{2+}, isolated ES7 forms tertiary interactions. The thermal melting shows that folded ES7 is highly stable and that the multistate folding/unfolding process is reversible. Combined, the results provide experimental confirmation of the secondary structure of *S. cerevisiae* ES7, and suggest that our experiments employed a native-like structure of ES7 rRNA.

2.5 Conclusions

We hypothesize that expansion segments could be sites of functional interactions with other macromolecules given their proximity to the ribosome surface. ES7 is the largest expansion segment and the most variable region of rRNA over phylogeny. We performed *in vitro* characterization of *S. cerevisiae* ES7 by measuring its stability, and its interactions with cellular proteins. The results here support our hypothesis. Here, we demonstrate that Asp-, Lys- and TyrRS bind directly to *S. cerevisiae* ES7. aaRSs charge amino acids with their cognate tRNAs catalyzing the first reaction needed for protein synthesis (150). Therefore, expansion segments likely play a variety of important roles in protein synthesis.

CHAPTER 3

ELONGATED EXPANSION SEGMENTS EXTEND THE CAPABILITIES OF HUMAN RIBOSOMES

3.1 Introduction

Secondary and three-dimensional structures of rRNA in the ribosomal common core are universally conserved in all living systems (4, 19, 22), as are mechanisms of mRNA decoding and aminoacyl transfer. By contrast, translation initiation, termination and regulation, quality control, and ribosome biogenesis are accomplished by diverse structures and mechanisms (111, 151-157). In many cases, variability in ribosomal function appears to be related to specific differences in ribosomal structure beyond the common core. Relative ribosomal size is a predictor of differential ribosomal function.

The prokaryotic ribosome is a reasonable approximation of the common core. With a surface patchwork of rRNA and ribosomal proteins (rProteins) (16, 158), the common core is universal to all cytoplasmic ribosomes. Larger and more complex ribosomes are found in eukaryotes due to rRNA expansion segments (ESs) (2-3, 32) that emerge at the surface of the common core (Figure 3.1A). Eukaryotic ESs and eukaryote-specific rProteins encase much of the common core, forming a secondary eukaryotic rRNA–rProtein shell (21, 159-160). Sizes and secondary structures of ES7 in various organisms are shown in Figure 3.1B.

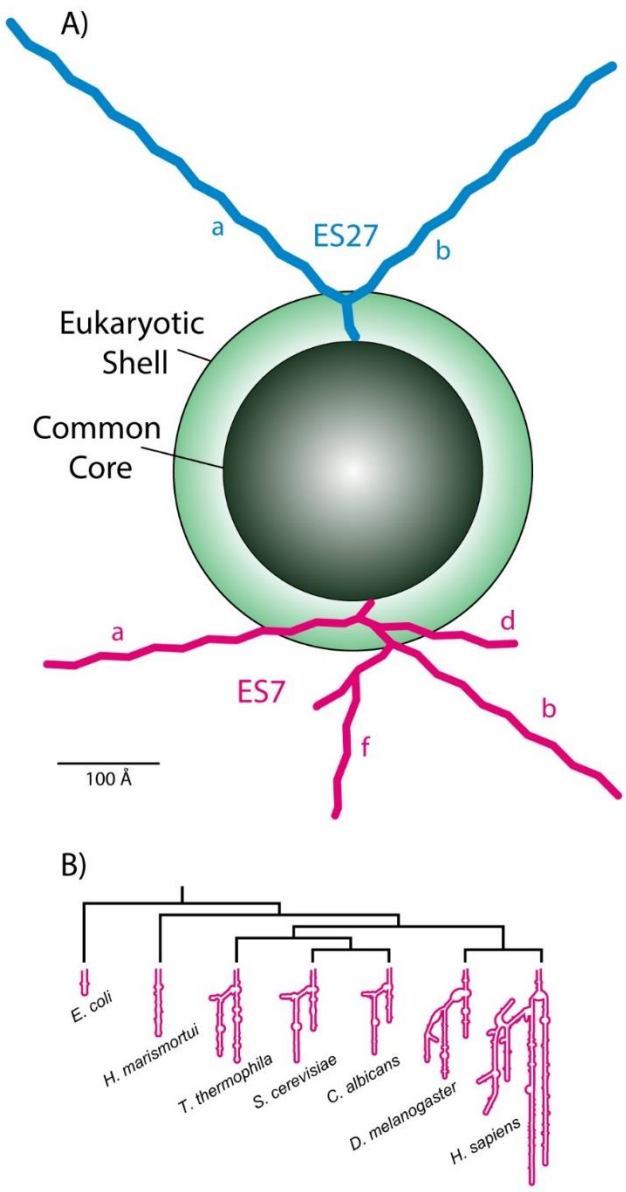


Figure 3.1 ESs from the human ribosome are the biggest ones in nature.A) Schematic diagram of the human LSU, showing the relative sizes of the common core (dark green) and the eukaryotic shell (light green). Human ES7 (magenta) and ES27 (blue), which contain the longest ESs are shown. The lengths of ES7 and ES27 are correctly proportioned relative to the size of the common core and eukaryotic shell. Human ES7 contains two primary tentacles (ES7a and ES7b) which are both estimated to be ~325 Å in length (~125 bp). Human ES27 contains two tentacles (ES27a and ES27b) which are estimated to be ~450 Å and ~350 Å in length (ES27a, ~180 bp; ES27b, ~145 bp). This schematic is not intended to accurately indicate ES/tentacle conformation. B) Secondary structures of Helix 25/ES7 from throughout the phylogenetic tree. Each of these secondary structures has been experimentally determined by diffraction methods or chemical footprinting.

In protists, ESs are important for ribosome biogenesis (115-116), rRNA processing and stabilization (47, 60), translation initiation (68) and recruitment of non-ribosomal factors such as chaperones, enzymes or signal recognition particles (69-70). ESs can be described as docking regions for proteins that enhance and extend ribosomal function (161). Previously we characterized the secondary structure and protein binding capabilities of ES7 of *S. cerevisiae* (shown in Figure 3.1B) and suggested that ES7 confers auxiliary functionality to the ribosome. In *S. cerevisiae,* ES7-associated proteins include aminoacyl tRNA synthetases (aaRSs), quality control proteins and chaperones.

ESs of metazoan ribosomes contain characteristic rRNA 'tentacles', which are long unbranched helical structures that attach to the surface of the eukaryotic LSU (20, 162) (Figure 3.1). Tentacles are laden with defects such as bulges and mismatches. Except at their attachment sites, rRNA tentacles are not tightly integrated with the ribosomal surface and appear to be dynamic and positionally disordered. These properties preclude their observation by structure determination by X-ray diffraction and cryo-EM. The lengths, dynamics, nucleotide compositions and lack of helical branching of human rRNA tentacles are anomalous among known ESs. Human ESs (and its tentacles) are highly enriched in C and G (approximately 85% CG content).

Here we investigate the functions of the largest ESs of the human ribosome (Figure 3.1 and 3.2), which is the largest ribosome known in nature (4), with the longest tentacles. Our results reveal association with ESs of a variety of critical proteins and assemblies, including the ubiquitin-proteasome system (UPS). The UPS directly associates with rRNA ESs. The results suggest that protein synthesis in humans is physically coupled with protein degradation. Human ESs are also associated with a variety of other functional elements,

including those involved in metabolism, transcription, translation, mRNA stability and regulation, and cell structure.

The distribution of ES-associated proteins appears to depend on cell line. We observe differences in protein-ES association in a non-carcinoma cell line and in a carcinoma cell line. In addition, the distribution of ES-associated proteins appears to depend on the identity of the ES, suggesting specific rRNA structures and sequences could be important determinants of protein association. We have characterized the stability, secondary structure and folding of isolated ES7. We note that the tentacles of ES7 contain putative G-quadruplex forming sequences and that in our experiments ES7 is seen to associate with known G-quadruplex binding proteins.

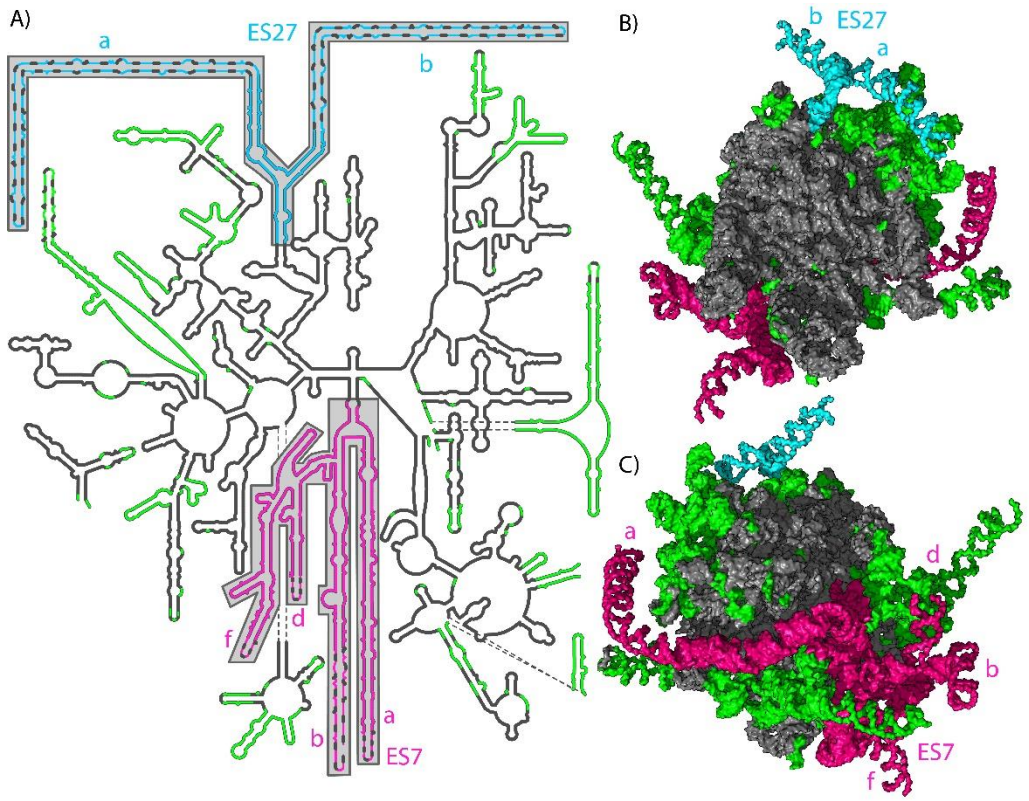


Figure 3.2 Human LSU rRNA with highlighted expansion segments. A) Secondary structure is shown. B) Three-dimensional structure with front view (SSU interface) and C) three-dimensional structure with back view (solvent accessible region) of the LSU are shown. Common core rRNA, which is contained in all cytoplasmic ribosomes, is gray. ES7 (magenta) and ES27 (blue) are enclosed in shaded boxes. Dashed lines on the ESs in panel A indicate segments of tentacles that are not observed by three-dimensional structure determination and which are absent from panels B and C. Other ESs are shown in green. The images in panel B and C are upside down relative to the conventional ribosome representation so that they are oriented in accordance with the secondary structure in panel A. These images are made from PDB entry 4UG0. Ribosomal proteins are omitted for clarity.

3.2 Materials & Methods

3.2.1 Human ES7 and ES27 sequences

ES7 corresponds to nucleotides (nts) 436 to 1311 of the LSU rRNA (Figure 3.2). ES27 corresponds to nts 2903 to 3591 (Figure 3.2). Tentacles ES27a and ES27b are composed of nts 2914 to 3279 (ES27a) and nts 3285 to 3582 (ES27b).

3.2.1.1 Plasmid construction

The genes encoding human ES7, ES27, ES27a and ES27b (Table B.1) in pUC57-Kan resistant plasmids were purchased from Genewiz. All constructs were confirmed by sequencing (Eurofins MWG Operon). Each construct contains the T7 promoter and restriction sites for EcoRI (5’ end) and HindIII (3’ end). Once expressed, the RNA contains a 3’ single stranded region that acts as a primer binding site for SHAPE reverse transcription (RT) reactions and as a complement to a biotin-linked DNA oligomer for anchoring the RNA to beads in RNA-protein pull-down experiments. The appended 3’ single-stranded RNA, taken from Weeks (124), allowed unobstructed SHAPE mapping of the complete ES7 sequence. ES27 was appended to a seven nucleotide base pair clamp (5’ CAAGTAC 3’) designed to promote correct secondary structure at the terminus. The high

CG content of all the sequences used in this work makes the RNA and DNA difficult to manipulate via molecular biology techniques.

$Gq1_{ES7}$ and $Gq2_{ES7}$ (Table B.2), which contain putative G-quadruplex forming sequences, were cloned into pUC19 and overexpressed as RNA. Restriction digests and purifications were performed by standard procedures. Restricted DNA was inserted into previously restricted, dephosphorylated pUC19 plasmid.

3.2.1.2 *In-vitro* RNA synthesis

RNAs were produced by transcription (HiScribe™ T7 High Yield RNA Synthesis Kit, New England BioLabs) as described (161). RNA purity and integrity were verified by 8 M Urea gel in TBE buffer with 5% acrylamide for human RNAs and 12% acrylamide for the shorter putative G-quadruplex RNAs.

3.2.2 SHAPE

SHAPE was adapted from published protocols (118-119). *In vitro*-transcribed RNA (400 nM) in nuclease-free water was dissolved in 200 mM NaOAc, 50 mM Na-HEPES (pH 8.0) and 1 mM 1,2-trans-cyclohexane-N,N,N'N'-diaminetetraacetic acid or 5 mM Mg^{2+}. RNAs were annealed by slow cooling from 75 °C to 25 °C at 1 °C/min. SHAPE reactions were conducted in 200 mM Na^{+} and the absence of Mg^{2+} for secondary structure determination. Tertiary interactions and divalent binding sites were characterized by changes on SHAPE reactivity upon addition of Mg^{2+}.

3.2.2.1 Benzoyl cyanide modification

Chemical modification of human ES7 rRNA and subsequent RNA purification were performed as described (161) except that benzoyl cyanide concentration was decreased by half (500 mM) to promote longer reads of the reverse transcriptase.

3.2.2.2 Reverse transcription of benzoyl cyanide-modified human ES7 RNA

Reverse transcription using [6-FAM]-labeled primers and Maxima Reverse Transcriptase (Thermo Fisher Scientific) provided SHAPE data on 91% of ES7. Modified RNA was added independently to 40 pmol of primer 1, 80 pmol of primer 2 or 60 pmol of primer 3 and RT buffer to yield 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT and 625 µM of each dNTP in 50 µL.

Primers were annealed to modified RNA following a stepwise incubation procedure at 95 °C for 1 min, 60 °C for 5 min, 55 °C for 5 min and 4 °C for 10 min in a thermocycler. Primer sequences are shown in Table B.3. RT mixtures were pre-incubated at 52 °C for 2 min before addition of 200 U of reverse transcriptase. RT reactions were run for 2 h at 52 °C and were quenched by stepping the temperature to 85 °C for 5 min. Control RT reactions were performed in un-modified RNAs.

RNA sequencing reactions used dideoxynucleotides (ddNTPs) with modified ddNTP to dNTP ratios that were established empirically to account for the high CG content of these rRNAs. For primer 1 the reaction ratios were 1:8 ddNTP to dNTP for guanosine and cytosine and 1:1 for thymidine and adenosine. For primers 2 and 3 the ratios were 1:15 ddNTP to dNTP for G and C and 1:1 for T and A. Ten pmol of human ES7 RNA in TE buffer and 40 pmol of primer 1, 80 pmol of primer 2 or 60 pmol of primer 3 were used in sequencing reactions. Reactions lacking ddNTPs were run as controls. RT reactions were analyzed by capillary electrophoresis as described elsewhere (119).

3.2.2.3 SHAPE data processing

SHAPE reactivities were obtained after background subtraction and normalization. Nucleotides were clustered by SHAPE reactivity. Unreactive nucleotides (<0.4) are confined to helical regions in the secondary structure. Moderately (0.4-0.8) and highly reactive nucleotides (>0.8) are concentrated in bulges, loops and helical defects. SHAPE data were processed as described (125) and were mapped onto secondary structures with RiboVision v.1.15 (126).

3.2.3 RNA folding/unfolding

Thermal melting of rRNA fragments was conducted by monitoring UV absorbance at 260 nm as temperature was varied in a thermostated Varian Cary-1E UV spectrophotometer. Melting solutions contained 0.3 ODs of RNA, 18 mM of monovalent cation (Li^+, Na^+, K^+, Rb^+ or Cs^+) and 20 mM Tris-HEPES (pH 8.0).

Samples were initially heated from 15 °C to 95 °C at 3 °C/min to unfold the RNA. Equilibrium folding and unfolding reactions were conducted by cooling or heating at 0.5 °C/min. Superimposition of heating and cooling curves was used to confirm equilibrium during melting. A slow temperature ramp (0.5 °C/min) was used to accommodate the large RNA. Fitting of melting data to models was performed as described (161). The normalized absorbance vs. temperature data were subject to binomial smoothing before and after differentiation in the program IGOR Pro.

3.2.4 Circular dichroism spectroscopy

Samples were prepared at 1 µM strand concentration in 100 mM monovalent cation (Li^+, Na^+ or K^+) and 10 mM Tris-Cl (pH 7.5). Samples were annealed with a linear gradient from 90 °C to 25 °C at 1 °C/min in a thermocycler. CD spectra in 1 cm cuvettes were acquired at 20 °C on a Jasco J-810 spectropolarimeter with a Peltier temperature controller. Data from 200-320 nm was acquired at a rate of 100 nm/min with 1 sec response, a bandwidth of 4 nm, and averaged over four measurements. The buffer spectrum was subtracted. Smoothing was performed with the program Origin 9.1.

3.2.5 Pull-down assays and LC-MS/MS analysis

A sequence equivalent to primer 1, but linked to biotin at the 5'end, served as a DNA anchor (Table B.4). Ten µg of rRNA (ES7, ES27 or ES27b), appended to the RNA complement of the DNA anchor, were hybridized with 1.20 µg of the DNA anchor. DNA-RNA hybridization was performed in 180 mM NaCl and 20mM Tris-HEPES (pH 8.0), mimicking the annealing conditions used for SHAPE RT. SoftLink™ Soft Release Avidin Resin (Promega) was equilibrated three times with lysis buffer [10 mM HEPES (pH 8.0), 10 mM KCl, 1.5 mM $MgCl_2$ and 1x protease arrest from G-Biosciences] containing 10 U of RNAse inhibitor, 5 mg/mL of heparin and 0.5 mg/mL of yeast tRNA (Roche) at 4 °C. The hybrid complex was shaken gently at room temperature with 10 µL of Avidin resin previously equilibrated in the lysis buffer. Attachment of RNA to beads was confirmed by electrophoresis.

Pull-down experiments were performed with cell lysates from HEK293T and MDA-MB-231 with RNAs corresponding in sequence to human ES7, ES27 or ES27b. HEK293T and MDA-MB-231 were grown in Dulbecco's Modified Eagle's Medium,

supplemented with 10% fetal bovine serum and harvested at 90% confluency. Pellets were washed once with phosphate buffered saline and stored at -80 °C. Cells were resuspended in lysis buffer with 0.5 mm Zirconia/Silica beads (BioSpec) and lysed with a Mini BeadBeater 16 with three 30 sec cycles interleaved with 2 min on ice. Lysates were separated from beads by centrifugation at 1000xg for 3 min in a new micro-centrifuge tube, vortexed and centrifuged at 15,000xg for 10 min. Supernatants were collected and used for pull-down experiments. RNA-protein incubation was performed for 1 h with slow vortexing at room temperature. Description of controls to assess for non-specific RNA-protein binding, protein digestion, LC-MS/MS runs and data analysis can be found elsewhere (161). All MS/MS spectra were searched against the UniProt Human Database (*Homo sapiens*) downloaded in February 2014.

3.3 Results

3.3.1 Pull-down assays and LC-MS/MS analysis

Protein binding to human rRNA expansion segments

The goal here is to characterize protein interactions of ES rRNAs in human cells. Cell lysates were screened to determine which proteins associate with ES7 and ES27. We investigated a non-carcinoma cell line (HEK293T) and a highly metastatic carcinoma cell line (MDA-MB-231) to determine whether protein-rRNA interactions might vary with cell line. The assay used here does not distinguish between primary proteins, which bind directly to ES rRNAs, and secondary proteins, which bind to primary proteins. Some of the proteins observed here in association with ESs are known to interact with each other *in vivo* suggesting secondary association in our experiments.

Isolated ESs, or isolated fragments of ESs, were attached to beads, incubated with cell lysates and washed to remove non-specific and loosely associated proteins. ES-associated proteins were proteolytically digested and the resulting peptides identified by LC-MS/MS and analyzed by tandem mass spectra. Clustering of these proteins based on biological process and/or molecular function was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (135-136).

For each rRNA fragment, four replica experiments were performed with each cell line. The proteins detected were partitioned into Group I and II. Group I contains proteins observed in all four-replica experiments. Proteins in Group I are identified with the greatest confidence. Group II contains proteins observed in two or three of the four replica experiments. Group II appears to contain proteins that bind to the ESs less directly, with lower affinity, or more transiently, than Group I proteins. Groups associated with HEK293T cells are indicated with a subscript H (i.e. Group I_H). Groups associated with MDA-MB-231 cells are indicated with a subscript M (i.e. Group I_M).

Human ESs are CG-rich. ES7 and ES27 contain sequences suggesting the possibility of G-quadruplex formation (see below). Therefore, proteins observed here to bind to human ESs were evaluated for known affinity towards G-quadruplexes.

ES7-associated proteins

Many important proteins bind to ES7. The ubiquitin-proteasome system (UPS), including the ubiquitin-proteasome complex (UPC) and ubiquitin-like proteins (UBLs) associate with ES7 rRNA primarily in HEK293T cells and to a lesser extent in MDA-MB-231 cells. Proteins involved with cell-cycle regulation and division, mRNA stability and mRNA processing also associate with ES7 rRNA.

ES7-associated proteins in HEK293T cells

Human ES7 associates with multiple members of the UPS in HEK293T cells. Nine of the fifteen Group $I_{ES7,H}$ proteins are known to be involved in ubiquitination or sumoylation processes (Table B.6). Group $II_{ES7,H}$ contains eleven additional proteins with known involvement in the UPS (Table B.7).

Group $I_{ES7,H}$ contains three E1-like enzymes, which are SUMO-activating enzyme subunit 2 (SAE2), NEDD8-activating enzyme E1 regulatory subunit (ULA1) and Ubiquitin-like modifier-activating enzyme 5 (UBA5). This Group also contains Ubiquitin-conjugating enzyme E2O (UBE2O), which is a hybrid between E2/E3 enzymatic activities. Group $I_{ES7,H}$ contains several proteins that are known to bind to deubiquitinating enzymes (163). These proteins are Transcription elongation factor A protein-like 4 (TCAL4), Ankyrin repeat domain-containing protein 17 (ANR17) and Nucleoside diphosphate kinase A (NME1) (163). Group $I_{ES7,H}$, in addition, contains CD2 antigen cytoplasmic tail-binding protein 2 (CD2B2) and Neurofilament light polypeptide (NFL). CD2B2 is a protein-binding partner for E3-ligase NEDD4 (164) and NFL interacts with E3 SUMO-protein ligase PIAS4 (165).

Three Group $I_{ES7,H}$ proteins are distinguished by their roles in transcriptional or translational regulation; these include CD2B2, ERF1, and TCAL4. Some of these proteins were previously identified as binding partners for ubiquitination-related processes. TCAL4 is involved in transcriptional regulation (166), CD2B2 helps to regulate pre-mRNA splicing (167) and ERF1 recognizes the three in-frame stop codons in human mRNA and induces termination of protein biosynthesis (168).

Group $II_{ES7,H}$ (Table B.7) includes ribosomal proteins eL6 and eL33, which are binding partners of ES7 in the assembled ribosome *in vivo*. Additional members of Group $II_{ES7,H}$ include DICER, three Ras-related proteins, eukaryotic initiation factor 4E (EIF4E) and associated components, and Fragile X mental retardation protein (FMR1). A complex formed by FMR1, EIF4E and additional proteins inhibits translation ignition in neurons (169). FMR1 is known to bind G-quadruplexes (170-171). Only two aminoacyl tRNA synthetases (SerRS and AsnRS) were identified in Group $II_{ES7,H}$ in contrast to the previous observation of seven ES7-associated aaRSs in *S. cerevisiae* (161).

ES7-associated proteins in MDA-MB-231 cells

Proteins involved with the UPS appear to associate less extensively with ES7 in MDA-MB-231 cells in comparison to HEK293T cells. Group $I_{ES7,M}$ contains nine proteins, none of which are involved in the UPS (Table B.8). Around half of Group $I_{ES7,M}$ proteins are involved in mRNA stability, nucleo-cytoplasmic trafficking and mRNA decay. These proteins include Polyadenylate-binding protein 2 (PABP2), DnaJ homolog subfamily C member 8 (DNJC8), Protein quaking (QKI) and Superkiller viralicidic activity 2-like 2 (SK2L2). Ras-related protein Rab-7a is also found within Group $I_{ES7,M}$. Two additional members of the Ras protein family are contained in Group $II_{ES7,M}$.

SAE2, ULA1, UBA5 and Phosphoglucomutase-2 (PGM2) were found in both Group $II_{ES7,M}$ and Group $I_{ES7,H}$. Ribosomal proteins uL13 and eL28, known binding partners of ES7 in the assembled ribosome *in vivo*, FMR1 and twelve UPSs were also identified in Group $II_{ES7,M}$ (Table B.9).

Clustering of ES7-associated proteins

Clustering of ES-associated proteins was performed based on molecular function and biological process using DAVID (Figure 3.3). Group $I_{ES7,H}$ shows four biological process clusters. These clusters are i) protein catabolic processes, ii) metabolic processes, iii) ribonucleoside triphosphate biosynthetic processes and iv) cellular response to stress. Clustering by molecular function gives only a single cluster, which is ligase.

Group $I_{ES7,M}$ shows three biological process clusters, i) RNA processing, ii) gene expression and iii) cellular process. Clustering Group $I_{ES7,M}$ by molecular function gives only a single cluster, which is protein binding. As described previously (161), if an identical group of proteins is assigned to multiple clusters, we consider only the cluster with the lowest *P*-value.

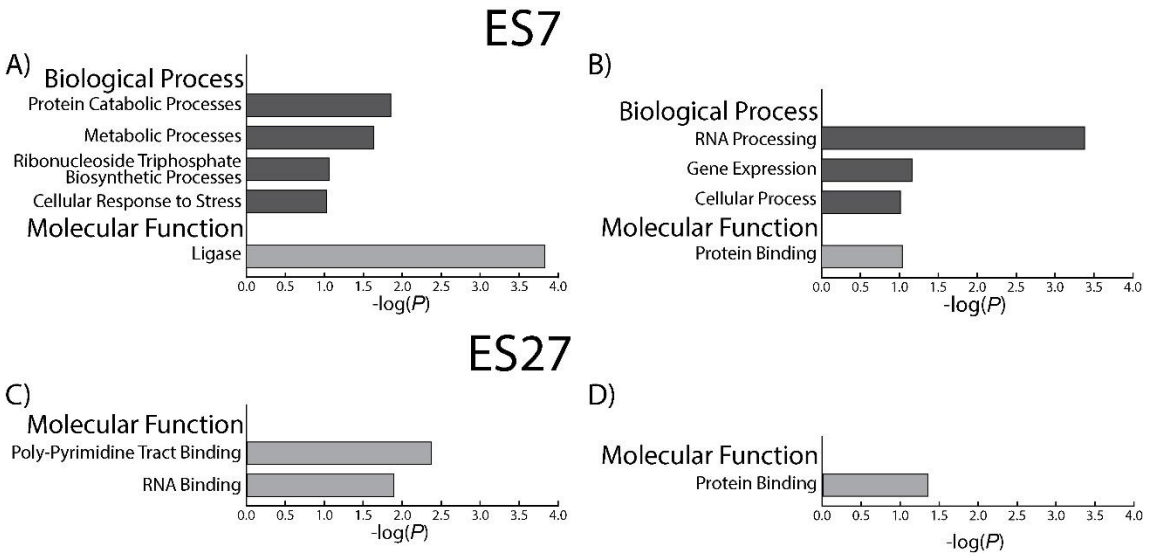


Figure 3.3 Clustering of human ES7- and ES27-associated proteins obtained from pull-down experiments with HEK293T and MDA-MB-231 cells based on biological process and molecular function. Clusters are listed in order of their *P*-values. A greater –log(*P*) corresponds to higher statistical significance. Clustering of proteins of A) Group $I_{ES7,H}$, B) Group $I_{ES7,M}$, C) Group $I_{ES27,H}$, and D) Group $I_{ES27,M}$. Clusters by biological process were not obtained for Group $I_{ES27,H}$ and Group $I_{ES27,M}$.

ES27-associated proteins

Human ES27 appears to be a hub for RNA binding proteins known to be involved in transcription, translation and protein folding. The interactions of ES27 with UPS components are attenuated compared to the UPS interactions with ES7.

ES27-associated proteins in HEK293T cells

Group $I_{ES27,H}$ contains sixteen proteins (Table B.10). Seven Group $I_{ES27,H}$ proteins are related to regulation of RNA transcription and protein translation. These include RNA-binding protein Musashi homolog 1 and 2 (MSI1H and MSI2H), putative ATP-dependent RNA helicase DDX11-like protein 8 (D11L8), B-cell CLL/lymphoma 9-like protein (BCL9L), Zinc finger CCCH-type antiviral protein 1 (ZCCHV), Phenylalanine tRNA ligase (SYFB) and Transcription intermediary factor 1-alpha (TIF1A). Only two Group $I_{ES27,H}$ proteins are involved in the UPS. These are ubiquitin carboxyl-terminal hydrolase 7 (UBP7) and TIF1A, which has dual functionality and is involved in both transcription and protein degradation (172-174). The rest of Group $I_{ES27,H}$ is comprised of proteins with a broad array of functions in metabolism. These proteins include ADP-ribosylation factor GTPase-activating protein 2 (ARFG2), which plays a role in COPI vesicle formation (175) and Importin subunit alpha-5 (IMA5), which is involved in nucleo-cytoplasmic protein export (176-177).

Like Group $I_{ES27,H}$, Group $II_{ES27,H}$ contains proteins involved in cellular organization and RNA processing and metabolism (Table B.11). Group $II_{ES27,H}$ contains twenty UPS-associated proteins. Group $II_{ES27,H}$ also contains Eukaryotic translation initiation factor 4E transporter (4ET) and Translation initiation factor eIF-2B subunit gamma (EI2BG). Several translation initiation factors are also contained in Group $II_{ES27,M}$

(below). Other proteins in Group $II_{ES27,H}$ include DICER, Ribosomal biogenesis protein LAS1L, two Ras-related proteins and several proteins involved in RNA splicing. Consistent with results with ES7 and HEK293T cells, few aaRSs were identified here (AlaRS, PheRS and GlyRS). FMR1 was not observed either in Group $I_{ES27,H}$ or Group $II_{ES27,H}$.

Here ES27 was cleaved into two fragments, corresponding to ES27a and ES27b (Figure 3.1 and 3.2), each of which contains a single tentacle, in an attempt to determine the locus of RNA-protein interactions within ES27. However, isolated ES27a is poorly behaved and did not tend to fold properly. Therefore, pull-down experiments were conducted with ES27b only. Further information is provided in Appendix B.

ES27-associated proteins in MDA-MB-231 cells

Group $I_{ES27,M}$ contains seven proteins (Table B.12). Two of these, Splicing factor 1 (SF01) and protein SON (SON) are involved in pre-mRNA splicing (178-179). Other two proteins, Transcription factor BTF3 (180) and Eukaryotic translation initiation factor 3 subunit J (EIF3J), are involved in transcription and translation (181). Several subunits of the eukaryotic translation initiation factor 3 (eIF3) complex were identified in Group $II_{ES27,M}$, including subunits B, D and E. eIF3 is known as a binding partner for members of the eukaryotic translation initiation factor 4 (eIF4) complex (182). Two members of the eIF4 complex, eIF4E2 and eIF4G2, were identified in Group $II_{ES27,M}$. Our experiments do not distinguish whether the binding of eIF4 to ES27 is direct or mediated through intermediate protein factors. Similarly, Group $II_{ES27,M}$ contains Eukaryotic translation initiation factor 2 subunit 2 (IFB2) and Eukaryotic translation initiation factor 5B (IF2P). Fragile X mental retardation syndrome-related protein 1 (FXR1) which associates directly

with FMR1 (183) also forms part of Group $II_{ES27,M}$. FMR1 was identified in Group $II_{ES7,H}$ and Group $II_{ES7,M}$ (above).

It has been suggested that eIF3 interacts with proteins in the proteasome core particle (181, 184). Ten UPS- or UBL-associated proteins are contained in Group $II_{ES27,M}$ (Table B.13). Other proteins with affinities for the UPS in Group $II_{ES27,M}$ include SF01, Heterogeneous nuclear ribonucleoproteins C1/C2 (HNRPC), Peptidyl-prolyl cis-trans isomerase B (PPIB) and Septin-9 (SEPT9). SAE2, ULA1 and FRX1 identified in association with ES7 (above) are also contained in Group $II_{ES27,M}$.

Clustering of ES27-associated proteins

Group $I_{ES27,H}$ and Group $I_{ES27,M}$ (Figure 3.3C and 3.3D) failed to cluster by biological process due to the small number of proteins identified. Molecular function of HEK293T proteins yielded two clusters, i) poly-pyrimidine tract binding and ii) RNA binding. Clustering of MDA-MB-231 proteins yielded a single group, which is protein binding.

3.3.2 Secondary structure of human ES7

The secondary structure of isolated human ES7 was characterized by SHAPE footprinting (139-140) in conjunction with computational folding (141) and secondary structure homology modeling. SHAPE provides a powerful test for secondary structural models. Nucleotides that are paired within helices are mostly unreactive, whereas nucleotides in bulges, loops, mismatches and other unpaired regions are reactive (Figure 3.4A). Ultimately, approximately 91% of nucleotides were characterized, showing SHAPE reactivities consistent with the secondary structure of Beckman and coworkers (19),

confirming the validity of their secondary structure model (Figure 3.4A). Metazoan ESs are characterized by high CG content, which pose special problems to biophysical characterization, including SHAPE characterization. CG-rich RNA can be difficult to synthesize and here required iterative rounds of primer design, synthesis and testing.

3.3.3 Magnesium-dependent interactions of human ES7

The Mg^{2+}-dependence of the SHAPE reactivity of human ES7 suggests changes in conformation and molecular interactions upon addition of Mg^{2+} ions (Figure 3.4B). SHAPE reactivity in Na^{+} alone was subtracted from reactivity in the presence of Na^{+} and Mg^{2+}. The resulting heat map (Figure 3.4C) indicates that helical regions are generally invariant to addition of Mg^{2+} ions. The Mg^{2+}-dependence of the SHAPE reactivity correlates with locations of helical defects. Nucleotides with high intrinsic SHAPE reactivity in the presence of Na^{+} only are more likely to interact with Mg^{2+} ions, indicated by either increases or decreases in SHAPE reactivity. This pattern of Mg^{2+} ion dependence on SHAPE reactivity has been observed previously (125, 185-186).

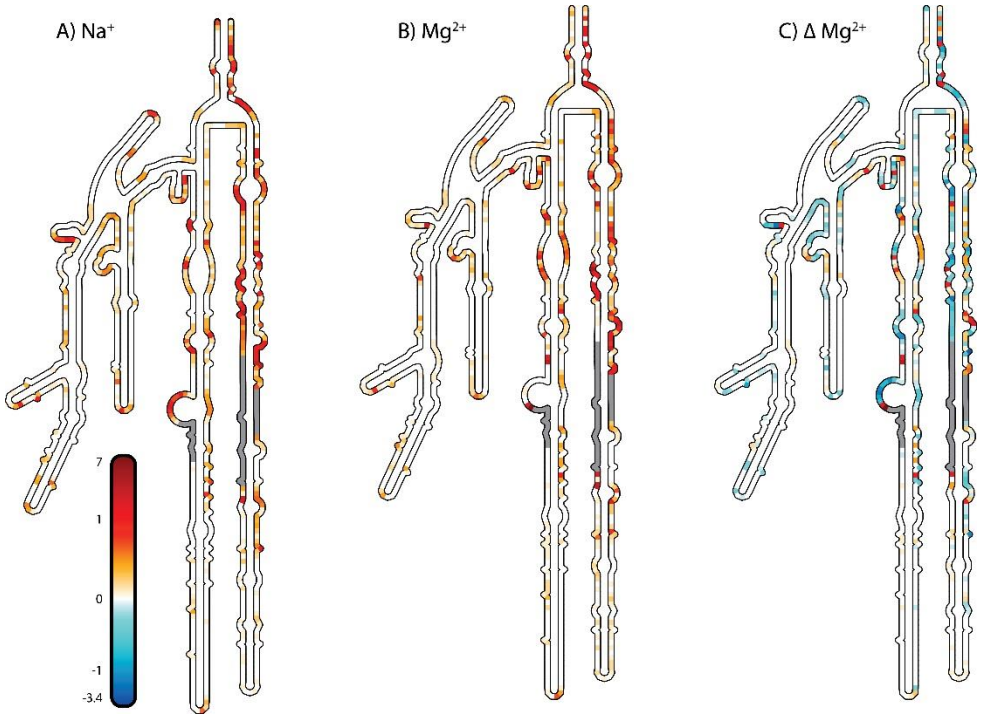


Figure 3.4 SHAPE reactivities of isolated human ES7 rRNA. SHAPE reactivities in the presence of A) Na^+ alone, or B) Na^+ and Mg^{2+} are mapped onto the ES7 secondary structure. Orange or red indicates nucleotides with moderate to high SHAPE reactivity. White indicates low SHAPE reactivity. No SHAPE data could be obtained for regions in gray. C) Mg^{2+}-induced changes in SHAPE reactivity mapped onto human ES7 rRNA secondary structure. Positions with positive values (red) indicate increased reactivity upon addition of Mg^{2+}. Positions with negative values (blue) indicate decreased reactivity.

3.3.4 Thermal unfolding of human ES7

Human ES7 is extremely stable because of its high CG content. In low salt conditions the unfolding transition temperatures are accessible (Figure 3.5, Table B.5). The unfolding of human ES7 proceeds in a sequential multi-state process (Figure B.1) as observed previously for ES7 of *S. cerevisiae* (161) and other large RNAs (143, 187). Distinct unfolding transitions are consistent with the proposed ES7 secondary structure, which contains resolved secondary structural elements.

The best fit of the human ES7 unfolding data to a melting model suggests seven distinct transitions (Figure 3.5) in the presence of Na^+ ions. Thermodynamic parameters

for melting were estimated by fitting the unfolding data to a multiple state model with independent transitions as described previously (161, 187). Derivative plots of melting profiles were baseline corrected, fit to multiple melting transition models, and analyzed for component transitions using IGOR Pro. The most probable model is assumed to be the simplest model, with the fewest transitions, that yielded good fit of model to data. More complex folding models, with greater numbers of transitions, cannot be excluded by the methods used here.

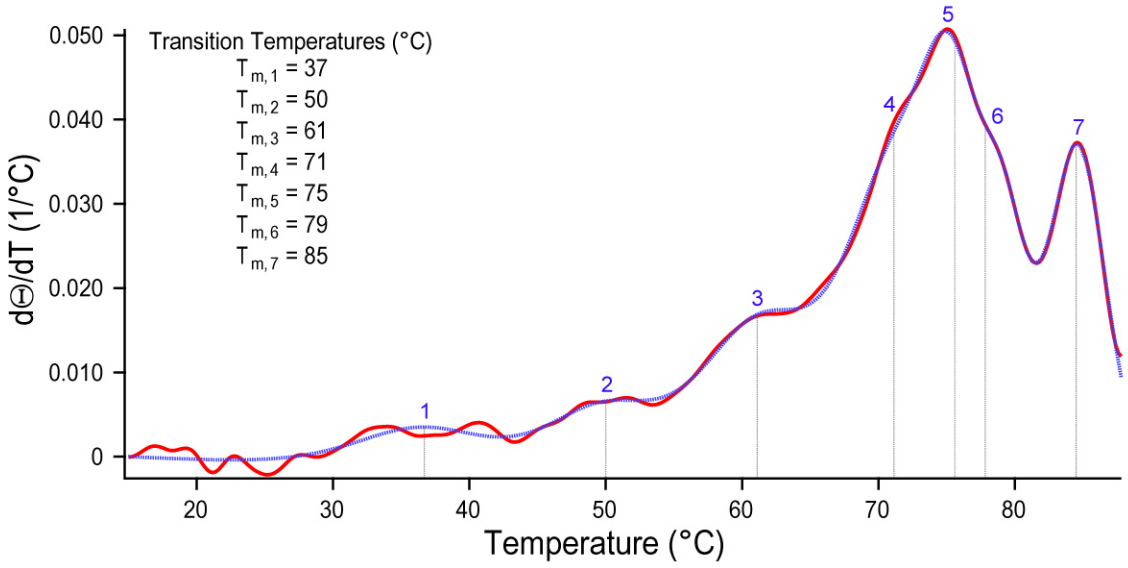


Figure 3.5 Melting profile of human ES7 rRNA. The derivative of fraction unfolded with respect to temperature was computed in NaCl at 260 nm. The derivative was fit to a non-sequential, independent transition model with IGOR Pro. The enthalpy (H_n), melting temperature ($T_{m,n}$) and the relative hyperchromicity (A_n) were estimated for each transition, n. The observed derivative is red, the fit is dotted blue and the estimated melting temperature for each transition is highlighted with dotted gray lines. Data were obtained at 18 mM NaCl and 20 mM Tris-HEPES (pH 8.0).

The melting temperatures, $T_{m,n}$, (where n indicates the transition number) of ES7 at 18 mM Na^+ were estimated as 37, 50, 61, 71, 75, 79 and 85 °C (Figure 3.5). The $T_{m,n}$'s are

independent of RNA concentration, indicating that the unfolding is unimolecular and that the rRNA does not self-associate before or during the unfolding process under the conditions of the experiments. Superimposition of repeated RNA folding and unfolding curves confirms that the rRNA is effectively at equilibrium during unfolding and that the processes are reversible. $T_{m,n}$ estimations in the presence of different ions were also performed (Appendix B, Figure B.2, Table B.5). In general, there is no evidence of monovalent ion-specific interaction with particular areas of ES7, instead there is a comprehensive ion-dependence in ES7 as evidenced by the complete shifting of the melting curves.

3.3.5 Putative G-quadruplexes in human rRNA expansion segments

The sequence of human ES7 rRNA suggests potential G-quadruplex formation. Two segments of ES7 contain $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ motifs (where N is any nucleotide). Probabilities of G-quadruplex formation can be estimated with QGRS Mapper (188), where sequences are ranked by their 'G-score'. G-scores provide rough estimates of probabilities of forming G-quadruplexes based on number of possible guanine tetrads, length of loops, and length uniformity within loops. Higher G-scores indicate more probable G-quadruplex formation.

One segment of ES7 ($Gq1_{ES7}$, nts 583-652), contains ten tandem G-tracts ranging from three to five nucleotides. Within $Gq1_{ES7}$, the sequence 5' GGGGCCGGGGGUGGGGUCGGCGGGG 3' (nts 623-647) gives a G-score of 60 and the sequence 5' GGGUGCGGGGGUGGGCGGG 3' (nts 603- 621) gives a G-score of 40. A second segment of ES7 ($Gq2_{ES7}$) contains the sequence 5'

GGGAGGGCGCGCGGGUCGGGG 3' (nts 829- 849) which gives a G-score of 38. $Gq1_{ES7}$ and $Gq2_{ES7}$ have G-scores similar to those of established RNA G-quadruplex forming sequences. MT3-MMP, in the 5'-UTR of the mRNA of MMP16, and ARPC2, in the 5'-UTR of the mRNA of ARPC2 (189), gives a G-score of around 40. However in ES7, the complementary strand, which is absent in mRNAs, would probably decrease the possibility of G-quadruplex formation. In addition, G-quadruplexes in DNA and more so in RNA are complex, polymorphic and cannot be predicted with certainty.

Experimental investigation of G-quadruplexes

We synthesized $Gq1_{ES7}$ and $Gq2_{ES7}$ along with controls $C1_{ES7}$ and $C2_{ES7}$. $C1_{ES7}$ is analogous to $Gq1_{ES7}$ in sequence composition and length, except that the G-tracts are disrupted. $C2_{ES7}$ is analogous to $Gq2_{ES7}$, also with disrupted G-tracts. These two control sequences give G-scores of 0 (Table B.2).

The CD spectra of $Gq1_{ES7}$ and $Gq2_{ES7}$ are similar to spectra of known RNA G-quadruplexes (189-193). Cation dependencies of the CD spectra of $Gq1_{ES7}$ and $Gq2_{ES7}$ are also consistent with G-quadruplex formation. It is known that G-quadruplexes are stabilized by monovalent cations in the order $K^+ > Na^+ > Li^+$ (194). $Gq1_{ES7}$ gives the same CD spectrum in Li^+ or Na^+, showing a λ_{max} at 267 nm and a λ_{min} at 212 nm. However, in K^+, the intensity of the $Gq1_{ES7}$ maximum is increased by 15% relative to Li^+ or Na^+ and shifts slightly to the blue to a λ_{max} of 265 nm, whereas the intensity of the minimum decreases by 35% (Figure 3.6A). The amplitude of the negative band follows the order $Li^+<Na^+<K^+$. The CD spectrum of the control, $C1_{ES7}$, does not show monovalent cation dependence (Figure 3.6A).

The CD spectrum of $Gq2_{ES7}$ shows a distinctive K^+ dependence that is less pronounced than for $Gq1_{ES7}$ (Figure 3.6). $Gq2_{ES7}$ shows a λ_{max} at 266 nm in Li^+, Na^+, and K^+. This peak increases in intensity by around 14% in the presence of K^+. Like that of $C1_{ES7}$, the CD spectrum of $C2_{ES7}$ does not show cation dependence (Figure 3.6B).

We searched for additional G-quadruplex forming segments in the sequences of all ESs of human LSU. ES27 emerged as the only ES, besides ES7, containing putative G-quadruplex forming sequences, with two sequences identified. Both sequences are located within ES27a. The sequence 5' GGGGAGAAGGGUCGGGGCGGCAGGG 3' (nts 3122 to 3149) gives a G-score of 40 and the sequence 5' GGGCUGGGGCGCGAAGCGGGGCUGGG 3' (nts 2910 to 2935) gives a G-score of 37.

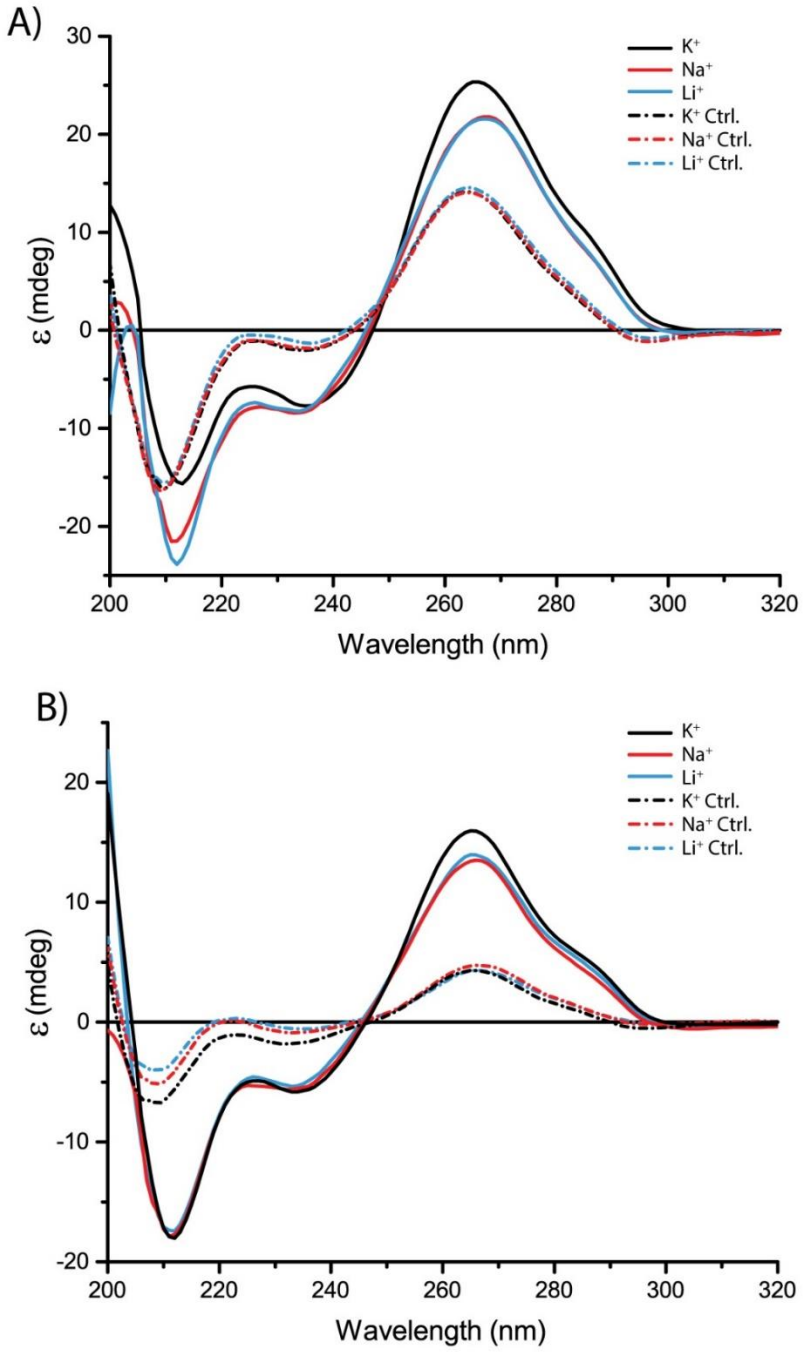


Figure 3.6 CD spectra of putative G-quadruplex forming sequences extracted from human ES7 rRNA in the presence of three different monovalent cations. CD spectra of A) $Gq1_{ES7}$, which contains 10 tandem G-tracts (solid lines) and control $C1_{ES7}$ (dashed lines) and B) $Gq2_{ES7}$, which contains 4 tandem G-tracts (solid lines) and control $C2_{ES7}$ (dashed lines). Spectra obtained in K^+ are black, Na^+ are red and Li^+ are blue.

3.4 Discussion

We believe that human ESs are dynamic and positionally polymorphic docking stations for proteins and protein assemblies. A variety of systems that enhance and extend translational function appear to be localized by association with ESs.

3.4.1 The ubiquitin-proteasome system

We observe that many components of the UPS associate with ES7 and ES27. The robustness of UPS-ESs association is underscored by observation of interactions of multiple UPS components with multiple ESs in multiple cell lines. UPS association with ESs appears modestly more extensive in HEK293T cells than in MDA-MB-231 cells. In general, the ES-protein associations observed here appear to be particularly reliable because competition with non-ES RNAs is intrinsic to these experiments. rRNA accounts for over 80% of cellular RNA (149) and is highly abundant during our assays. The proteins identified in this study survive that competition and therefore appear representative of proteins with affinity for ESs. However, we cannot exclude additional binding partners.

It is known that many proteins in mammalian cells are co-translationally degraded during quality control. Approximately 12-15% of nascent polypeptides are co-translationally ubiquitinated (195) and 30% of newly synthesized proteins are quickly degraded by the proteasome (196-197). Here we observe that the UPS is physically associated with the ribosome. Proteins involved in the UPS system bind to human ES7 and ES27 rRNA directly or through intermediates. This association suggests that an important function of the human ESs is to maintain proximity of the ribosome to the UPS.

The UPS targets proteins for degradation (198-200). Ubiquitin is activated by the ubiquitin-activating enzyme, E1. In the second step, ubiquitin is conjugated to a ubiquitin-carrier enzyme or ubiquitin-conjugating enzyme, E2. In the third step, a ubiquitin-protein ligase, E3, catalyzes the conjugation of the protein target to ubiquitin.

E1-like enzymes, like SAE2, ULA1 and UBA5, are shown here to associate with human ESs. These ubiquitin-like proteins (UBLs) share high degree of structural similarity with ubiquitin and are conjugated to target proteins in mechanisms similar to ubiquitination (201). UBLs are involved not only in protein degradation but in a variety of other functions (201-202). SAE2 activates SUMO 1, 2 and 3 (small ubiquitin-like modifiers). ULA1, which activates NEDD8 (Neural precursor cell expressed, developmentally down-regulated 8), has high sequence similarity to ubiquitin (60% identity). UBA5 activates SUMO2 and ubiquitin fold-modifier-1 (UFM1) (201, 203).

To our knowledge, there is no previous evidence for direct interactions of the UPS with cytoplasmic ribosomes. However, around ten percent of the UPS-proteins identified here, across both cell lines and RNAs, were also identified by Baltz in the mRNA-bound proteome in HEK293T cells (204). These proteins are exclusively ubiquitination and deubiquitination enzymes. UBLs and proteasome components were not identified in the mRNA-bound proteome.

The interdependence and tight linkage of protein synthesis and protein degradation might have broad implication in disease and development. The UPS is involved in ribosome biogenesis and required for maturation of the 90S pre-ribosome particle (205). Many factors involved in proteasome maturation are also involved in ribosome biogenesis

and it has been suggested that participation redundancy of these factors could link these systems (206).

3.4.2 Other ES-associated proteins

Many additional proteins appear to associate with ES7 and ES27 in both non-carcinomic HEK293T cells and breast cancer MDA-MB-231 cells. All of these non-ES specific proteins have been shown previously to be directly involved in RNA processing, protein synthesis, degradation or modification. These proteins (MSI1H, SAE2, ULA1, ZCCHV and Protein phosphatase methylesterase 1 (PPME1)) range in size but uniformly contain open twisted α/β domains. ZCCHV, which contains a zinc finger, is an exception. Additional factors might assist in regulating protein binding and positioning of human ESs.

In HEK293T cells, ES7 and ES27 also associate with proteins involved in transcriptional and translational regulation. Differences in human ES7 and ES27 protein association are most pronounced in MDA-MB-231 cells, where proteins involved in mRNA stability, mRNA decay and nucleo-cytoplasmic trafficking prefer association with ES7, and proteins involved in pre-mRNA splicing, transcription, and translation prefer association with ES27.

3.4.3 Folding of isolated ESs

The results obtained for human (here) and *S. cerevisiae* ES7 (161) with SHAPE, computational folding, homology modeling and thermal melting suggest that isolated ESs fold to native-like structures. By analogy with ES7 we assume that isolated ES27 also folds to a native-like state. Our SHAPE results reinforce the role of magnesium in the folding of

essentially all large, complex RNAs (120, 207-208). Thermal melting shows that ES7 folds to a stable native state *via* a fully reversible process. The thermodynamic parameters of melting displayed mild dependence on monovalent cations (Appendix B). Previous experimental structures of the human ribosome do not fully capture the positions and orientations of the tentacle regions of the ESs. The tentacles appear to be dynamically disordered in solution and statically disordered during structure determination.

3.4.4 G-quadruplexes

Human rRNA tentacles contain sequences that suggest the potential to form G-quadruplexes. These sequences possess high scores in QGRS Mapper (188), a program designed to predict G-quadruplexes. Experimental support for G quadruplex formation is seen in the K^+ dependence of CD spectra of a subset of single-stranded ES fragments. Finally, proteins with affinity for G-quadruplexes were observed to associate with ES7. These proteins include FMR1 and FXR1 (170-171), which are known to associate with the human LSU (209). Additional proteins observed here known to bind G-quadruplexes include Tropomyosin alpha 4, RNA binding protein 14, EF-hand domain-containing protein D2, Serine/arginine-rich splicing factor 1, Serine/arginine-rich splicing factor 9, Heterogeneous nuclear ribonucleoproteins and ribosomal proteins uL11, uL30, uS5 and eS6 (189, 210). It seems possible that these proteins associate with the LSU via interactions with G-quadruplexes in ESs.

Combined, the results from computation, CD spectroscopy and pull-down experiments suggest that human ESs could form G-quadruplexes that interact with specific

proteins *in vivo*. However, the results remain suggestive because we have not unequivocally demonstrated the stability of G-quadruplexes in human rRNA.

3.4.5 Clustering

DAVID (135-136) was used to cluster of ES7- and ES27-associated proteins by biological process and by molecular function. Among ES7-associated proteins in HEK293T, four clusters were detected by biological process: i) protein catabolic processes, ii) metabolic processes, iii) ribonucleoside triphosphate biosynthetic processes and iv) cellular response to stress. Ligase was the only detected molecular function cluster (Figure 3.3A). ES7-associated proteins in MDA-MB-231 clustered by biological process to: i) RNA processing, ii) gene expression and iii) cellular process. Protein binding was the only molecular function cluster detected (Figure 3.3B). ES27-associated proteins in HEK293T clustered by biological process clusters to: i) poly-pyrimidine tract binding and ii) RNA binding (Figure 3.3C), while in MDA-MB-231, protein binding was the only cluster detected (Figure 3.3D).

3.5 Conclusions

In previous work we hypothesized that expansion segments could be sites of functional interactions with other macromolecules and used ES7 of *S. cerevisiae* as a model (161). This work expands upon that knowledge by investigating the stability and protein binding partners of human ESs. Here, we demonstrate that proteins associated with the UPS bind to human ESs. In addition, binding of UPS-associated proteins seems to depend on cell line, which suggest that protein binding to ESs could also depend on a variety of

factors, like the stage of the cell life cycle. It seems likely that ESs could have different functions in different organisms. The specificity in sequence and structure of human ESs identified unique protein binding partners not identified in yeast ESs and suggest that expansion segments could have different roles in protein synthesis or quality control in different species.

CHAPTER 4

RIBOSOMAL EXPANSION SEGMENTS AS ANTIMICROBIAL TARGETS

4.1 Introduction

Many antibiotics bind to regions of rRNA that are universal to all organisms. This universal rRNA, known as the common core (21-22), contains the A site, the P site, the exit tunnel and the peptidyl transfer center (211). These functional centers of the ribosome are targeted by macrolides, aminoglycosides, tetracyclines, streptogramins, pleuromutilins, chloramphenicol, linezolid and puromycin. Increasing resistance to these antibiotics (212-213) provides a compelling rationale for seeking new antibiotic targets (214).

Eukaryotic pathogens, including various fungi, are gaining drug resistance in parallel with bacterial pathogens. Mortality caused by eukaryotic pathogens has been increasing since the 1980s (215). Unlike those of prokaryotes, eukaryotic ribosomes contain large amounts of species-specific rRNA in addition to the common core. Eukaryotic rRNA expansion segments (ESs) (3, 21, 25, 32-33, 114) are substantially unexplored as drug targets. ESs emerge from the surface of the common core (21) and are characterized by idiosyncratic structures and functions that are essential and specific to groups of species. ESs are important for ribosome biogenesis (115-116), rRNA processing and stabilization (47, 60), translation initiation and recruitment of factors such as chaperones, enzymes, aminoacyl tRNA synthetases (aaRSs) and signal recognition particles (68-70, 161).

Here we investigate ES7 of *Candida albicans* ($ES7_{CA}$) as a model for eukaryotic drug targets. *C. albicans*, along with *Cryptococcus neoformans*, *Coccidioides immitis* and *Pneumocystis jiroveci* affect immunocompromised populations including AIDS and cancer patients, organ-transplant recipients, low-birth-weight infants and individuals with inherited diseases (216-217). Candidiasis is the fourth leading cause of healthcare associated blood infections in the US, with higher mortality rate than infections caused by well-known bacteria including *E. coli* and *Enterobacter* species (218). *C. albicans* and several other *Candida* species show resistance to antimicrobial agents such as fluconazole (219-220).

ES7 of fungi averages around 210 nucleotides in length with conserved GC content and secondary structure (Table C.3, Figure C.1). $ES7_{CA}$ is 209 nucleotides with ~55% GC content (*C. neoformans*, 234 nts, ~50% GC; *C. immitis*, 210 nts, ~60% GC; *P. jiroveci*, 223 nts, ~50% GC). By contrast, ES7 of *Homo sapiens* ($ES7_{HS}$) is four times larger (876 nts) than $ES7_{CA}$ and is different in sequence, composition (~80% GC) and structure.

Here we use $ES7_{CA}$ to investigate the general utility of ESs as targets for chemotherapeutic agents. We demonstrate that isolated $ES7_{CA}$, like ES7 from human ($ES7_{HS}$) (Chapter 3) and *S. cerevisiae* ($ES7_{SC}$) (161), folds reversibly to a native state. Aminosugars are a class of compounds well known for their binding to the bacterial ribosome. An assay to screen for binding of an amino sugar library (peptidic aminosugars) to ES7 was developed. The affinities of peptidic aminosugar (PA) conjugates (Figure 4.1) for $ES7_{CA}$ were assessed with fluorescence displacement (FD) assays and differential thermal melting studies. Members of the PA library with high binding affinity for $ES7_{CA}$ were investigated further to determine the minimum inhibitory concentration (MIC)

needed to induce mortality on *C. albicans*. PAs with the lowest MIC values were tested for cytotoxic effects in HEK293T cells. Molecules that bind with high affinity and specificity to $ES7_{CA}$ *in vitro*, in comparison to $ES7_{HS}$, induce mortality in *C. albicans* but not in HEK293T cells. The results are consistent with the hypothesis that ESs represent useful targets for development of new chemotherapeutics directed against eukaryotic pathogens.

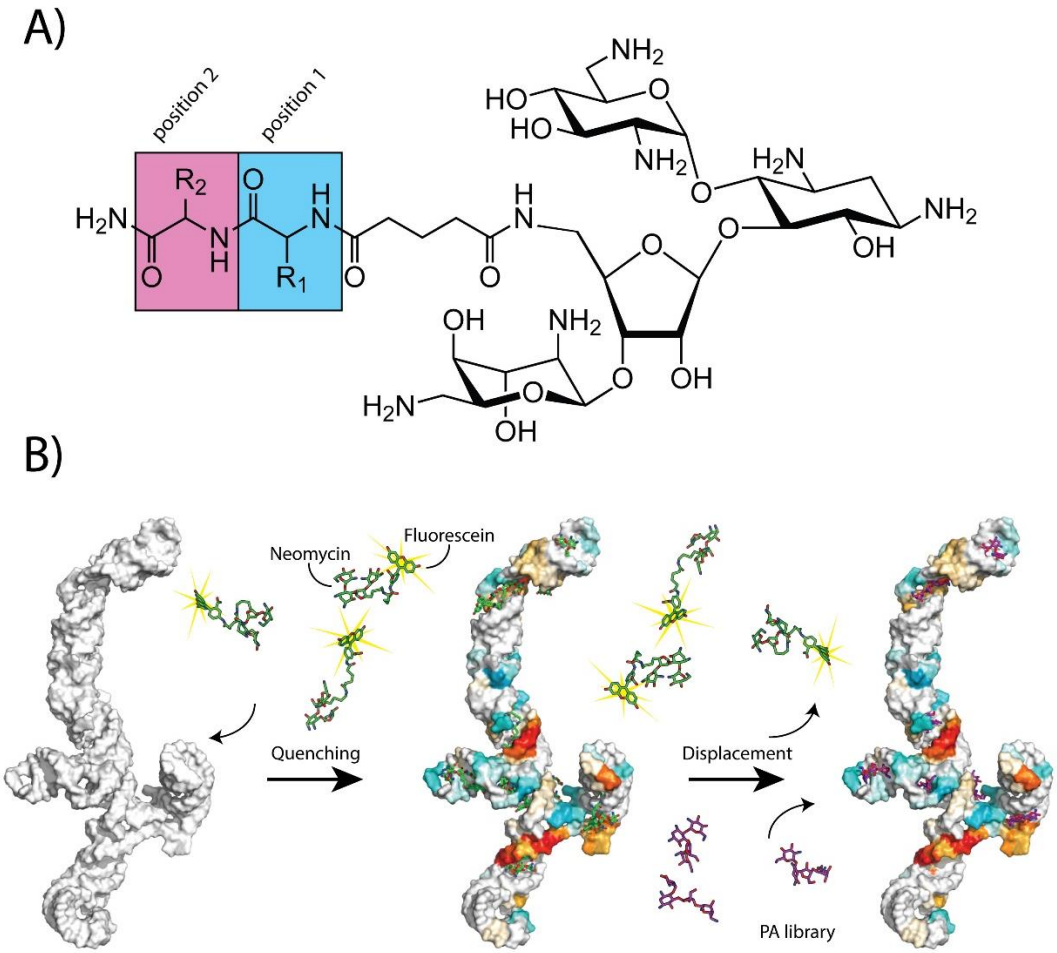


Figure 4.1 Affinities of a library of peptidic aminosugar (PA) conjugates for $ES7_{CA}$ were investigated by a fluorescence displacement (FD) assay. A) Generic structures of peptidic aminosugar conjugates. Position 1 is highlighted in blue. Position 2 is highlighted in pink. R_1 and R_2 are amino acid sidechains. B) Schematic diagram of the FD assay showing displacement of F-neo (green) by a competitive ligand (PA-conjugates, purple) from a three dimensional model of $ES7_{CA}$. The yellow star indicates fluorescence.

4.2 Materials & Methods

4.2.1 *C. albicans* and human ES7 DNA and RNA

The DNA encoding $ES7_{CA}$, appended to the T7 promoter and restriction sites for EcoRI (5' end) and HindIII (3' end), was constructed by recursive PCR (221). The DNA was inserted into previously restricted, dephosphorylated pUC19 plasmid. Ligated plasmid was transformed into DH5α competent cells and colonies were selected based on X-gal Blue-White screening (Thermo Scientific). Oligomer and primer sequences used in the recursive PCR are shown in Tables C.1 and C.2. Oligomers, primers and DNA sequencing were obtained from Eurofins-MWG Operon.

The Weeks cassette (118) was incorporated onto the 3' end of $ES7_{CA}$ with the Q5® Site-Directed Mutagenesis Kit (New England BioLabs) to allow chemical mapping of the entire $ES7_{CA}$. Primer sequences are shown in Table C.4. The $ES7_{HS}$ coding sequence was constructed as described previously (in Chapter 3). $ES7_{CA}$ and $ES7_{HS}$ RNAs were transcribed using the HiScribe™ T7 High Yield RNA Synthesis Kit (New England BioLabs) as described previously (161).

4.2.2 SHAPE

Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE) methods were adapted from published protocols (118-119). SHAPE monitors local nucleotide flexibility through reactivity of 2' oxygens upon addition of an electrophile. Experiments and data analysis on $ES7_{CA}$ were performed as described previously (161). Data were collected in 200 mM Na^+ and no Mg^{2+} to promote and analyze secondary structure

formation (120-122). RNA folding and long-range tertiary interactions were characterized in the presence of Mg^{2+} ions.

Small-molecule SHAPE was performed by addition of neomycin (Neo) to the folding buffer without further changes to the experimental procedure. Since the hydroxyl groups of the neomycin ribose might compete for adduct formation with the 2'-hydroxyl groups in the RNA ribose, increased Neo was used (10 mM). $ES7_{CA}$ secondary and tertiary structures were studied in the presence and the absence of Neo.

4.2.3 Thermal folding/unfolding

Melting profiles for $ES7_{CA}$ RNAs were determined by monitoring the absorbance at 260 nm as a function of temperature with a thermostated Varian Cary-1E UV spectrophotometer. RNA was concentrated to reach an initial absorbance of 0.30-0.33. The melting buffer consisted of 180 mM NaCl, 20 mM Tris-HEPES (pH 8.0) and Neo at various concentrations. Samples were initially heated to 95 °C with a ramp of 3 °C/min. Melting transitions were obtained by repeated heating and cooling at 0.5 °C/min. The experimental temperature ramp was slower than normal (1 °C/min) because of the size of the RNA (127). To prevent in-line cleavage, Mg^{2+} was excluded from melting buffers (144). Superimposition of heating and cooling profiles was used to confirm equilibrium during melting. Thermodynamic model information as well as information on fitting and parameter estimation are described elsewhere (161).

4.2.4 Interactions of $ES7_{CA}$ with small molecules

A 215 compound peptidic amino sugar (PA) library, containing Neo as the model aminosugar (222), was screened against $ES7_{CA}$ using a fluorescence displacement assay (FD) described by Watkins et al. (223). Fluorescein-neomycin (F-neo) was used as the reporter molecule and synthesized as described elsewhere (223).

4.2.4.1 K_d's of neomycin and ES7

Binding studies of F-neo and $ES7_{CA}$ were performed in triplicates in binding buffer [50 mM NaCl, 0.4 mM EDTA, 10 mM HEPES (pH 7.0)]. F-neo concentration was maintained constant in a 96 well-plate (10 nM) and titrated with serial dilutions of $ES7_{CA}$ RNA (1.58E-8 M to 7.63E-12 M, 2x dilutions). Binding of F-neo to $ES7_{HS}$ RNA was performed in duplicates in binding buffer. Similarly, F-neo was maintained constant (100 nM) and titrated with serial dilutions of $ES7_{HS}$ RNA (25E-8 M to 1.22E-8 M, 2x dilutions). F-neo excitation was carried out at 485 nm and emission was measured at 525 nm using the Genios Pro plate reader (Tecan). Emission of fluorescence intensity decrease with increasing concentration of RNA since the interaction between F-neo and RNA quench the emission of fluorescence.

The stoichiometry of the F-neo-ES7 complex was determined from saturation plots. *K_d's* were calculated from the mass balance describing the relationship between bound F-neo, $[\text{F-neo}]_b$, and the total concentration of F-neo, $[\text{F-neo}]_{total}$, and RNA in the well, $[\text{RNA}]_{total}$,:

$$[\text{F-neo}]_b^2 - [\text{F-neo}]_b ([\text{F-neo}]_{total} + [\text{RNA}]_{total} + K_d) + [\text{F-neo}]_{total}[\text{RNA}]_{total} = 0$$

The observed emission, I, is the combination of the emission intensities from bound F-neo, I_b, and free F-neo, I_f, and provides the relationship to obtain the concentration of F-neo bound:

$$I = I_b [F\text{-}neo]_b + I_f [F\text{-}neo]_f$$

4.2.4.2 Small molecule high throughput screening

High throughput screening (HTS) was performed in duplicates in binding buffer using the FD assay (223). The PA-conjugates (1 µM final concentration) were added to the F-neo-ES7 complex (100 nM F-neo and 11.6 nM ES7 RNA, final concentrations), previously aliquoted into 96 well-plates. All plates contain two set of controls, the F-neo-ES7 complex and the F-neo-ES7 complex with Neo added. The displacement of F-neo by PA-conjugates was measured by the increase in fluorescence after addition of PAs and compared against the fluorescence of the F-neo-ES7 complex (ΔF). The percent of displacement of F-neo by PAs was calculated from the change in fluorescence by the addition of PAs (ΔF_{PA}) divided by the change in fluorescence obtained by the addition of Neo to the F-neo-ES7 complex (ΔF_{Neo}).

$$\% \text{ binding} = (\Delta F_{PA}/\Delta F_{Neo}) \times 100$$

4.2.4.3 IC_{50} experiments

IC_{50} experiments were performed using the FD assay. The F-neo-ES7 complex was maintained at constant concentration (100 nM F-neo and 11.6 nM ES7 RNA, final concentrations) and titrated with PA-conjugates in a 96 well-plate. Curve fitting was performed with the sigmoid function of Igor Pro. The IC_{50} is given by the mid-point value in the x-axis obtained from the curve fitting.

4.2.5 Determination of minimum inhibitory concentration

Stocks of PA-conjugates were stored at 1 µM in purified water at -80 °C until use. *C. albicans* SC5314 was cultured in yeast extract-peptone-dextrose (YPD) agar plates at

30 °C for 24 h. Colonies were picked and diluted in YPD media to read an OD_{600} ~ 0.1. Eighty µL of cultivated media were aliquoted in 96-well clear flat bottom plates. Double-fold dilutions of PAs were performed in water and 20 µL of diluted compound was added to each well (200, 100, 50, 25, 12.5, 6.25, 3.13, 1.57, 0.78, 0.39, 0.20 and 0.10 µM, final concentrations) to obtain a final volume of 100 µL per well. Controls in each plate include 20 µL of water and 80 µL of YPD media to monitor sterility, and 20 µL of water and 80 µL of cultivated YPD to measure standard microbial growth. All plates were continuously shaken at 30 °C for 24 h in a Synergy™ H4 Hybrid Multimode plate reader (Biotek). Absorbance at 600 nm was measured every 10 min. The percent of microbial growth was calculated by monitoring the change in absorbance of the culture with PA ($A_{600,PA}$) compared with the absorbance of media only ($A_{600,media}$) divided by the change in absorbance of the microbial culture ($A_{600,CA}$) and media:

$$\% \text{ growth} = (A_{600,PA}-A_{600,media})/ (A_{600,CA}-A_{600,media}) \times 100$$

MIC values were determined after incubation of cells with PA-conjugates for 24 h and growth rates of ≤ 5 %.

4.2.6 Cytotoxicity studies

HEK293T cells were grown in Dulbecco's Modified Eagle's Medium, supplemented with 10% fetal bovine serum. Cells were incubated at 37 °C in a humidified atmosphere of 5% CO_2. Three thousand cells were seeded in 96-well plates (100 µL total vol.). Cell adhesion to the plates was allowed in an initial incubation period of 24 h. PA-conjugates were added to each well (100 µM final concentrations) and incubation was

performed for 72 h. The CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega) was used to determine cell viability. Absorbance was measured at 490 nm.

4.2.7 Three dimensional modeling of $ES7_{CA}$

The three dimensional structure of ES7CA was generated using RNAComposer (http://rnacomposer.cs.put.poznan.pl) in its automated mode (224). The following user-defined secondary structure was used as an input:

((((((((((.....((((..((((.((((.((.....)).)))).....)))).))))..))))..((((((((((((((((.(((((((.....(.((((((.....((((((....
..))))))).....)))))).)....)))))..)))))))))).....)))))))))). Ten 3D models were predicted and subsequently ranked based on the final energy of the 3D structure (225). The best structures were evaluated for the conservation of the introduced pairing patterns (http://rnapdbee.cs.put.poznan.pl) and used in docking studies (226).

4.3 Results

The affinities of peptidic aminosugars (PAs) for $ES7_{CA}$ were assessed with FD assays, SHAPE and thermal melting. Members of the PA library with high affinity for $ES7_{CA}$ were further investigated to determine the minimum inhibitory concentration (MIC) needed to induce mortality on *C. albicans*. PAs with the lowest MIC values were tested for cytotoxic effects in HEK293T cells.

4.3.1 Affinities of F-neo for $ES7_{CA}$ and $ES7_{HS}$

F-neo binds with high affinity and specificity to $ES7_{CA}$ relative to $ES7_{HS}$. The extent of binding was estimated from changes in fluorescence of F-neo upon titration with RNA. The dissociation constant (K_d) for binding to $ES7_{CA}$ was estimated from non-linear fitting

of data to mass balance equations. The K_d of F-neo for $ES7_{CA}$ is 2.5 x 10^{-9} M ± 9 x 10^{-10} M (Figure C.2A). The K_d of F-neo for $ES7_{HS}$ could not be accurately determined because the affinity is lower than the practical limits of the assay. A limiting K_d could be estimated. The K_d of F-neo for $ES7_{HS}$ is estimated to be greater than 10 μM (Figure C.2B) The affinity F-neo for $ES7_{CA}$ is 4,000 times greater than for $ES7_{HS}$. By contrast K_d's for association of Neo with A site RNA, its cognate binding site, were previously estimated to be 0.053 μM for *E. coli* and 0.26 μM for human (227). This difference (4-fold) is much less than the difference observed for association of F-neo and $ES7_{CA}$ versus the association of F-neo and $ES7_{HS}$.

The stoichiometry of binding of neomycin (Neo) to $ES7_{CA}$ was estimated from saturation binding plots of F-neo (228). Fluorescence of F-neo during $ES7_{CA}$ titration was plotted as a function of the $ES7_{CA}$ to F-neo molar ratio. The inflection point at ~0.077 suggests an approximate stoichiometric ratio of Neo to $ES7_{CA}$ of 13:1 (Figure C.3). The binding stoichiometry for $ES7_{HS}$ could not be obtained due to the low affinity of F-neo for $ES7_{HS}$.

4.3.2 The Library

We investigated the affinity of members of a library of PAs (222) for $ES7_{CA}$ using several methods. The library consists of mono and dipeptides linked by glutarate to Neo at the 5' position of the ribose ring. The amino acid positions are numbered from the N-terminus and are linked at their N-terminus. Some members of the PA library are mono-peptidic, with amino acids only at position one and some are di-peptidic, with amino acids at positions one and two (Figure 4.1).

4.3.3 High throughput screening of the PA library for affinity for $ES7_{CA}$

Binding of PAs to $ES7_{CA}$ was estimated through high-throughput screening (HTS) using a FD assay (Figure 4.1 and 4.2) (222-223). In this assay, $ES7_{CA}$ is first incubated with F-neo, which is then competitively displaced by a PA. The assay allows ranking of members of the PA library, and is normalized so that F-neo displacement by Neo equals 100%. PAs with displacements greater than 100% bind to $ES7_{CA}$ with greater affinity than Neo. PAs with displacements less than 100% bind with less affinity than Neo. Sixty three compounds from the PA library showed greater affinity than Neo for $ES7_{CA}$ in the HTS (Figure C.4). Four of these are single amino acid PAs; the other 59 are two amino acid PAs.

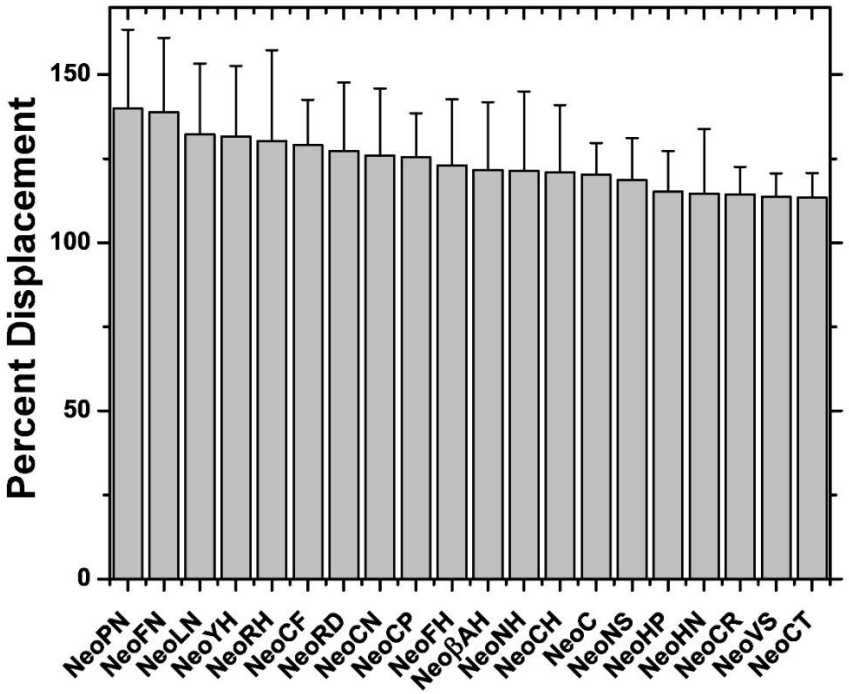


Figure 4.2 Ranking of the PA-conjugates by their tendency to displace F-neo from $ES7_{CA}$ relative to displacement of F-neo by Neo. The 20 molecules that most effectively displace

F-neo are shown here. Each experiment was performed in duplicate. Standard deviation for each compound is shown with error bars.

4.3.4 IC_{50} measurements of affinity for $ES7_{CA}$

IC_{50} measurements of PA binding with $ES7_{CA}$ were used to validate selected HTS results. IC_{50} experiments were performed with three compounds from the top five HTS binders (Figure 4.3). In the IC_{50} assay, complexes of F-neo and $ES7_{CA}$ were titrated with PA-conjugates and displacement of F-neo was monitored at various ratios of PA to $ES7_{CA}$. In comparison to Neo, binding affinities are 2.2-fold greater for NeoRH, 1.5-fold greater for NeoYH and 1.3-fold greater for NeoFN. IC_{50} measurements confirm the high affinity of these molecules for $ES7_{CA}$. In contrast to HTS, IC_{50} experiments demonstrate that NeoRH is a better binder than NeoFN.

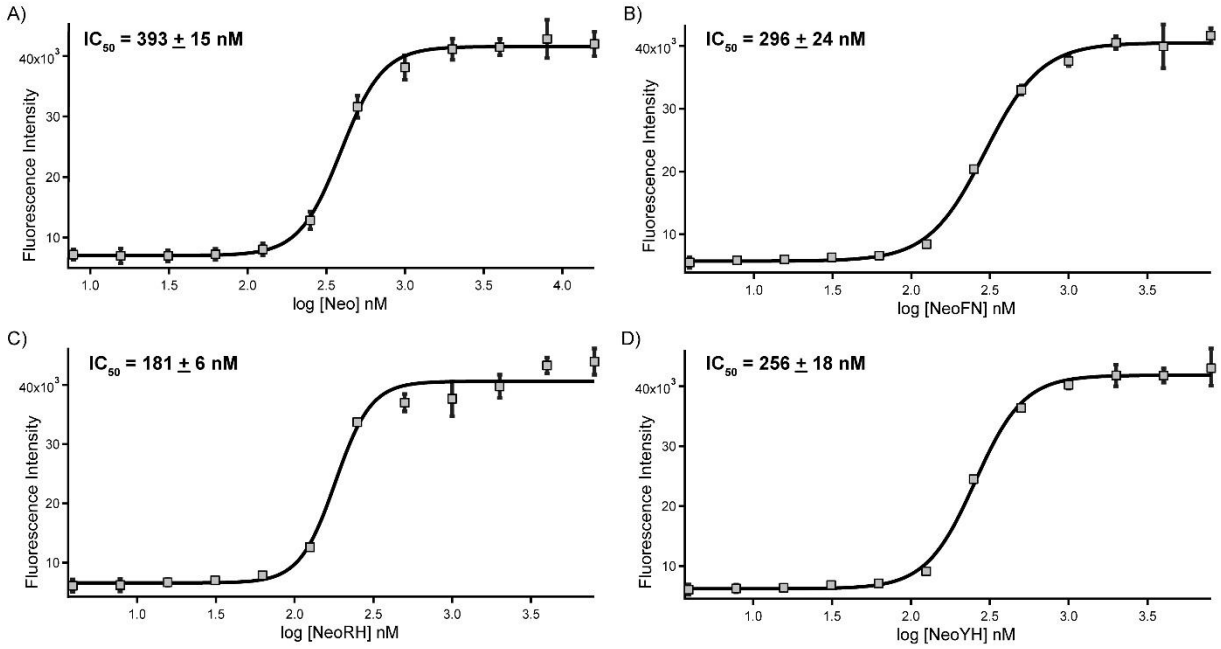


Figure 4.3 IC_{50} measurements of PA-conjugates. Several PA-conjugates displayed higher affinity than Neo towards $ES7_{CA}$ RNA. Titration of ES7 was performed with A) Neo, B) NeoFN, C) NeoRH and D) NeoYH and F-neo displacement was monitored. Measured IC_{50}

values are estimated as: Neo = 393 ± 15 nM, NeoFN = 296 ± 24 nM, NeoRH = 181 ± 6 nM and NeoYH = 256 ± 18 nM.

4.3.5 Determination of minimum inhibitory concentration

Minimum inhibitory concentration (MIC) values, which quantitate inhibition of microbial growth, are consistent with the results of a single-dose assay (Table 4.1). Inhibition of microbial growth by PA compounds was initially assessed through a single-dose assay. The ten strongest binders identified in HTS (above) were tested at 200 µM for inhibition of growth of *C. albicans.* All the PA-conjugates except NeoRD and NeoPN inhibited *C. albicans* growth at these conditions.

MIC studies were performed on PA-conjugates that showed activity in single dose assays. In this work MIC values were defined as the lowest concentration of compound that inhibit microbial growth by 95% after 24 h. NeoFH, NeoRH and NeoYH exhibited MICs of 100 µM, which are the lowest MIC values obtained from our assays. Most compounds from the PA library have MICs of 200 µM; these compounds include NeoFN, NeoLN, NeoCN, NeoCF, and NeoCP. Neo, without any amino acid conjugation, NeoPN and NeoRD have a MICs higher than 200 µM. After 24 h of incubation at 200 µM, none of these compounds inhibit microbial growth. Preliminary experiments performed in our laboratory (unpublished observations) indicated that Neo have a MIC greater than 5mM in *C. albicans.*

Fluconazole is a standard treatment against fungal infections. MIC values for fluconazole were determined at our experimental conditions and at experimental conditions used by others in YPD media with the same strain of *C. albicans* (229). The results obtained

here are in general agreement with those obtained by Vasicek et al. Most PA-conjugates exhibit MIC values that are lower than fluconazole.

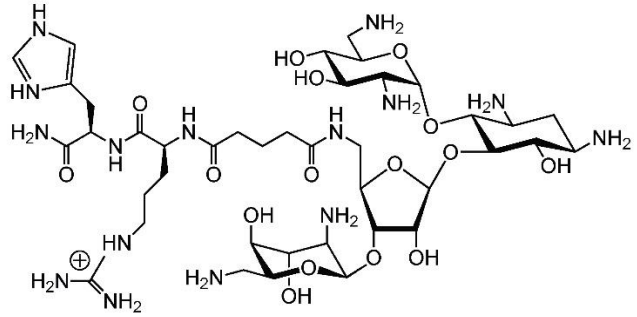
Table 4.1 Minimal inhibitory concentration (MIC) values determined in YPD at 30 °C for 24 h. MIC values were defined as the lowest concentration of compound that inhibit microbial growth by 95%.

Name	**PA No.**	***C. albicans***
Neo	---	>200
NeoPN	1151	>200
NeoFN	1150	200
NeoLN	1149	200
NeoYH	1182	100
NeoRH	1173	100
NeoCF	1212	200
NeoRD	1159	>200
NeoCN	1156	200
NeoCP	1226	200
NeoFH	1178	100
Fluconazole	---	>200

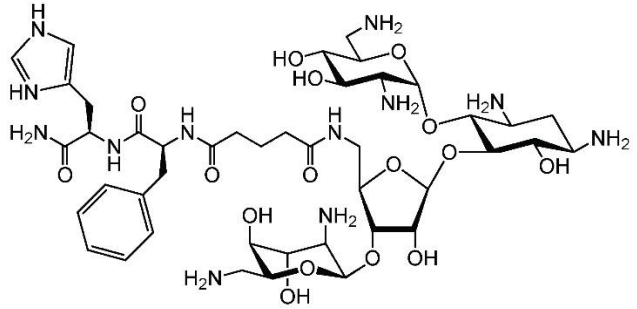
4.3.6 Cytotoxicity assays

Cytotoxicity studies reveal that PA-conjugates are not lethal to human cells (Figure 4.4). Neo, NeoRH, NeoFH and NeoYH at 100 μM were incubated with HEK293T cells for 72 h. These concentrations were set at the MIC values obtained in this work. The results indicate a survival rate of more than 69% for all PA-conjugates and for Neo. Useful antimicrobial agents must have target selectivity and no-toxicity towards the human host.

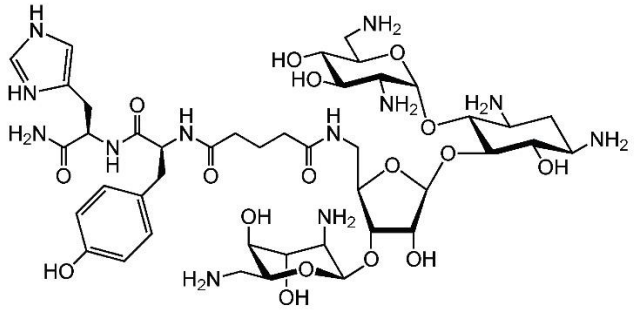
A) NeoRH



B) NeoFH



C) NeoYH



D)

Name	DPA No.	Viability (%)
Neo	---	69
NeoRH	1173	69
NeoFH	1178	82
NeoYH	1182	94

Figure 4.4 Cytotoxicity studies performed with PA-conjugates in human cells. PAs were incubated with HEK293T cells for 72 h at 100 μM. Structures of PAs used in this study, A) NeoRH, B) NeoFH and C) NeoYH are shown. D) Cell viability is reported in comparison to the viability of PA-free cells.

4.3.7 General trends in mono- and di-peptide composition

The HTS results suggest that ten of the twenty best binders in the PA library contain aromatic amino acids. The HTS results suggest that single amino acid conjugates exhibiting higher affinity than Neo for $ES7_{CA}$ (NeoS, NeoC and NeoN) contain amino acids that are polar and uncharged, including serine, cysteine and asparagine (Figure 4.2). An exception to this is NeoβA, which contains a non-polar amino acid. Single amino acid conjugates NeoD and NeoP show the lowest affinity for $ES7_{CA}$, displacing F-neo at 68% under our experimental conditions (Figure C.4). The observed low displacement by NeoD is expected since RNA is negatively charged.

Double amino acid conjugates with positively charged and polar amino acids, including PAs containing asparagine, histidine or cysteine at either the first or second position, exhibit higher affinities for $ES7_{CA}$. NeoPN displays the highest affinity for $ES7_{CA}$ with 140% displacement, followed by NeoFN with 139% displacement, and NeoLN and NeoYH with 132% of displacement. The conjugation of two positive charges to the PA-conjugates does not seem to enhance the binding to $ES7_{CA}$ (Figure C.4). Three of the ten worst binders contain an aspartic acid in the first position. Conjugation to a tryptophan is also detrimental for binding to $ES7_{CA}$; seven of the ten worst binders belong to this category. The worst binding double amino acid conjugates contain negatively charged or bulky amino acids.

IC_{50} results are generally consistent with the HTS results. IC_{50} results obtained with double amino acid conjugates confirm that binding to $ES7_{CA}$ is strengthened when the Neo scaffold is conjugated to positively charged and polar amino acids. IC_{50} results also suggest that conjugation to amino acids with smaller aromatic systems and greater hydrogen

bonding capabilities can improve the affinity towards $ES7_{CA}$. Of all the conjugates tested, NeoRH displays the greatest affinity. This PA contains the smallest aromatic system and the most hydrogen donors/acceptors of the three PAs tested for IC_{50}.

MIC experiments suggest that the second amino acid position is important for PA activity *in vivo*. Compounds conjugated to a histidine in the second amino acid position exhibit the lowest MICs. Overall, the amino acids conjugated at this position follow a well-defined pattern; most of them are polar, with hydrogen bonding capabilities. NeoRD is an exception and exhibited higher values of MIC. Histidine is the smallest aromatic amino acid and possesses two sites available for hydrogen bonding. It seems likely that a combination of geometry and hydrogen bonding capabilities at the second amino acid position offer adequate interactions within the PA-$ES7_{CA}$ complex *in vivo*. No specific contribution could be deduced for the first amino acid position, as no definite trend is observed. Based on the results this position could have less of an impact in MIC determination.

4.3.8 Secondary structure of $ES7_{CA}$

Isolated $ES7_{CA}$ folds to a native-like secondary structure in the presence of Na^+, with four primary helices (ES7a, ES7b and ES7c and Helix 25, Figure 4.5). The secondary structure of isolated $ES7_{CA}$ was determined by SHAPE footprinting (140), computational folding (141) and secondary structure homology modeling. As shown in Figure 4.5A, nucleotides in helical regions are the least reactive in the SHAPE assay.

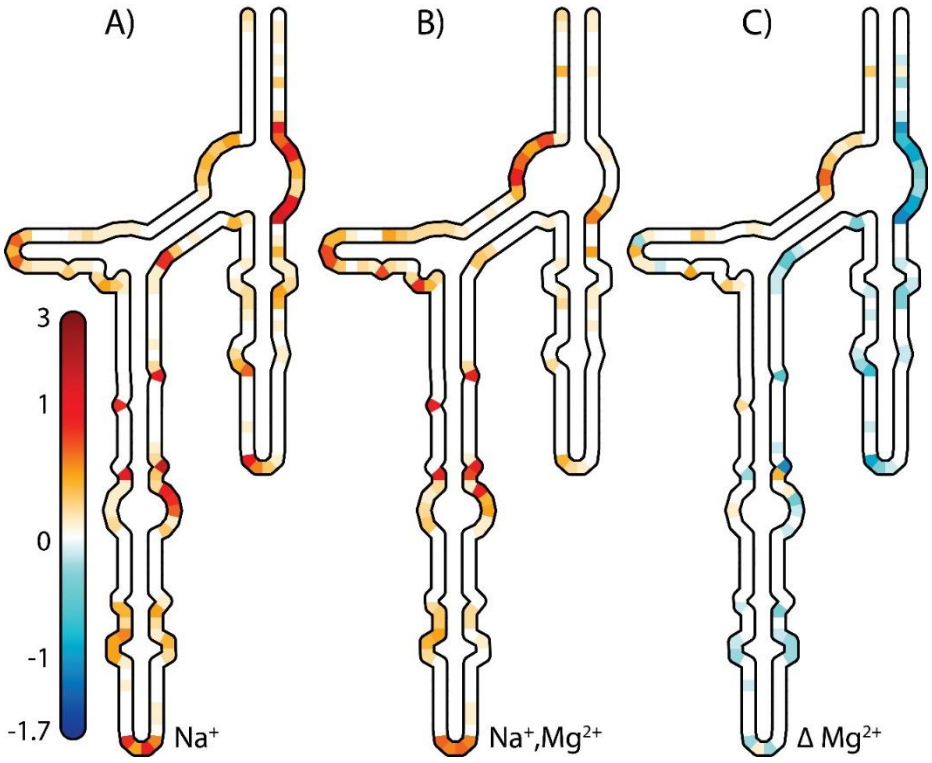


Figure 4.5 SHAPE reactivities of $ES7_{CA}$. A) SHAPE reactivities mapped onto the $ES7_{CA}$ secondary structure confirm the correctness of secondary structural model of $ES7_{CA}$ in the presence of Na^+ alone. Helical regions are unreactive (white) while loops and bulges are reactive (red). B) Addition of Mg^{2+} does not disrupt secondary structure. C) A delta map (Δ), showing changes in reactivity upon addition of Mg^{2+} indicating Mg^+-induced changes in the structure of $ES7_{CA}$. Nucleotides with positive values (red) indicate higher reactivity upon addition of Mg^{2+}. Nucleotides with negative values (blue) indicate decreased reactivity upon addition of Mg^{2+}. All samples contained 200 mM NaOAc, 50 mM Na-HEPES (pH 8.0). Data were mapped and figures generated with the program RiboVision (126).

4.3.9 ΔMg^{2+} heat maps of $ES7_{CA}$

The structure of isolated $ES7_{CA}$ changes upon addition of Mg^{2+}. SHAPE reactivity obtained in the presence of Na^+ alone (Figure 4.5A) was subtracted from the SHAPE reactivity obtained in the presence of both Na^+ and Mg^{2+} (Figure 4.5B) to give a ΔMg^{2+} heat map (Figure 4.5C). The ΔMg^{2+} heat map identifies nucleotides that exhibit Mg^{2+}-dependence of SHAPE reactivity, which can either increase or decrease. Helical regions

are invariant to Mg^{2+} (white on the ΔMg^{2+} heat map) as expected (120, 122), indicating that secondary structure forms in the presence of Na^{+} alone and is maintained upon Mg^{2+} addition. Mg^{2+}-specific changes in SHAPE reactivity are focused in bulges, loops and non-helical regions. It seems probable that these nucleotides are involved in long-range interactions in the native state, which are favored by the addition of Mg^{2+} (120, 122). This work provides the first structural data on any rRNA of *C. albicans.*

Comparison of SHAPE data obtained for $ES7_{CA}$ (Figure 4.4) and $ES7_{SC}$ (161) show that ES7 of these fungi share similar secondary and tertiary structures. Our SHAPE data obtained previously for isolated $ES7_{SC}$ matches the predicted models of the $ES7_{SC}$ secondary structure (161) and the SHAPE reactivity of the $ES7_{SC}$ of the assembled *S. cerevisiae* ribosome (142). ES7 for both organisms consist of four helices and a main junction where Helices ES7a, ES7c and 25 are joined. This junction is the region of greatest changes in SHAPE upon addition of Mg^{2+}. Many of the bulges and helical mismatches show changes in SHAPE upon addition of Mg^{2+}. The loops and mismatches are conserved between $ES7_{CA}$ and $ES7_{SC}$, as are changes in SHAPE reactivity of these regions upon Mg^{2+} addition.

4.3.10 ΔNeo heat maps of $ES7_{CA}$

Neo binds to $ES7_{CA}$ with high stoichiometry, as indicated by binding assays (see above) and is therefore assumed to occupy multiple binding sites on $ES7_{CA}$. SHAPE reactivity obtained in the presence of saturating Neo was subtracted from the SHAPE reactivity obtained in the absence of Neo to give a ΔNeo heat map (Figure 4.6). We have obtained ΔNeo heat maps in the presence and absence of Mg^{2+}. These data, obtained under

saturating Neo, do not allow determination of differential affinities of the multiple Neo binding sites.

ΔNeo heat maps in both the presence and absence of Mg^{2+} (Figure 4.6) show that Neo dependent-changes are broadly distributed throughout $ES7_{CA}$. This distribution is consistent with Neo association at multiple binding sites. Nucleotides located in the main junction, helical loops, single-stranded and mismatch regions exhibit the most intense signals in both ΔNeo heat maps. The locations of Neo-dependent changes in SHAPE reactivity are consistent with greater affinity of Neo for non-canonical secondary structure (230-234). Similarities in the ΔNeo heat maps obtained in the absence and presence of Mg^{2+} shows that binding is not substantially altered by formation of tertiary structure. Mg^{2+} does not appear to compete with Neo and it is not required for Neo binding to ES7 (Figure 4.6A and 4.6B).

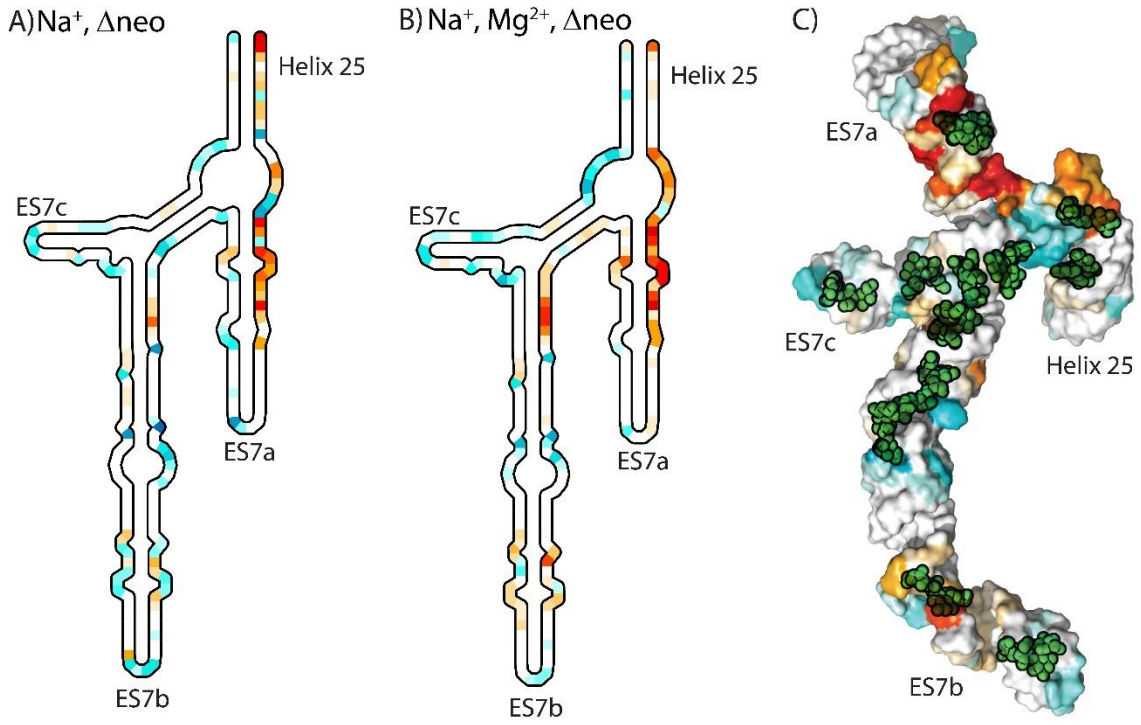


Figure 4.6 Delta maps exhibiting the changes in SHAPE reactivity of $ES7_{CA}$ upon addition of Neo. A) in Na^+ only and B) in Na^+/Mg^{2+}. C) Docking of Neo (green) on a three-dimensional model of $ES7_{CA}$. Data from panel B is mapped onto the three-dimensional structure of $ES7_{CA}$. Nucleotides with positive values (red) indicate higher reactivity in the presence of Neo. Nucleotides with negative values (blue) indicate less reactivity with Neo. The experimental conditions and data analysis are the same as in Figure 4.5.

4.3.11 Thermal folding/unfolding of $ES7_{CA}$

Isolated $ES7_{CA}$ is well-behaved in solution. As temperature changes, $ES7_{CA}$ folds and unfolds reversibly in multi-state processes. The number of melting transitions of $ES7_{CA}$ were estimated, along with the T_m's and thermodynamic parameters, from non-linear fitting of observed melting profiles to predictions of a multi-state model with independent transitions (Figure 4.7) (143). Thermodynamic parameters were estimated previously for $ES7_{SC}$ (161) and $ES7_{HS}$ (Chapter 3). Melting profiles are plots of the derivative of the fraction unfolded ($d\Theta/dT$) versus temperature.

The unfolding of $ES7_{CA}$ is similar to that of $ES7_{SC}$ (161). The best fits for both these fungal ESs were obtained with four transition models. For $ES7_{CA}$ in 180 mM NaCl, the melting temperatures of the four unfolding transitions ($T_{m,n}$) are $T_{m,1}$ = 52 °C, $T_{m,2}$ = 59 °C,

$T_{m,3}$ = 68 °C and $T_{m,4}$ = 78 °C. Under the same conditions, $T_{m,1}$ is the same for $ES7_{CA}$ and $ES7_{SC}$. However, $T_{m,2}$ $T_{m,3}$, and $T_{m,4}$ for $ES7_{CA}$ are around ~ 5 °C less than the corresponding T_m's of $ES7_{SC}$. The results suggest that $ES7_{CA}$ is less stable than $ES7_{SC}$. As observed for $ES7_{SC}$, the T_m's of $ES7_{CA}$ are constant with varying RNA concentrations indicating that the transitions are unimolecular.

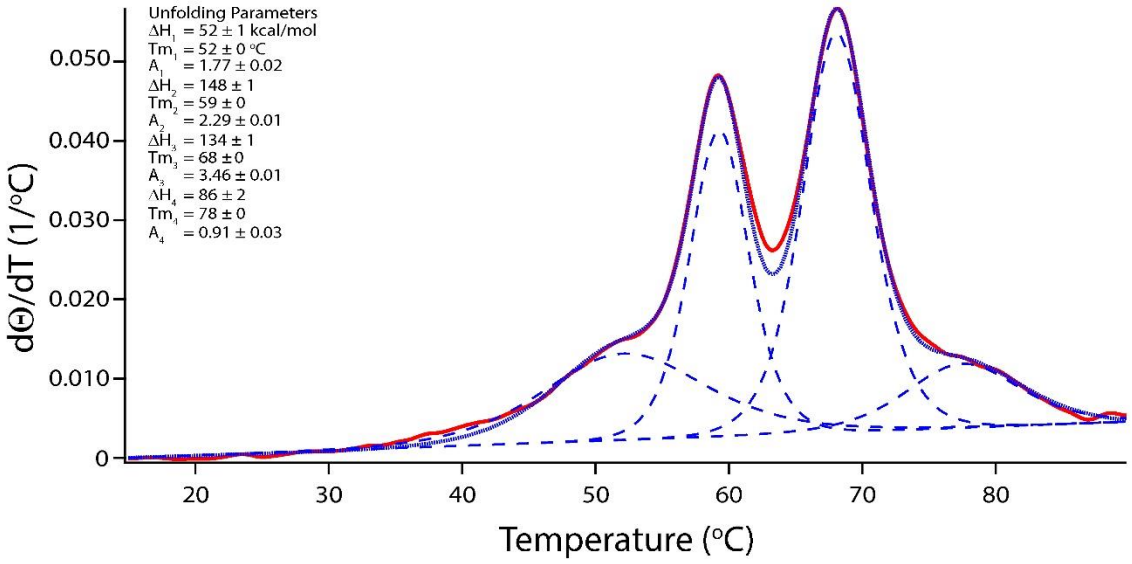


Figure 4.7 Melting profile of $ES7_{CA}$ rRNA. The derivative of the fraction unfolded with respect to temperature was computed at 260 nm. The profile was fit to a non-sequential, independent transition model, with four transitions with the program Igor Pro. The enthalpy (ΔH_n), melting temperature ($T_{m,n}$) and the relative hyperchromicity (A_n) were estimated for each transition, n. The observed profile is red, the fit is dotted blue and the deconvolution of the fit is dashed blue.

Neo stabilizes $ES7_{CA}$. Melting profiles for $ES7_{CA}$ were obtained at five molar ratios of Neo to $ES7_{CA}$ (0 µM Neo, 2.78 µM, 27.8 µM, 139 µM, and 278 µM, at constant 278 nM $ES7_{CA}$). The overall stability of $ES7_{CA}$ increases with increasing concentration of Neo as indicated by monotonic changes in all four $T_{m,n}$'s (Figure 4.8). The results are consistent with a model (above) in which Neo binds to $ES7_{CA}$ at broadly distributed sites. It appears that Neo binds to both native and partially unfolded intermediates. Additional information on fitting procedures and parameter estimations are provided in Appendixes A and C.

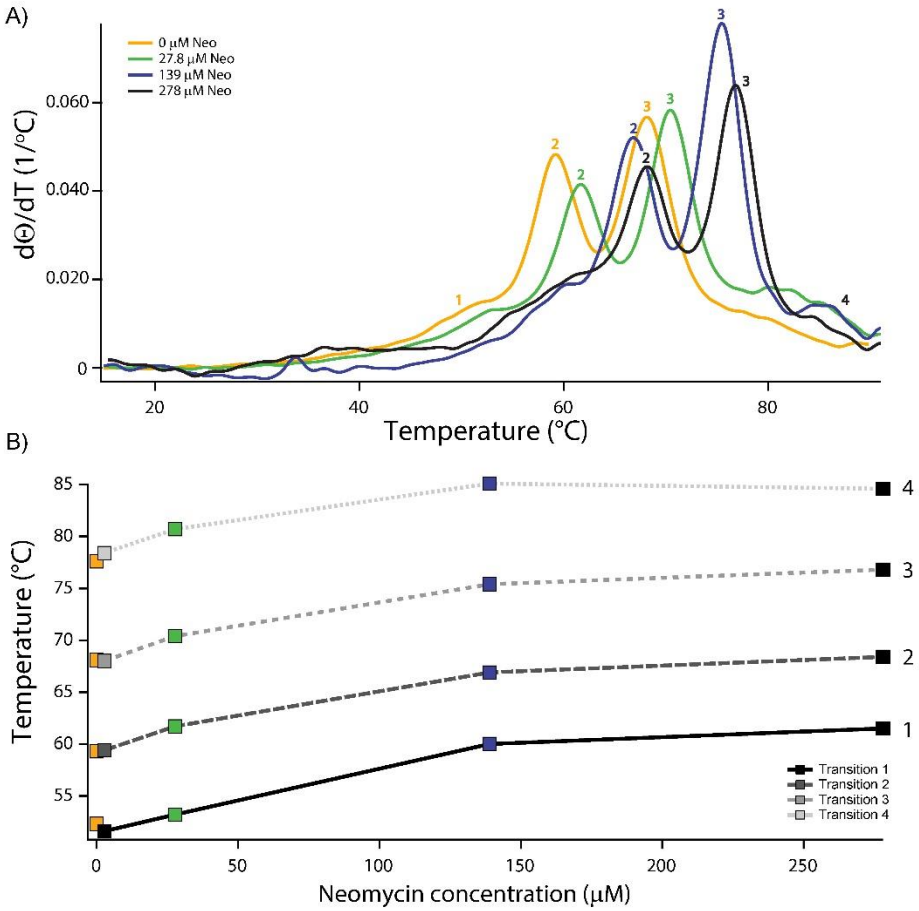


Figure 4.8 Melting profiles of $ES7_{CA}$ in the presence of Neo. A) The derivative of the fraction unfolded with respect to temperature at 260nm reveals four transitions. Each transition is indicated with a number at the top of the melting profiles. Increased Neo causes the $T_{m,n}$ to increase. B) Changes in melting temperatures ($T_{m,n}$) for $ES7_{CA}$ at increased concentrations of Neo for each transition. Each transition is indicated with a number.

4.4 Discussion

Ribosomes are primary targets for antibacterial agents (235) but thus far, have not proven to be useful targets for drugs against pathogenic fungi. Fungal ribosomes are poor targets in part because of acute similarities in sequence and structure between fungal and human ribosomes in the A site, P site, exit tunnel and peptidyl transfer center regions. Instead, chemotherapeutics target fungal cell walls, membranes, and nucleotide biosynthetic pathways (90, 236). The diversity of RNA function, which has been only

recently understood, suggests that small molecule binding to RNA may provide new avenues to useful therapeutics (237). Targets have emerged including ribozymes, riboswitches (238), microRNAs (239) and RNA promoters (240).

Human and fungal ESs are highly divergent, and therefore fungal ESs are inviting drug targets. Our goal is to determine if ES rRNAs can be developed as useful targets for chemotherapeutics directed at eukaryotic pathogens. Fungal translation can be targeted with specificity as evidenced by the activity of Sordarins, which inhibit ribosomal translation in fungi by interacting with elongation factor 2 (EF-2) (103).

Aminoglycosides such as Neo are useful lead compounds for a discovery process for targeting fungal ribosomes. Aminoglycosides bind to rRNA and are active against gram-positive and gram-negative bacteria (241). In bacteria, Neo associates with the A site decoding region of 16S rRNA, inducing changes in mRNA-tRNA conformation leading to amino acid misincorporation in the nascent protein (242-243). Aminoglycoside-induced mistranslation is thought to promote protein misfolding and cell death (244-245).

Here we show that Neo binds tightly and selectively to $ES7_{CA}$ *in vitro*. The K_d of F-neo for $ES7_{CA}$ is in the low nanomolar range, similar to the K_d of Neo for the bacterial A site (246). Neo appears to be specific for *C. albicans* ribosomes over human ribosomes. F-neo binds to $ES7_{CA}$ with 4000x higher affinity than to $ES7_{HS}$. $ES7_{CA}$ and $ES7_{HS}$ share low similarity in sequence, size and structure. $ES7_{HS}$ is four-times larger than $ES7_{CA}$ with 30% greater GC content.

High throughput screening of a 215 member PA library (Figure 4.1, Figure C.4) reveals sixty three compounds that bind with greater affinity than Neo to $ES7_{CA}$. This library was previously screened for antibacterial activity in gram-negative and gram-

positive bacteria and for binding to *E. coli* A site rRNA (222). PA compounds in which polar and positively charged amino acids are conjugated at the second position (Figure 4.2, Figure C.4) exhibit the greatest affinity for $ES7_{CA}$. NeoFH and NeoYH have lower IC_{50} values than Neo and NeoFN. MIC values were determined for the ten best $ES7_{CA}$ binders from the HTS using *C. albicans*. NeoFH, NeoRH and NeoYH exhibited the lowest MIC values (100 μM). Our results suggest that the geometry and hydrogen bonding capabilities of the second amino acid position are important for binding and antifungal activities. PA compounds with histidine at the second position exhibit the lowest IC_{50} values and MIC values.

Thermal melting, SHAPE and binding experiments suggest that $ES7_{CA}$ has multiple binding sites for Neo. Saturation plots suggest that there are ~13 binding sites for Neo in ES7. Neo is known to interact with various rRNA structures including hairpins (230-231), internal bulges (231), single-stranded RNA (232), telomeric DNA (247), DNA triple helices (234, 248-250), RNA triple helices (233-234) and DNA-RNA hybrid helices (233). Neo-RNA recognition is generally mediated by structure and not by sequence. SHAPE experiments suggest that Neo binds predominantly to internal loops, hairpins, single-strand regions and mismatches of $ES7_{CA}$.

The results here are consistent with a model in which $ES7_{CA}$ is a target for certain PA-conjugates. This model is supported with tight binding to $ES7_{CA}$ *in vitro* and with the MIC results. Binding of PA-compounds to $ES7_{CA}$ *in vivo* might disrupt core ribosomal function or could disturb interactions with non-ribosomal proteins that are essential for cell viability (161). In sum, ESs appear to be reasonable targets for future drug target studies. ESs may emerge as useful new targets for anti-microbials.

It seems unlikely that Neo or PA-compounds bind to the *C. albicans* A site, which is very similar in sequence and structure to the human A site (Figure C.6). It has been shown by Arya that Neo binds with very low affinity to the human A site (222). Our cytotoxicity studies here suggest that PA-conjugates are non-toxic to human cells.

4.5 Conclusions

Isolated $ES7_{CA}$ is stable, folds autonomously and adopts a near-native state. PA-conjugates bind with high affinity and selectivity to $ES7_{CA}$. Moreover, PA-conjugates with high affinity for $ES7_{CA}$ induce mortality on *C. albicans* but not in HEK293T cells. Additional experiments are required to confirm that the mechanism of mortality in *C. albicans* is related to binding of the lead compounds to ES7.

CHAPTER 5

RECOMMENDATIONS AND CONCLUSIONS

5.1 Recommendations

5.1.1 *In vivo* identification of ES-associated proteins

RNA-protein pull-down experiments provided evidence for the interaction of ESs with non-ribosomal proteins/protein assemblies. As stated in Chapter 2 and Chapter 3 our experimental procedure identified primary and secondary proteins. The identification of multiple complexes essential to cell viability (i.e. ubiquitin-proteasome system and eIF-complexes) was facilitated by the experimental setup used in our work. However, in order to better understand the functions of ESs within the cell and its associations with proteins, *in vivo* experiments will be required.

Recent development of fluorescent RNA aptamers like Spinach (251) and Broccoli (252) could help to determine the fate of RNAs *in vivo*. Gene constructs that include the covalent attachment of fluorescent aptamers to ESs will provide real-time monitoring of the fate of ESs inside the cells through confocal microscopy. Moreover, previously identified ES-associated proteins (i.e. identified *via* pull-down experiments) could also be fused with fluorescent proteins, like m-cherry and m-strawberry (253), which will open up the possibility of simultaneous RNA-protein visualization upon transfection of constructs to human cells. The RNA will be observed as green, the proteins in red and the region of co-localization will be observed as yellow. Monitoring of ES-protein interactions *in vivo* will help to elucidate the nature of ES-proteins interactions within the cell (steady vs. transient).

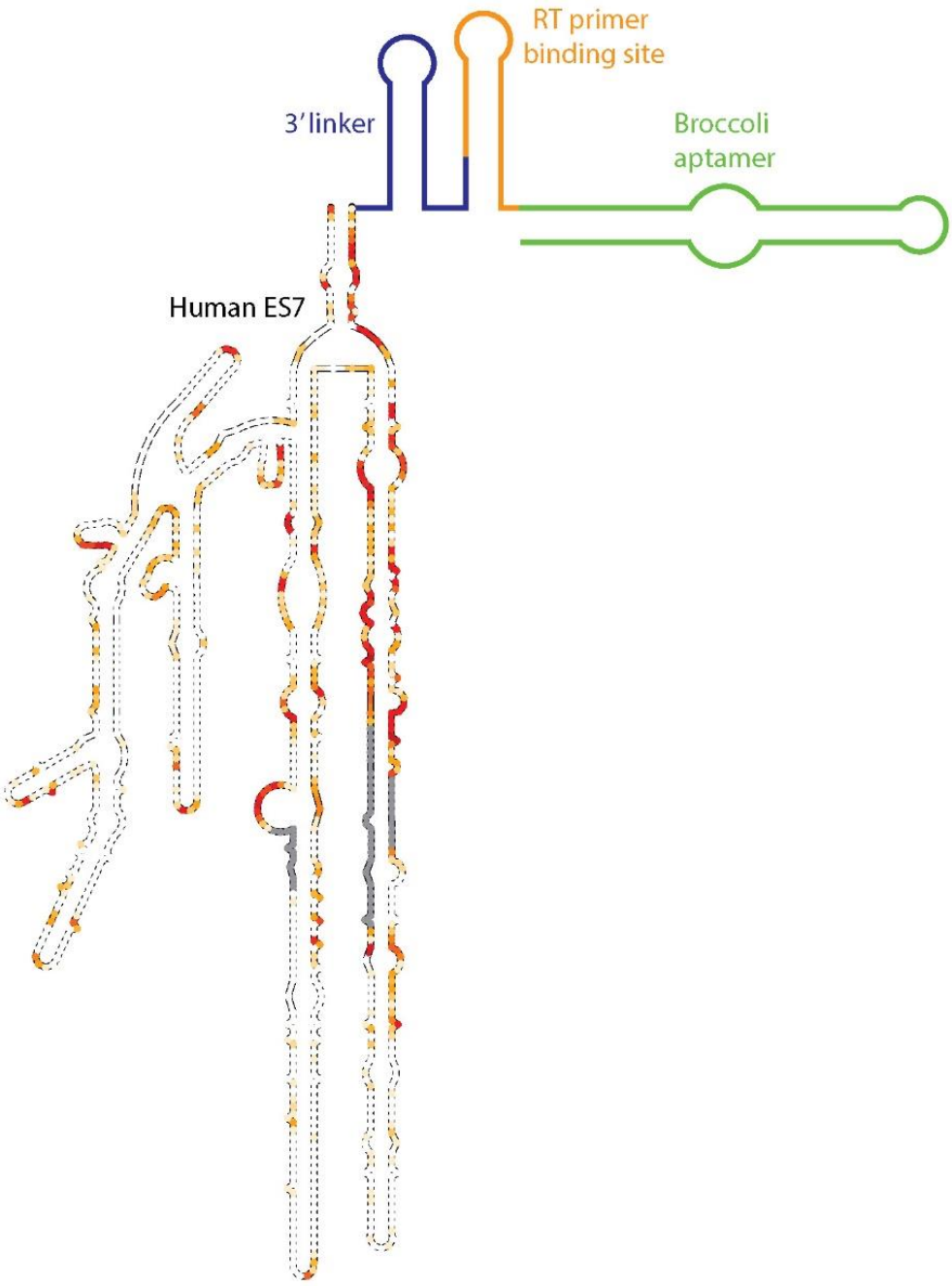


Figure 5.1 Schematic diagram of human ES7 attached to the SHAPE primer-linker region and the broccoli aptamer. Human ES7 secondary structure is shown. The SHAPE primer-linker region taken from Weeks (124) consist of two regions: the RT primer binding site (orange) and the 3' linker (dark blue). The broccoli aptamer is shown in green and should be inserted at locations where it does not disturb the RNA structure (i.e. 5' or 3' terminus). Aptamer incorporation at the Weeks region ensure that secondary- and three-dimensional structures determined by SHAPE are maintained upon transfection.

5.1.2 Post-transcriptional modifications of ESs

rRNAs are subjected to many post-transcriptional modifications, including methylation and pseudouridylation (254-255). Post-transcriptional modifications in *E.coli* (256) and *Thermus thermophilus* (257) are clustered in regions of functional importance in the ribosome, with high levels of nucleotide conservation and lacking contacts with rProteins. Over 100 modifications have been found in vertebrate 28S cytoplasmic rRNA (258), mainly in the common core region.

ESs rRNAs used in this work were transcribed *in vitro* and therefore lack naturally-occurring post-transcriptional modifications. Although several studies (reviewed by Maden et al.) do not support the existence of post-transcriptional modifications in ESs, it will be important to confirm their presence or absence in ES rRNAs with current experimental techniques. Existence of post-transcriptional modifications can alter the RNA-protein interactions identified in pull-down experiments and it will be important to investigate alternative functions of ESs that could be attributed to these modifications, if any.

5.1.3 Confirmation of G-quadruplexes in human rRNAs

G-quadruplexes are located on telomeric DNAs (259-260), DNA promoter regions (261-262) and mRNAs (193, 263). Numerous putative G-quadruplex sequences have been identified in the human genome (264-265), but none of them correspond to sequences of ribosomal genes. Chapter 3 describes multiple, putative G-quadruplex sequences in human rRNA, which are clustered in ES7 and ES27. $Gq1_{ES7}$ and $Gq2_{ES7}$ sequences, from ES7, were studied *via* CD experiments and the results suggest G-quadruplex formation in human

rRNA. Further research in this field is essential to confirm the existence of G-quadruplexes in human rDNA or rRNA.

Balasubramanian et al. visualized DNA and RNA G-quadruplexes in the cytoplasm of human cells using the BG4 antibody (266-267). Identification of BG4 as a selective G-quadruplex binder expands the repertoire of experiments that can be performed for G-quadruplex identification. Binding curves to measure the affinity of BG4 for G-quadruplexes can be acquired with ELISA for all putative G-quadruplex sequences and their respective controls. EMSA experiments performed with ESs and BG4 along with CD experiments will also assess the formation of *in vitro* G-quadruplexes.

In vivo confirmation of G-quadruplexes in human rRNA is necessary to assess the validity of *in vitro* assays. Co-transfection of human cells with a plasmid encoding human ES7/ES27 rRNA and BG4 will allow visualization of human rRNA G-quadruplexes in cells, if formed, through confocal microscopy. Recently, RNA foot printing techniques have also been used to detect and map RNA G-quadruplexes (268-269). Both techniques aim to identify G-quadruplexes in functional conditions and represent alternative methods for *in vivo*/functional G-quadruplex identification.

Preliminary NMR experiments are being performed by our collaborator Dr. Janez Plavec on ES7 putative G-quadruplex sequences and its respective controls to confirm or discard G-quadruplex formation. Existence of G-quadruplexes in human ribosomes will be of high impact to the RNA field as the specific roles of these structures are not completely elucidated, and they have not been identified in rRNA.

5.1.4 Validation of expansion segments as antimicrobial targets

Successful discovery of antimicrobials requires both, good targets and good libraries of compounds (270). Good targets are essential to the pathogenic organism, possess remarkable selectivity, prevent or stop pathogenic growth, and insofar as possible promote the development of broad-spectrum antimicrobials. ES7 of pathogenic *C. albicans*, *C. neoformans*, *C. immitis* and *P. jiroveci*, share sequence and structural similarities among themselves but differ significantly, in sequence and structure, from ES7 of human (Figure 2.5, Figure 3.4, and Figure 5.2).

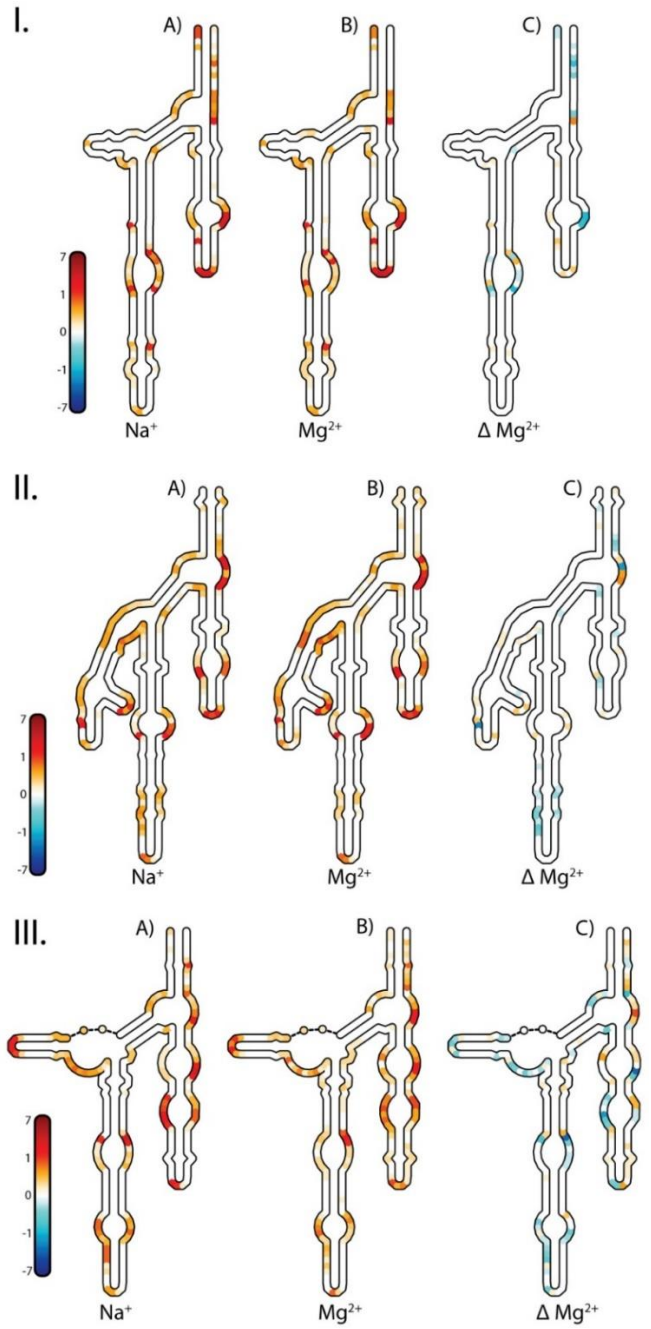


Figure 5.2 SHAPE reactivities of *C. immitis*, *C. neoformans* and *P. jiroveci*. SHAPE reactivities are mapped onto I. *C. immitis*, II. *C. neoformans* and III. *P. jiroveci* ES7 rRNA secondary structures. SHAPE reactivities were determined in presence of A) Na^+ or B) Na^+ and Mg^{2+}. Orange or red nucleotides show moderate to high SHAPE reactivity. Ivory or white nucleotides show low or no SHAPE reactivity. C) Mg^{2+} induces changes in SHAPE reactivity. Nucleotides in red show increased reactivity in the presence of Mg^{2+}. Nucleotides in blue show decreased reactivity. These figures were generated with RiboVision. All samples contained 200 mM NaOAc, 50 mM Na-HEPES (pH 8.0).

Chapter 4 discusses the HTS of a peptidic aminosugar library of compounds for its affinity to $ES7_{CA}$. Although binding data correlates with MIC calculations, the relationship between binding to ES7 and pathogen mortality has yet to be understood. Neomycin targets the A site in the SSU of bacteria with good selectivity (242). However, the similarity between A sites in eukaryotic species is remarkable (Figure C.6) and it does not represent a good target for antifungals. There is no difference between the A sites of *C. albicans* and human ribosomes. It therefore seems unlikely to consider that binding of PA-conjugates to the A site of *C. albicans* will induce mortality on these organisms, but not in human cells. Our results suggest that binding of PA-conjugates to ES7 interferes with vital functions of the cell, but more studies will have to be performed to fully understand how it does so.

Chapter 4 of this thesis described work performed using only *C. albicans* and ES7 as test models. It will valuable to investigate if the binding results obtained with $ES7_{CA}$ extend to other ES7 of pathogenic fungi including, *C. neoformans*, *C. immitis* and *P. jiroveci* and if it will produce similar mortality rates. Further research into the binding and selectivity of small molecules to other ESs of pathogenic organisms will provide information as to whether binding of small molecules to ESs will cause mortality of fungal pathogens.

5.2 Conclusions

This work investigates the role of various ESs within the ribosome, and therefore, its importance for the appropriate functioning of cells. Among the variety of ESs present in rRNA, ES7 was chosen as a model of study since it is the largest ES on the LSU, and the largest ES on the assembled ribosome. As shown throughout this work, ESs maintain their structural conformation in isolation (i.e. in the absence of the complete ribosomal

machinery). Thermal denaturation experiments demonstrated that ESs are stable, autonomously folded, and contain multiple sub-domains.

RNA-protein pull-down experiments demonstrated that ES7 of *S. cerevisiae* bind multiple non-ribosomal proteins, including class I and class II aaRSs, nucleolar, and metabolic proteins. Pull-down experiments performed with human ES7 and ES27 indicate that multiple non-ribosomal proteins also bind to human ESs. Proteins related with the UPS seem to be in association with human ESs. Combined, these results suggest that ESs could play important roles in protein synthesis and degradation.

Experiments with ES7 of *C. albicans* suggest that ESs could have potential as targets for antimicrobials. Neomycin bind tightly to $ES7_{CA}$, but showed low affinity towards $ES7_{HS}$. High-throughput screening of PA-conjugates identified several lead molecules. Lead molecules showed improved binding for $ES7_{CA}$ and lower MIC values, in comparison to Neo. These molecules can be further developed to induce higher mortality rates in pathogenic microorganisms.

Future steps in the study of ESs should focus on: i) conducting *in vivo* experiments to investigate the nature of the interactions of ESs with cellular proteins (steady vs. transient), and how these interactions vary at different stages of the cell cycle, ii) understanding the functions of ESs in different organisms and among themselves, and iii) studying the correlation (i.e mechanism of action) of small molecule-ES interactions on pathogenic mortality. The work presented here provided evidence that ESs are an important component of the ribosome assembly, and set the path for more investigations in this field. Due to the increasing rate of resistance of fungal pathogens to commonly used

antimicrobials and the potential role of these rRNAs in therapeutic treatments further resources should be allocated toward the investigation of ESs.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 MATERIALS AND METHODS

A.1.1 PCR reactions

Primers were dissolved in nuclease-free water to working concentrations of 10 µM or 20 µM (strand). Final PCR conditions consisted of 0.2 mM of each dNTP (New England BioLabs), 1x Complete Long PCR Buffer (Ambion), 1 µM each of forward and reverse primers, 3U of *Pfu* DNA polymerase and at least 50 ng of DNA template. Reaction mixtures were heated to 95 °C for 5 min during an initial denaturation step and cycled 30 times through denaturation (98 °C for 30 sec), annealing (52.0 °C for 1 min), and elongation (72 °C for 2 min/kb). The final elongation step was performed at 72 °C for 10 min. The target amplification products were purified by preparative gel electrophoresis with standard protocols.

Table A.1 Primers used to isolate *S. cerevisiae* ES7 from the yeast genome and to insert flanking restriction sites.

Gene	Primer	Sequence (5' to 3')[a]
S. cerevisiae ES7	Forward	GGCATTTGATCAGACATGGTGTTTT
	Reverse	GGCATATAACCATTATGCCAGCATC
	EcoRIT7_FWD	GTAAGAATTC*TAATACGACTCACTATAGGG*CATTTGATCAGAC
	HindIII_REV	CACCAAGCTTGGCATATAACCATTATGCCAG

a) Restriction sites are underlined and the T7 promoter sequence is in italics.

Table A.2 Primers used to generate the SHAPE structural cassette.

Gene	Primer	Sequence (5' to 3')
S. cerevisiae ES7 Weeks	Forward	AATCGGGCTTCGGTCCGGTTCAAGCTTGGCGTAATCATG
	Reverse	TGGATCCGGCGAACCGGATCGAGGCATATAACCATTATGCC

Table A.3 Primer used in SHAPE RT reactions and the DNA anchor used in RNA-protein pull-down assays.

Gene	Primer	Sequence (5' to 3')
S. cerevisiae ES7	5' [6-FAM] or Biotin label	GAACCGGACCGAAGCCCG

Table A.4 Primers used to insert aminoacyl tRNA-synthetases in pET21C.

Gene	Primer	Sequence (5' to 3')[b]
AspRS	Forward	CGGCAGCCATATGATGTCTCAAGACGAAAATATTGTCAAAG
	Reverse	GTGGTGCTCGAGTGGTCTTAATCTCTTTGGATCTCTTGG
LysRS	Forward	CGGCAGCCATATGATGTCTCAACAAGATAATGTCAAAGCC
	Reverse	GTGGTGCTCGAGATTTTCTTCTTCCTTTTTGACTTCCTCT
TyrRS	Forward	CGGCAGCCATATGATGTCCTCTGCTGCCACGGT
	Reverse	GTGGTGCTCGAGCAATTTGGTTTCCTCTAGTTTCGTG

b) Restriction sites are underlined.

A.1.2 Pull-down assays and LC-MS/MS analysis

Samples were loaded into a WPS-3000TPL RS autosampler (Thermostatted Pulled Loop Rapid Separation Nano/Capillary Autosampler, Dionex) and separated by reversed-phase chromatography on a C18-packed microcapillary column (Magic C18AQ, 3 µm, 100 µm x 16 cm) using an UltiMate 3000 binary pump with a 110 min gradient of 2-25% acetonitrile containing 0.125% formic acid. Samples were detected in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific) using a data-dependent Top 20 method. Cycles included one full MS scan with a resolution of 60,000 in the Orbitrap at the Automatic Gain Control target of 1 million and up to 20 MS/MS of the most intense ions in the LTQ. Selected ions were omitted from further sequencing for 90 sec. Ions with a single or unassigned charge were excluded from fragmentation and maximum ion accumulation times were 1000 msec for each full MS scan and 50 msec for MS/MS scans.

Raw MS files were converted into mzXML format. Precursors for MS/MS fragmentation were checked to minimize incorrect monoisotopic peak assignments. All MS/MS spectra were searched with the SEQUEST algorithm v.28 against the Yeast ORF database (downloaded from http://www.yeastgenome.org/) which include sequences of all proteins in this category and common contaminants (271). To estimate the false discovery rate (FDR) of peptide identification, all protein sequences were listed in forward and reverse orientation (272-273). The following parameters were used in the database search: 10 ppm precursor mass tolerance, 1.0 Da product ion mass tolerance, and Lys-C enzymatic digestion. The parameter file included one differential modification, the oxidation of methionine (+15.9949), and one fixed modification, carbamidomethylation of cysteine (+57.0215).

To evaluate and further control FDRs of peptide identification, the target-decoy method was applied (272-273). Linear discriminant analysis was utilized to distinguish correct and incorrect peptide identifications based on numerous parameters, including XCorr, ΔC_n, and precursor mass error (274-275). Separate linear discriminant models were trained for each raw file using forward and reversed peptide identifications to provide positive and negative training data, which is similar to other approaches in literature (276-277). After scoring, peptides less than six amino acids in length were discarded and peptide spectral matches were filtered to a less than 1% FDR at the peptide level based on the number of decoy sequences in the final dataset.

A.1.3 Thermodynamic model used in UV melting studies

Thermodynamic parameters were estimated from reversible, concentration-independent melting curves. The data was fitted to multiple linear, independent, two-state models to account for the presence of multiple transitions in RNA folding/unfolding (127, 278). The model was first derived for a single, two-state transition and adjusted to account for multiple transitions.

$$N \rightleftharpoons U$$

$$-RTlnK = \Delta H - T\Delta S = \Delta G$$

$$K = \frac{\theta_U}{1-\theta_U}$$

$$ln\left[\frac{\theta_U}{1-\theta_U}\right] = \Delta G = \frac{-\Delta H + T\Delta S}{RT}$$

$$\frac{\theta_U}{1-\theta_U} = exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)$$

$$\theta_U = exp\left(\frac{-\Delta H + T\Delta S}{RT}\right) - \theta_U\left[\exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)\right]$$

$$\theta_U = E - \theta_U E$$

$$\theta_U + \theta_U E = \theta_U(1+E) = E$$

$$\theta_U = \frac{E}{1+E}$$

$$\theta_U = \frac{exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}{1 + exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}$$

$$Multiple\ States$$

$$\theta_U = \sum A_n \left[\frac{exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}{1 + exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}\right]$$

$$\theta_U = \sum A_n \left[\frac{exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}{1 + exp\left(\frac{-\Delta H + T\Delta S}{RT}\right)}\right]$$

$$\frac{d\theta_U(T)}{dT} = \sum A_n \frac{\frac{\Delta H(n)}{2RT^2}}{1 + cosh\left(\frac{\Delta H(n)(T_m(n) - T)}{RTT_m(n)}\right)}$$

Legend:
N – Native RNA
U – Unfolded RNA

T – Temperature
ΔG – Gibbs free energy of melting
ΔS – Entropy of Melting
K – Equilibrium constant
R – Universal gas constant
Θ_u – Fraction of unfolded RNA
n – Transition number
ΔH(n) – Enthalpy of transition n
T_m(n) – Melting temperature of transition n
A_n – Hyperchromicity of transition n
E – $exp\left(\frac{-\Delta H+T\Delta S}{RT}\right)$

A.1.4 Melting parameter estimation and fitting

Data processing and fitting was performed with IgorPro v.6.37. The normalized absorbance vs. temperature data was subjected to central differentiation and binomial smoothing prior fitting of the calculated derivative. For each transition three parameters were varied to fit the calculated derivative to the observed derivative: the enthalpy, the hyperchromicity and the T_m of the transition (n).

A.2 RESULTS

A.2.1 Definitions of expansion segments on *S. cerevisiae* LSU

ESs drawn onto the secondary and tertiary structure of the *S. cerevisiae* LSU particle (Figure 2.1) followed the definitions outlined by Petrov et al. (4) which were obtained by comparing pre- and post-expanded rRNA structures of *E. coli* and *S. cerevisiae* (http://apollo.chemistry.gatech.edu/RibosomeGallery/). ESs defined here are ES3, ES4, ES5, ES7, ES9, ES12, ES19, ES20, ES26, ES27, ES30, ES31, ES39, and ES41.

A.2.2 Protein Binding to *S. cerevisiae* ES7

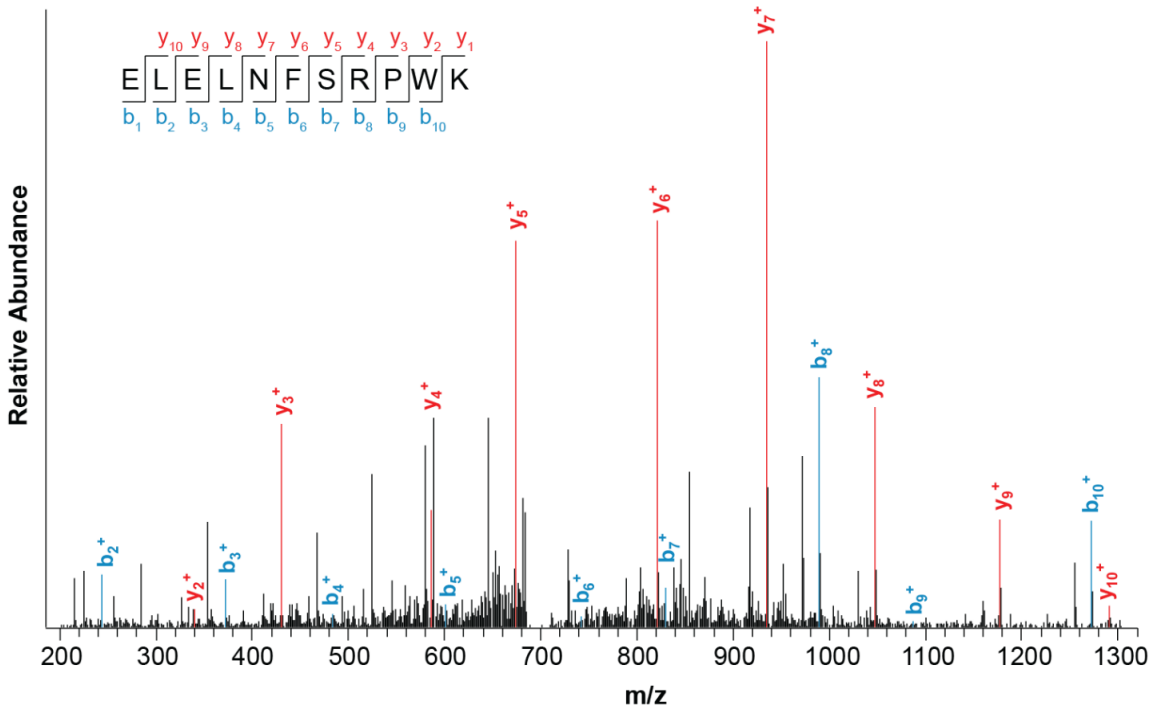


Figure A.1 Tandem mass spectrum of peptide ELELNFSRPWK from LysRS. The peptide was identified with an XCorr of 3.0 and a mass accuracy of -0.69 ppm.

Table A.5 Identified peptides corresponding to amino acyl tRNA-synthetases.

Gene name	Annotation	Unique peptides identified
ALA1	Alanyl-tRNA ligase, cytoplasmic	4
YDR341C	Arginyl-tRNA ligase, cytoplasmic	2
ASN1	Asparagine-tRNA ligase, cytoplasmic	3
DPS1	Aspartyl-tRNA ligase, cytoplasmic	2
CDC60	Leucyl-tRNA ligase, cytoplasmic	2
KRS1	Lysyl-tRNA ligase, cytoplasmic	2
TYS1	Tyrosyl-tRNA ligase, cytoplasmic	3

Table A.6 Proteins identified in the RNA-protein pull-down assay performed with *S. cerevisiae* cell lysates and ES7.

Reference	Annotation
YER164W	Chromo domain-containing protein 1
YBR142W	ATP-dependent RNA helicase MAK5
YDR060W	Ribosome biogenesis protein MAK21
YDR150W	Nuclear migration protein NUM1
YFR053C	Hexokinase-1
YJL053W	Vacuolar protein sorting-associated protein 26

YKL080W	V-type proton ATPase subunit C
YMR049C	Ribosome biogenesis protein ERB1
YOL145C	RNA polymerase-associated protein CTR9
YAL060W	(R,R)-butanediol dehydrogenase
YBL050W	Alpha-soluble NSF attachment protein
YBR299W	Alpha-glucosidase MAL32; Alpha-glucosidase MAL62
YDR002W	Ran-specific GTPase-activating protein 1
YDR129C	Fimbrin
YDR164C	Protein transport protein SEC1
YGL037C	Nicotinamidase
YGR116W	Transcription elongation factor SPT6
YGR130C	Uncharacterized protein YGR130C
YGR175C	Squalene monooxygenase
YGR185C	Tyrosyl-tRNA synthetase, cytoplasmic
YGR234W	Flavohemoprotein
YGR266W	Uncharacterized protein YGR266W
YJL128C	MAP kinase PBS2
YJR024C	Putative aldolase class 2 protein YJR024C
YJR137C	Sulfite reductase [NADPH] subunit beta
YKR067W	Glycerol-3-phosphate O-acyltransferase 2
YLL018C	Aspartyl-tRNA synthetase, cytoplasmic
YLL034C	Ribosome biogenesis ATPase RIX7
YLR287C-A	40S ribosomal protein S30
YLR327C	Translation machinery-associated protein 10
YMR114C	Uncharacterized protein YMR114C
YNL015W	Protease B inhibitors 2 and 1
YNL248C	DNA-directed RNA polymerase I subunit RPA49
YNR034W-A	Uncharacterized protein YNR034W-A
YOL010W	RNA 3'-terminal phosphate cyclase-like protein
YOR014W	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform

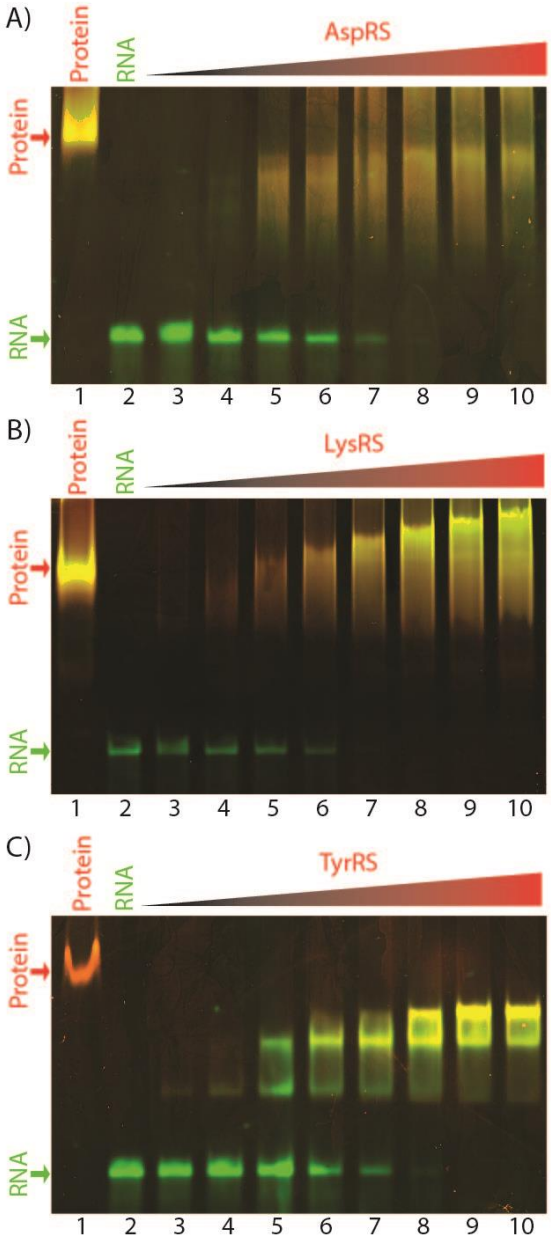


Figure A.2 Gel shift analysis of *T. thermophilus* Domain III^{Fragment} rRNA with aaRSs. Interaction of rRNA with A) AspRS, B) LysRS and C) TyrRS is shown *in vitro*. RNA, protein and RNA-protein complexes were visualized by two color EMSA in a 5% native-PAGE gel containing 3% glycerol. Binding reactions were performed with 1 μM RNA and various protein concentrations at room temperature for 20 min in 20 mM Tris-HEPES, pH 8.0. Lane 1 contains 10 μM of protein and no RNA. Lane 2 consists of 1 μM RNA only. Lanes 3-10 contain 1 μM RNA and increasing protein concentration: lane 3 – 0.25 μM, lane 4 – 0.50 μM, lane 5 – 1 μM, lane 6 – 2 μM, lane 7 – 4 μM, lane 8 – 6 μM, lane 9 – 8 μM and lane 10 – 10 μM.

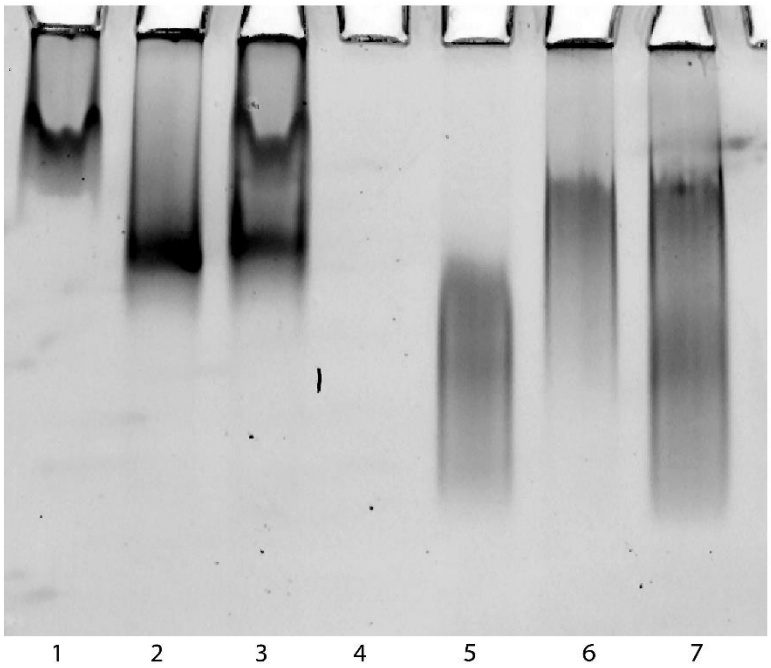


Figure A.3 Gel shift analysis of *S. cerevisiae* ES7 rRNA with multiple aaRSs. Interaction of ES7 rRNA with LysRS and TyrRS simultaneously could not be confirmed *in vitro*. Protein shifts, shown here, were visualized with SYPRO RED in a 5% native-PAGE containing 3% glycerol. Binding reactions were performed with 4 μM RNA and 10 μM aaRS at room temperature for 20 min in 20mM Tris-HEPES, pH 8.0. Lane 1-3 contains no RNA and different aaRSs: lane 1 – 10 μM TyrRS, lane 2 – 10 μM LysRS and lane 3 – 10 μM LysRS and 10 μM TyrRS. Lanes 4-7 contain 4 μM RNA and different aaRSs: lane 4 – no protein, lane 5 – 10 μM TyrRS, lane 6 – 10 μM LysRS and lane 7 – 10 μM TyrRS and 10 μM LysRS.

A.2.3 Thermal folding/unfolding of *S. cerevisiae* ES7

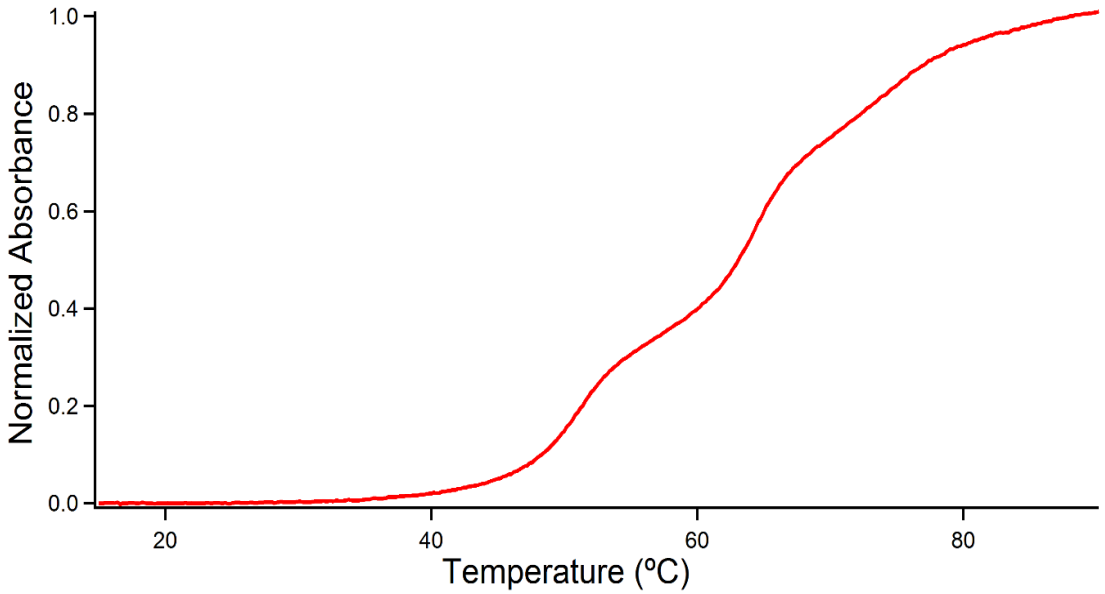


Figure A.4 Normalized, baseline corrected thermal denaturation curve of *S. cerevisiae* ES7 rRNA at 260 nm. Multiple melting transitions are observed during the unfolding process.

APPENDIX B

SUPPLEMENTARY INFORMARTION FOR CHAPTER 3

B.1 MATERIALS AND METHODS

Table B.1 DNA sequences encoding *H. sapiens* ES7, ES27, ES27a and ES27b.

Gene	Sequence (5' to 3')[a]
H. sapiens ES7	GAATTC*TAATACGACTCACTATAGGG*CGGGTGGGGTCCGCGCA GTCCGCCCGGAGGATTCAACCCGGCGGCGGGTCCGGCCGTGT CGGCGGCCCGGCGGATCTTTCCCGCCCCCCGTTCCTCCCGACC CCTCCACCCGCCCTCCCTTCCCCCGCCGCCCCTCCTCCTCCTC CCCGGAGGGGGCGGGCTCCGGCGGGTGCGGGGGTGGGCGGG CGGGGCCGGGGGTGGGGTCGGCGGGGGACCGTCCCCCGACC GGCGACCGGCCGCCGCCGGGCGCATTTCCACCGCGGCGGTGC GCCGCGACCGGCTCCGGGACGGCTGGGAAGGCCCGGCGGGG AAGGTGGCTCGGGGGGCCCCGTCCGTCCGTCCGTCCGTCCTC CTCCTCCCCCGTCTCCGCCCCCCGGCCCCGCGTCCTCCCTCGG GAGGGCGCGCGGGTCGGGGCGGCGGCGGCGGCGGCGGTGGC GGCGGCGGCGGCGGCGGCGGGACCGAAACCCCCCCCGAGTG TTACAGCCCCCCCGGCAGCAGCACTCGCCGAATCCCGGGGCC GAGGGAGCGAGACCCGTCGCCGCGCTCTCCCCCCTCCCGGCG CCCACCCCCGCGGGGAATCCCCCGCGAGGGGGGTCTCCCCCG CGGGGGCGCGCCGGCGTCTCCTCGTGGGGGGGCCGGGCCACC CCTCCCACGGCGCGACCGCTCTCCCACCCCTCCTCCCCGCGCC CCCGCCCCGGCGACGGGGGGGGTGCCGCGCGCGGGTCGGGG GGCGGGGCGGACTGTCCCCAGTGCGCCCCGGGCGGGTCGCGC CGTCGGGCCCGGGGGAGGTTCTCTCGGGGCCACGCGCGCGTC CCCCGAAGAGGGGGACGGCGGAGCGAGCGCACGGGGTCGGC GGCGACGTCGGCTACCCACCCG**TCGATCCGGTTCGCCGGAT** **CCAAATCGGGCTTCGGTCCGGTTC**AAGCTT
H. sapiens ES27	GAATTC*TAATACGACTCACTATAGGG*CAAGTACGTCGGTCGGGC TGGGGCGCGAAGCGGGGCTGGGCGCGCGCCGCGGCTGGACG AGGCGCCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCG GCCCTCCCCCGCCCCACCCCGCGCGCGCCGCTCGCTCCCTCCC CGCCCCGCGCCCTCTCTCTCTCTCTCTCCCCCGCTCCCCGTCCT CCCCCCTCCCCGGGGGAGCGCCGCGTGGGGGCGGCGGCGGG GGGAGAAGGGTCGGGGCGGCAGGGGCCGGCGGCGGCCCGCC GCGGGGCCCCGGCGGCGGGGGCACGGTCCCCCGCGAGGGGG GCCCGGGCACCCGGGGGGCCGGCGGCGGCGGCGACTCTGGA CGCGAGCCGGGCCCTTCCCGTGGATCGCCCCAGCTGCGGCGG GCGTCGCGGCCGCCCCCGGGGAGCCCGGCGGGCGCCGGCGC GCCCCCCCCCCCACCCCACGTCTCGTCGCGCGCGCGTCCGCTG GGGGCGGGGAGCGGTCGGGCGGCGGCGGTCGGCGGGCGGCG GGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCG

	TCCGCCCCCCGTTCCCCCCTCCTCCTCGGCGCGCGGCGGCGGC GGCGGCAGGCGGCGGAGGGGCCGCGGGCCGGTCCCCCCCGC CGGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGGCGCCTCGCC TCGGCCGGCGTACTTG**TCGATCCGGTTCGCCGGATCCAAAT** **CGGGCTTCGGTCCGGTTC**AAGCTT
H. sapiens ES27a	GAATTC*TAATACGACTCACTATAGGG*TGGGGCGCGAAGCGGGG CTGGGCGCGCGCCGCGGCTGGACGAGGCGCCGCCGCCCCCCC CACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCACC CCGCGCGCGCCGCTCGCTCCCTCCCCGCCCCGCGCCCTCTCTC TCTCTCTCTCCCCCGCTCCCCGTCCTCCCCCCTCCCCGGGGGA GCGCCGCGTGGGGGCGGCGGCGGGGGGAGAAGGGTCGGGGC GGCAGGGGCCGGCGGCGGCCCGCCGCGGGGCCCCGGCGGCG GGGGCACGGTCCCCCGCGAGGGGGGCCCGGGCACCCGGGGG GCCGGCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCC CGTGGATCGCCCCA**TCGATCCGGTTCGCCGGATCCAAATC** **GGGCTTCGGTCCGGTTC**AAGCTT
H. sapiens ES27b	GAATTC*TAATACGACTCACTATAG*GGCGGGCGTCGCGGCCGCCC CCGGGGAGCCCGGCGGGCGCCGGCGCGCCCCCCCCCCCACCC CACGTCTCGTCGCGCGCGCGTCCGCTGGGGGCGGGGAGCGGT CGGGCGGCGGCGGTCGGCGGGCGGCGGGGCGGGGCGGTTCG TCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCCCCGTTCCC CCCTCCTCCTCGGCGCGCGGCGGCGGCGGCGGCAGGCGGCGG AGGGGCCGCGGGCCGGTCCCCCCCGCCGGGTCCGCCCCCGGG GCCGCGGTTCCGCGCGGCGCCTCGCC**TCGATCCGGTTCGCC** **GGATCCAAATCGGGCTTCGGTCCGGTTC**AAGCTT

a) Restriction sites are underlined, the T7 promoter region is in italics and the Weeks cassette is in bold. Additional nucleotides were incorporated to ES27 to confer structural stability. These nucleotides are highlighted in gray.

Table B.2 DNA oligomers used in putative G-quadruplex studies.

Gene	Primer	Sequence (5' to 3')[b]
H. sapiens ES7 RNA	$Gq1_{ES7}$	GAATTC*TAATACGACTCACTATAGGG*CGGA**GGGG** **GCGGG**CTCCGGC**GGG**TGC**GGGGG**T**GGG**C**GGG** C**GGGG**CC**GGGGG**T**GGGG**TCGGC**GGGGG**ACCG AAGCTT
	$Gq2_{ES7}$	GAATTC*TAATACGACTCACTATAGGG*CCTC**GGG**A **GGG**CGCGC**GGG**TC**GGGG**CGGCAAGCTT
	$C1_{ES7}$	GAATTC*TAATACGACTCACTATAGGG*CGGA**GAGA** **GCAGT**CTCCGTC**AGA**TGC**GATAG**T**GAT**C**AGA**C **GTTG**CC**GTATG**T**GAAG**TCGGC**GAATG**ACCGAA GCTT
	$C2_{ES7}$	CCTC**AGT**A**TTG**CGCGC**AAG**TC**GATG**CGGC

b) Restriction sites are underlined, the T7 promoter region is in italics and G stacks are in bold for $Gq1_{ES7}$ and $Gq2_{ES7}$. Mutation of G stacks are also shown in bold for $C1_{ES7}$ and $C2_{ES7}$. $Gq1_{ES7}$, $Gq2_{ES7}$, $C1_{ES7}$ and their complimentary DNA strands were ordered from Eurofins MWG Operon. RNAs were transcribed *in vitro* with HiScribe™ T7 High Yield Synthesis Kit (New England BioLabs). $C2_{ES7}$ RNA was ordered from Integrated DNA Technologies.

Table B.3 Primers used in *H. sapiens* SHAPE RT reactions labeled with [6-FAM] at the 5'end.

Gene	Primer	Sequence (5' to 3')[c]
ES7 Weeks	1	GAACCGGACCGAAGCCCG
	2	GGATTCGGCGAGTGCTGCTG
	3	GCCGGTCGCCGGTCGG

c) Sequences ordered from Euro MWG Operon.

Table B.4 DNA anchor used in *H. sapiens* RNA pull-down assays (5' Biotin label).

Gene	Primer	Sequence (5' to 3')[c]
ES7, ES27, ES27b	1 - [5' Biotin label]	GAACCGGACCGAAGCCCG

B.2 RESULTS

B.2.1 Thermal folding/unfolding of human ES7

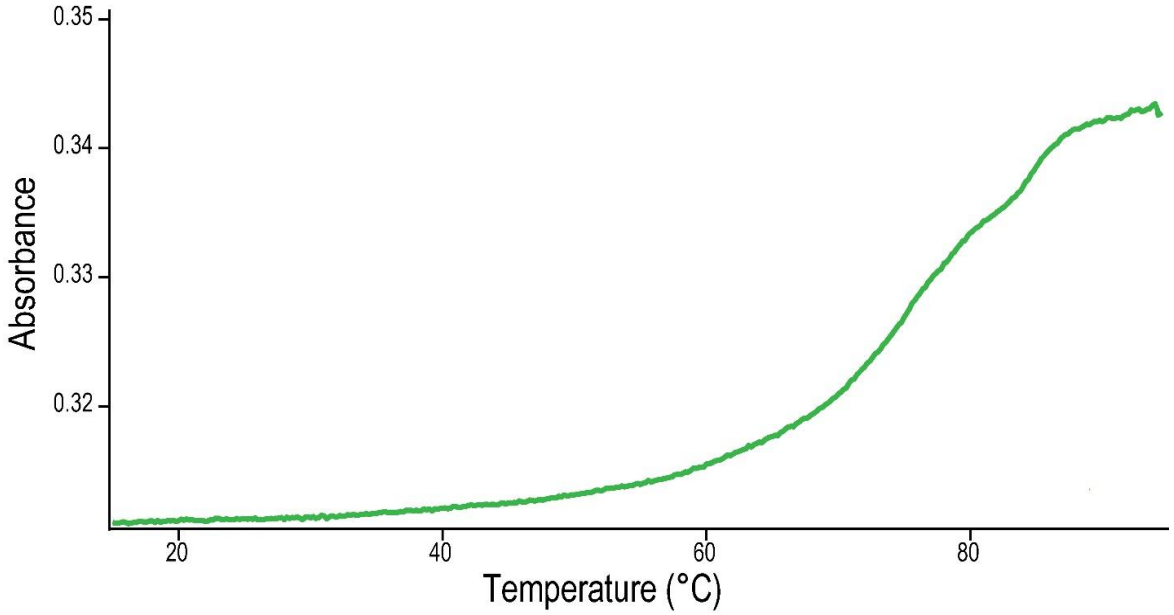


Figure B.1 Thermal denaturation curve of human ES7 rRNA at 260 nm. Multiple melting transitions are observed during the unfolding process.

Monovalent ion-dependence

We investigated the effect of alkali metal cations Li^+, Na^+, K^+, Rb^+ and Cs^+ on $T_{m,n}$ (Figure B.2, Table B.5). Most features of the unfolding profiles are conserved throughout the series. The peaks that are best defined and most prominent (third, fourth, fifth, sixth and seven peak using the melting curve of Na^+ as a reference) were observed for all cations. The two lowest temperature transitions which are broad and ill defined, are the least interpretable. All five prominent unfolding transitions of ES7 show the greatest $T_{m,n}$ in the presence of Li^+ and the lowest $T_{m,n}$ in the presence of Na^+. Cs^+ appears to be anomalous compared to the other alkali cations in that the highest temperature transition is shifted to lower $T_{m,n}$ than expected, relative to other $T_{m,n}$'s. In general, there is not strong evidence of specific interactions between the RNA and any given monovalent cation.

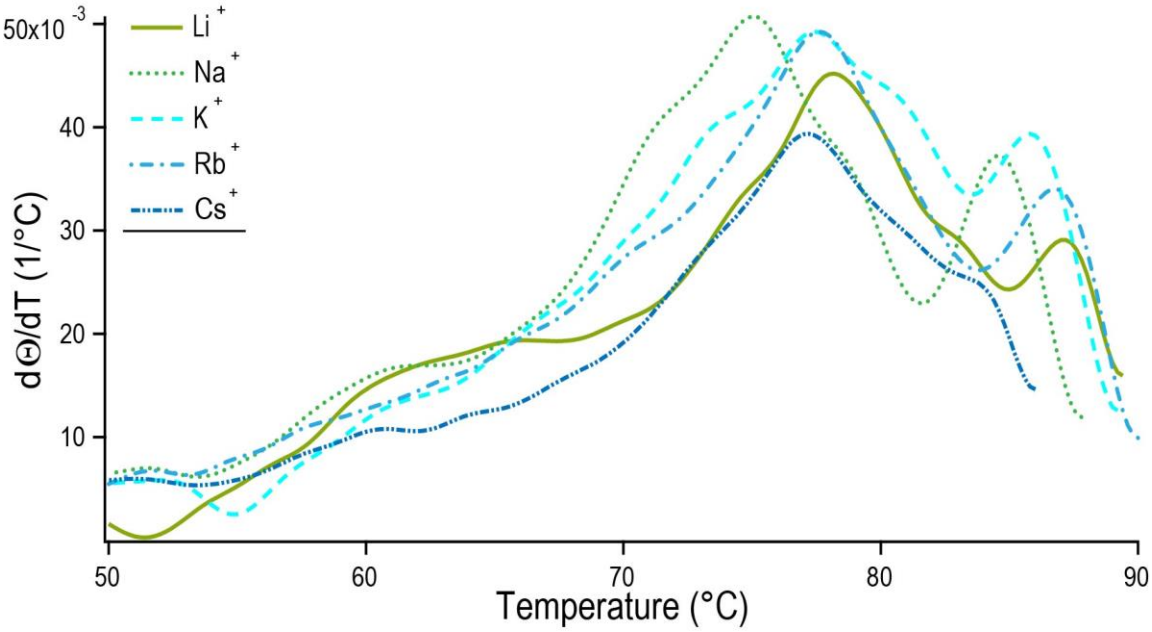


Figure B.2 Melting profiles of ES7 rRNA obtained at different alkali metal cations. The derivative of fraction unfolded at 260 nm with respect to temperature was computed. Data below 50°C have less statistical significance and is not shown. Experiments were performed in 18 mM Li^+, Na^+, K^+, Rb^+ or Cs^+ and 20 mM Tris-HEPES (pH 8.0). Figures were generated with IGOR Pro.

Table B.5 Parameter estimation of thermal unfolding curves of ES7 obtained in different monovalent cations. The data was fitted to the minimum number of transitions. Greater number of transitions cannot be discarded. Li^+, Na^+ and K^+ were fitted to a seven-state transition model. Rb^+ and Cs^+ were fitted to a six-state transition model. Transitions were numbered in order of appearance.

Parameter	Li^+	Na^+	K^+	Rb^+	Cs^+
H_1	50.4	56.5	49.1	31.2	63.0
$T_{m,1}$	36.1	36.7	36.2	44.7	49.3
A_1	1.9	0.7	0.6	1.5	0.4
H_2	223.7	71.7	53.5	42.4	73.0
$T_{m,2}$	46.3	49.9	50.3	63.0	62.0
A_2	0.2	0.9	1.7	3.4	0.9
H_3	47.8	73.1	72.2	80.6	100.0
$T_{m,3}$	65.1	61.4	62.8	71.3	73.0
A_3	4.3	2.2	2.0	2.1	1.7
H_4	97.7	96.4	66.4	116.1	165.0
$T_{m,4}$	74.7	71.0	73.8	77.8	77.2
A_4	1.8	3.1	6.4	2.6	1.1
H_5	147.5	151.4	174.6	66.9	120.0
$T_{m,5}$	78.5	75.4	77.7	82.3	81.3
A_5	2.1	2.3	1.2	3.1	1.3
H_6	172.4	208.3	229.4	206.8	190.0
$T_{m,6}$	82.9	79.2	80.9	87.1	84.9
A_6	1.1	1.0	0.9	1.3	0.5
H_7	191.4	176.4	142.4	N/A	N/A
$T_{m,7}$	87.4	84.7	86.0	N/A	N/A
A_7	1.6	2.4	3.3	N/A	N/A

B.2.2 ES-associated proteins

Human ES27a and ES27b rRNA

ES27 was cleaved in two fragments, ES27a and ES27b (Figure 3.2) to attempt to determine the locus of RNA-protein interactions within ES27. Since no pull-down experiments were performed with ES27a, the subset of proteins that associate with ES27

and not ES27a would appear to bind specifically to ES27b, although those assignments remain tentative.

Group $I_{ES27b,H}$ contains 12 proteins (Table B.14). Eight of those are identical or belong to the same protein family as Group $II_{ES27,H}$ (Table B.11). These Group $I_{ES27b,H}$ proteins include ADP-ribosylation factor 5 (ARF5), Cold-inducible RNA-binding protein (CIRBP), MSI1H, Rho GTPase-activating protein 17 (RHG17), Selenocysteine insertion sequence-binding protein 2 (SEBP2), Serine/arginine-rich splicing factor 9 (SRSF9), Ribonuclease H2 subunit A (RNH2A) and Tubulin beta-2A chain (TBB2A). E3 ubiquitin-protein ligase BRE1A and Protein phosphatase 1 regulatory subunit 12A (MYPT1) also from Group $I_{ES27b,H}$, share similarities with proteins identified in association with ES27 in all pull-down experiments. Conversely, seven of the sixteen proteins in Group $I_{ES27,H}$ belong to Group $II_{ES27b,H}$ (MSI1H, D11L8, BCL9L, ITSN2, SYFB, FA45B, RNH2B), while four proteins in Group $I_{ES27,H}$ share functions similar to proteins identified in Group $II_{ES27b,H}$ (MSI2H, ZCCHV, ARFG2, UBP7).

Group $II_{ES27b,H}$ contains at least twenty-four proteins classified as UBLs or UPSs or that are known to associate with UBLs or UPSs (Table B.15). These proteins include SAE2, UBE2O and ULA1, which are also seen to associate with ES7 and/or ES27. Group $II_{ES27b,H}$ contains several ADP ribosylation factors, translation initiation factors and aminoacyl tRNA synthetases (CysRS, ThrRS and PheRS). The proteins within Group $II_{ES27b,H}$ are in general consistent with those seen to associate with ES27 in both HEK293T and MDA-MB-231 cells.

Protein binding to human rRNA expansion segments

Table B.6 Proteins identified in all pull-down assays performed with ES7 and HEK293T cells (Group $I_{ES7,H}$).

Reference	Annotation[d]
O95400	CD2 antigen cytoplasmic tail-binding protein 2
P15531	Nucleoside diphosphate kinase A
P21281	V-type proton ATPase subunit B, brain isoform
P62495	Eukaryotic peptide chain release factor subunit 1
Q96G03	Phosphoglucomutase-2
Q9UBT2	SUMO-activating enzyme subunit 2
O75179	Ankyrin repeat domain-containing protein 17
O75496	Geminin
P07196	Neurofilament light polypeptide
Q13564	NEDD8-activating enzyme E1 regulatory subunit
Q32P28	Prolyl 3-hydroxylase 1
Q8NDI1	EH domain-binding protein 1
Q96EI5	Transcription elongation factor A protein-like 4
Q9C0C9	Ubiquitin-conjugating enzyme E2 O
Q9GZZ9	Ubiquitin-like modifier-activating enzyme 5

d) Proteins involved directly or through secondary interactions to the UBS are colored in gray.

Table B.7 Proteins identified in half of the pull-down assays performed with ES7 and HEK293T cells (Group $II_{ES7,H}$).

Reference	Annotation[d]
Q08752	Peptidyl-prolyl cis-trans isomerase D
P19623	Spermidine synthase
P49591	Serine-tRNA ligase, cytoplasmic
Q9NZ63	Uncharacterized protein C9orf78
P31150	Rab GDP dissociation inhibitor alpha
Q13439	Golgin subfamily A member 4
Q14694	Ubiquitin carboxyl-terminal hydrolase 10
Q8IWJ2	GRIP and coiled-coil domain-containing protein 2
Q8TCU4	Alstrom syndrome protein 1
Q9NR50	Translation initiation factor eIF-2B subunit gamma
Q9UPY3	Endoribonuclease Dicer
Q9Y6G9	Cytoplasmic dynein 1 light intermediate chain 1
O00154	Cytosolic acyl coenzyme A thioester hydrolase
O00487	26S proteasome non-ATPase regulatory subunit 14
O60934	Nibrin
O75914	Serine/threonine-protein kinase PAK 3
P00492	Hypoxanthine-guanine phosphoribosyltransferase

P09960	Leukotriene A-4 hydrolase
P35270	Sepiapterin reductase
P42766	60S ribosomal protein L35
P49189	4-trimethylaminobutyraldehyde dehydrogenase
P61011	Signal recognition particle 54 kDa protein
Q01581	Hydroxymethylglutaryl-CoA synthase, cytoplasmic
Q12849	G-rich sequence factor 1
Q14692	Ribosome biogenesis protein BMS1 homolog
Q16566	Calcium/calmodulin-dependent protein kinase type IV
Q3MHD2	Protein LSM12 homolog
Q5W0B1	RING finger protein 219
Q86U42	Polyadenylate-binding protein 2
Q8IYQ7	Threonine synthase-like 1
Q8TD19	Serine/threonine-protein kinase Nek9
Q92733	Proline-rich protein PRCC
Q99618	Cell division cycle-associated protein 3
Q9NUL3	Double-stranded RNA-binding protein Staufen homolog 2
Q9Y5X3	Sorting nexin-5
L0R819	Alternative protein ASNSD1
A6NDU8	UPF0600 protein C5orf51
O00429	Isoform 4 of Dynamin-1-like protein
O00507	Probable ubiquitin carboxyl-terminal hydrolase FAF-Y
O00512	B-cell CLL/lymphoma 9 protein
O14964	Hepatocyte growth factor-regulated tyrosine kinase substrate
O15305	Phosphomannomutase 2
O43423	Acidic leucine-rich nuclear phosphoprotein 32 family member C
O60508	Pre-mRNA-processing factor 17
O60814	Histone H2B type 1-K
O75122	CLIP-associating protein 2
O94776	Metastasis-associated protein MTA2
O94788	Retinal dehydrogenase 2
O95163	Elongator complex protein 1
O95219	Sorting nexin-4
O95232	Luc7-like protein 3
P00568	Adenylate kinase isoenzyme 1
P00918	Carbonic anhydrase 2
P07203	Glutathione peroxidase 1
P13861	cAMP-dependent protein kinase type II-alpha regulatory subunit
P19022	Cadherin-2
P22059	Oxysterol-binding protein 1
P30050	60S ribosomal protein L12
P30154	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform
P30566	Adenylosuccinate lyase
P35244	Replication protein A 14 kDa subunit

P37198	Nuclear pore glycoprotein p62
P43034	Platelet-activating factor acetylhydrolase IB subunit alpha
P43246	DNA mismatch repair protein Msh2
P49356	Protein farnesyltransferase subunit beta
P49585	Choline-phosphate cytidylyltransferase A
P51003	Poly(A) polymerase alpha
P51665	26S proteasome non-ATPase regulatory subunit 7
P54725	UV excision repair protein RAD23 homolog A
P61019	Ras-related protein Rab-2A
P61758	Prefoldin subunit 3
P62266	40S ribosomal protein S23
P63173	60S ribosomal protein L38
P68371	Tubulin beta-4B chain
P98170	E3 ubiquitin-protein ligase XIAP
Q01433	AMP deaminase 2
Q06210	Glutamine-fructose-6-phosphate aminotransferase [isomerizing] 1
Q06265	Exosome complex component RRP45
Q13126	S-methyl-5'-thioadenosine phosphorylase
Q13242	Serine/arginine-rich splicing factor 9
Q13542	Eukaryotic translation initiation factor 4E-binding protein 2
Q13554	Calcium/calmodulin-dependent protein kinase type II subunit beta
Q13642	Four and a half LIM domains protein 1
Q15057	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 2
Q15404	Ras suppressor protein 1
Q15637	Isoform 5 of Splicing factor 1
Q16531	DNA damage-binding protein 1
Q5PRF9	Protein Smaug homolog 2
Q5TBB1	Ribonuclease H2 subunit B
Q7Z2W4	Zinc finger CCCH-type antiviral protein 1
Q7Z3T8	Zinc finger FYVE domain-containing protein 16
Q7Z4Q2	HEAT repeat-containing protein 3
Q86U44	N6-adenosine-methyltransferase 70 kDa subunit
Q86V81	THO complex subunit 4
Q8N0Z6	Tetratricopeptide repeat protein 5
Q8NDV7	Trinucleotide repeat-containing gene 6A protein
Q8TBC4	NEDD8-activating enzyme E1 catalytic subunit
Q8WVM8	Sec1 family domain-containing protein 1
Q8WX92	Negative elongation factor B
Q92793	CREB-binding protein
Q92879	CUGBP Elav-like family member 1
Q969T9	WW domain-binding protein 2
Q96C36	Pyrroline-5-carboxylate reductase 2
Q96EK6	Glucosamine 6-phosphate N-acetyltransferase
Q96H79	Zinc finger CCCH-type antiviral protein 1-like
Q96S66	Chloride channel CLIC-like protein 1

Q99538	Legumain
Q99598	Translin-associated protein X
Q99873	Protein arginine N-methyltransferase 1
Q9BRA2	Thioredoxin domain-containing protein 17
Q9BUK6	Protein misato homolog 1
Q9H3P2	Negative elongation factor A
Q9H910	Hematological and neurological expressed 1-like protein
Q9NR12	PDZ and LIM domain protein 7
Q9P289	Serine/threonine-protein kinase MST4
Q9UJC3	Protein Hook homolog 1
Q9UJW0	Dynactin subunit 4
Q9ULJ3	Zinc finger and BTB domain-containing protein 21
Q9ULR0	Pre-mRNA-splicing factor ISY1 homolog
Q9UNP9	Peptidyl-prolyl cis-trans isomerase E
Q9Y263	Phospholipase A-2-activating protein
Q9Y2J2	Isoform B of Band 4.1-like protein 3
Q9Y3P9	Rab GTPase-activating protein 1
Q9Y6K9	NF-kappa-B essential modulator
Q9Y6X4	Soluble lamin-associated protein of 75 kDa
Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1
O15371	Eukaryotic translation initiation factor 3 subunit D
O60701	UDP-glucose 6-dehydrogenase
P47755	F-actin-capping protein subunit alpha-2
P53367	Arfaptin-1
P54578	Ubiquitin carboxyl-terminal hydrolase 14
P54920	Alpha-soluble NSF attachment protein
P61221	ATP-binding cassette sub-family E member 1
Q06787	Fragile X mental retardation protein 1
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1
O00161	Synaptosomal-associated protein 23
O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1
O43347	RNA-binding protein Musashi homolog 1
O75391	Sperm-associated antigen 7
P06737	Glycogen phosphorylase, liver form
P11413	Glucose-6-phosphate 1-dehydrogenase
P23921	Ribonucleoside-diphosphate reductase large subunit
P62701	40S ribosomal protein S4, X isoform
P63220	40S ribosomal protein S21
Q02878	60S ribosomal protein L6
Q12792	Twinfilin-1
Q15819	Ubiquitin-conjugating enzyme E2 variant 2
Q99436	Proteasome subunit beta type-7
Q9BVL2	Nucleoporin p58/p45
Q9NP97	Dynein light chain roadblock-type 1

Q9Y5Z4	Heme-binding protein 2
Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2
O14950	Myosin regulatory light chain 12B
O43395	U4/U6 small nuclear ribonucleoprotein Prp3
O43396	Thioredoxin-like protein 1
O75376	Nuclear receptor corepressor 1
O75475	PC4 and SFRS1-interacting protein
O75663	TIP41-like protein
O75886	Signal transducing adapter molecule 2
P06730	Eukaryotic translation initiation factor 4E
P09497	Isoform Non-brain of Clathrin light chain B
P21796	Voltage-dependent anion-selective channel protein 1
P26196	Probable ATP-dependent RNA helicase DDX6
P27361	Mitogen-activated protein kinase 3
P31483	Nucleolysin TIA-1 isoform p40
P42025	Beta-centractin
P42285	Superkiller viralicidic activity 2-like 2
P42696	RNA-binding protein 34
P45973	Chromobox protein homolog 5
P45985	Dual specificity mitogen-activated protein kinase kinase 4
P61106	Ras-related protein Rab-14
P62263	40S ribosomal protein S14
P62851	40S ribosomal protein S25
P82094	TATA element modulatory factor
Q00839	Isoform Short of Heterogeneous nuclear ribonucleoprotein U
Q04760	Lactoylglutathione lyase
Q13085	Acetyl-CoA carboxylase 1
Q13363	C-terminal-binding protein 1
Q14011	Cold-inducible RNA-binding protein
Q14195	Isoform LCRMP-4 of Dihydropyrimidinase-related protein 3
Q14527	Helicase-like transcription factor
Q15393	Splicing factor 3B subunit 3
Q15427	Splicing factor 3B subunit 4
Q8IY63	Angiomotin-like protein 1
Q8NEY8	Periphilin-1
Q92615	La-related protein 4B
Q96F24	Nuclear receptor-binding factor 2
Q99426	Tubulin-folding cofactor B
Q99576	TSC22 domain family protein 3
Q99966	Cbp/p300-interacting transactivator 1
Q9BPW8	Protein NipSnap homolog 1
Q9HD15	Steroid receptor RNA activator 1
Q9NPA8	Transcription and mRNA export factor ENY2
Q9NRA8	Eukaryotic translation initiation factor 4E transporter
Q9UBL3	Set1/Ash2 histone methyltransferase complex subunit ASH2

Q9UBW8	COP9 signalosome complex subunit 7a
Q9UKM9	RNA-binding protein Raly
Q9UL46	Proteasome activator complex subunit 2
Q9ULV3	Cip1-interacting zinc finger protein
Q9UMZ2	Synergin gamma
Q9UNH7	Sorting nexin-6
Q9Y2Q3	Glutathione S-transferase kappa 1
Q9Y570	Protein phosphatase methylesterase 1

Table B.8 Proteins identified in all pull-down assays performed with ES7 and MDA-MB-231 cells (Group $I_{ES7,M}$).

Reference	Annotation
Q8NBF2	NHL repeat-containing protein 2
Q86U42	Polyadenylate-binding protein 2
Q9Y570	Protein phosphatase methylesterase 1
Q969G3	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1
O75937	DnaJ homolog subfamily C member 8
Q96PU8	Protein quaking
P51149	Ras-related protein Rab-7a
P08865	40S ribosomal protein SA
P42285	Superkiller viralicidic activity 2-like 2

Table B.9 Proteins identified in half of the pull-down assays performed with ES7 and MDA-MB-231 cells (Group $II_{ES7,M}$).

Reference	Annotation[d]
Q92615	La-related protein 4B
P49189	4-trimethylaminobutyraldehyde dehydrogenase
P08758	Annexin A5
O15511	Actin-related protein 2/3 complex subunit 5
P27708	CAD protein
Q86TI2	Dipeptidyl peptidase 9
Q96C19	EF-hand domain-containing protein D2
Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2
P09382	Galectin-1
P43243	Matrin-3
Q9UHV9	Prefoldin subunit 2
Q96PK6	RNA-binding protein 14
P35244	Replication protein A 14 kDa subunit
P40429	60S ribosomal protein L13a
P35237	Serpin B6

O75410	Transforming acidic coiled-coil-containing protein 1
Q9UDY2	Tight junction protein ZO-2
Q15019	Septin-2
P38606	V-type proton ATPase catalytic subunit A
P20810	Isoform 5 of Calpastatin
P53675	Clathrin heavy chain 2
Q92905	COP9 signalosome complex subunit 5
P56545	C-terminal-binding protein 2
Q16531	DNA damage-binding protein 1
O43324	Eukaryotic translation elongation factor 1 epsilon-1
Q96RT1	Protein LAP2
Q2NKX8	DNA excision repair protein ERCC-6-like
O75369	Isoform 8 of Filamin-B
O60547	GDP-mannose 4,6 dehydratase
Q9Y2Q3	Glutathione S-transferase kappa 1
Q9NRV9	Heme-binding protein 1
Q9Y6K9	NF-kappa-B essential modulator
O00410	Importin-5
Q96PU5	E3 ubiquitin-protein ligase NEDD4-like
Q96IZ0	PRKC apoptosis WT1 regulator protein
P50479	PDZ and LIM domain protein 4
P08237	6-phosphofructokinase, muscle type
A2BFH1	Peptidyl-prolyl cis-trans isomerase A-like 4G
Q99873	Protein arginine N-methyltransferase 1
P25787	Proteasome subunit alpha type-2
P48556	26S proteasome non-ATPase regulatory subunit 8
P15927	Replication protein A 32 kDa subunit
P39019	40S ribosomal protein S19
Q15637	Splicing factor 1
Q9BWU0	Kanadaptin
P35270	Sepiapterin reductase
Q68CZ2	Tensin-3
Q96KR1	Zinc finger RNA-binding protein
Q16204	Coiled-coil domain-containing protein 6
Q7Z2W4	Zinc finger CCCH-type antiviral protein 1
P05556	Integrin beta-1
P09493	Isoform 10 of Tropomyosin alpha-1 chain
P61158	Actin-related protein 3
Q12802	A-kinase anchor protein 13
Q13625	Apoptosis-stimulating of p53 protein 2
Q96G03	Phosphoglucomutase-2
Q96NB3	Zinc finger protein 830
Q96RS6	NudC domain-containing protein 1
Q9UHX1	Poly(U)-binding-splicing factor PUF60
O43681	ATPase ASNA1

O95801	Tetratricopeptide repeat protein 4
P09211	Glutathione S-transferase P
P11172	Uridine 5'-monophosphate synthase
P12429	Annexin A3
P14649	Myosin light chain 6B
P25205	DNA replication licensing factor MCM3
P25789	Proteasome subunit alpha type-4
P29034	Protein S100-A2
P30154	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform
P30566	Adenylosuccinate lyase
P38117	Electron transfer flavoprotein subunit beta
P40763	Signal transducer and activator of transcription 3
P60900	Proteasome subunit alpha type-6
P68371	Tubulin beta-4B chain
P98082	Disabled homolog 2
Q08752	Peptidyl-prolyl cis-trans isomerase D
Q12849	G-rich sequence factor 1
Q13564	NEDD8-activating enzyme E1 regulatory subunit
Q14008	Cytoskeleton-associated protein 5
Q17RY0	Cytoplasmic polyadenylation element-binding protein 4
Q2M2I8	AP2-associated protein kinase 1
Q2TAL8	Glutamine-rich protein 1
Q5VT25	Serine/threonine-protein kinase MRCK alpha
Q6IN85	Serine/threonine-protein phosphatase 4 regulatory subunit 3A
Q8N6T3	ADP-ribosylation factor GTPase-activating protein 1
Q8TC07	TBC1 domain family member 15
Q8WWY3	U4/U6 small nuclear ribonucleoprotein Prp31
Q96K17	Transcription factor BTF3 homolog 4
Q99747	Gamma-soluble NSF attachment protein
Q9NQG5	Regulation of nuclear pre-mRNA domain-containing protein 1B
Q9NZZ3	Charged multivesicular body protein 5
Q9Y3Z3	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1
A6ND36	Protein FAM83G
O00244	Copper transport protein ATOX1
O15056	Synaptojanin-2
O15145	Actin-related protein 2/3 complex subunit 3
O15231	Isoform 2 of Zinc finger protein 185
O60361	Putative nucleoside diphosphate kinase
O75223	Gamma-glutamylcyclotransferase
O75525	KH domain-containing, RNA-binding, signal transduction-associated protein 3
O94905	Erlin-2
O95684	FGFR1 oncogene partner
O95747	Serine/threonine-protein kinase OSR1

P00167	Cytochrome b5
P00492	Hypoxanthine-guanine phosphoribosyltransferase
P02786	Transferrin receptor protein 1
P04049	RAF proto-oncogene serine/threonine-protein kinase
P04080	Cystatin-B
P10644	cAMP-dependent protein kinase type I-alpha regulatory subunit
P15153	Ras-related C3 botulinum toxin substrate 2
P18031	Tyrosine-protein phosphatase non-receptor type 1
P21399	Cytoplasmic aconitate hydratase
P23526	Adenosylhomocysteinase
P29692	Elongation factor 1-delta
P33240	Cleavage stimulation factor subunit 2
P40261	Nicotinamide N-methyltransferase
P42025	Beta-centractin
P46779	60S ribosomal protein L28
P49137	MAP kinase-activated protein kinase 2
P51148	Ras-related protein Rab-5C
P55795	Heterogeneous nuclear ribonucleoprotein H2
P61088	Ubiquitin-conjugating enzyme E2 N
P61221	ATP-binding cassette sub-family E member 1
P63151	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform
P83881	60S ribosomal protein L36a
Q04724	Transducin-like enhancer protein 1
Q04760	Lactoylglutathione lyase
Q06787	Fragile X mental retardation protein 1
Q13043	Serine/threonine-protein kinase 4
Q13107	Ubiquitin carboxyl-terminal hydrolase 4
Q13363	C-terminal-binding protein 1
Q13557	Calcium/calmodulin-dependent protein kinase type II subunit delta
Q13596	Sorting nexin-1
Q15366	Poly(rC)-binding protein 2
Q15370	Transcription elongation factor B polypeptide 2
Q15382	GTP-binding protein Rheb
Q16626	Male-enhanced antigen 1
Q3MHD2	Protein LSM12 homolog
Q52LJ0	Protein FAM98B
Q5VZF2	Muscleblind-like protein 2
Q6P3W7	SCY1-like protein 2
Q6WCQ1	Myosin phosphatase Rho-interacting protein
Q6YN16	Hydroxysteroid dehydrogenase-like protein 2
Q6ZSR9	Uncharacterized protein FLJ45252
Q7Z4H3	HD domain-containing protein 2
Q8IY63	Angiomotin-like protein 1
Q8N7H5	RNA polymerase II-associated factor 1 homolog

Q8ND56	Protein LSM14 homolog A
Q8TAT6	Nuclear protein localization protein 4 homolog
Q92804	TATA-binding protein-associated factor 2N
Q969T9	WW domain-binding protein 2
Q96C36	Pyrroline-5-carboxylate reductase 2
Q96FQ6	Protein S100-A16
Q96I25	Splicing factor 45
Q9BRP8	Partner of Y14 and mago
Q9GZZ9	Ubiquitin-like modifier-activating enzyme 5
Q9H223	EH domain-containing protein 4
Q9H3S7	Tyrosine-protein phosphatase non-receptor type 23
Q9H6S3	Epidermal growth factor receptor kinase substrate 8-like protein 2
Q9H8Y8	Golgi reassembly-stacking protein 2
Q9NUL3	Double-stranded RNA-binding protein Staufen homolog 2
Q9NVE7	Pantothenate kinase 4
Q9UBQ5	Eukaryotic translation initiation factor 3 subunit K
Q9UBT2	SUMO-activating enzyme subunit 2
Q9UJM3	ERBB receptor feedback inhibitor 1
Q9UMX5	Neudesin
Q9UMZ2	Synergin gamma
Q9UNF0	Protein kinase C and casein kinase substrate in neurons protein 2
Q9Y5Z4	Heme-binding protein 2
Q9Y678	Coatomer subunit gamma-1

Table B.10 Proteins identified in all pull-down assays performed with ES27 and HEK293T cells (Group $I_{ES27,H}$).

Reference	Annotation[d]
O43347	RNA-binding protein Musashi homolog 1
Q96DH6	RNA-binding protein Musashi homolog 2
A8MPP1	Putative ATP-dependent RNA helicase DDX11-like protein 8
Q86UU0	B-cell CLL/lymphoma 9-like protein
Q7Z2W4	Zinc finger CCCH-type antiviral protein 1
Q9NZM3	Intersectin-2
O15164	Transcription intermediary factor 1-alpha
Q9NSD9	Phenylalanine-tRNA ligase beta subunit
P12277	Creatine kinase B-type
O00151	PDZ and LIM domain protein 1
Q8N6H7	ADP-ribosylation factor GTPase-activating protein 2
P52294	Importin subunit alpha-5
Q6NSW5	Protein FAM45B
Q96EA4	Protein Spindly
Q93009	Ubiquitin carboxyl-terminal hydrolase 7
Q5TBB1	Ribonuclease H2 subunit B

Table B.11 Proteins identified in half of the pull-down assays performed with ES27 and HEK293T cells (Group $II_{ES27,H}$).

Reference	Annotation[d]
O00505	Importin subunit alpha-4
P46939	Utrophin
Q92905	COP9 signalosome complex subunit 5
A0MZ66	Shootin-1
O60343	TBC1 domain family member 4
O75116	Rho-associated protein kinase 2
O75369	Filamin-B
P22314	Ubiquitin-like modifier-activating enzyme 1
P35221	Catenin alpha-1
Q92804	TATA-binding protein-associated factor 2N
Q96LB3	Intraflagellar transport protein 74 homolog
Q9UHD1	Cysteine and histidine-rich domain-containing protein 1
Q9UNN5	FAS-associated factor 1
O14929	Histone acetyltransferase type B catalytic subunit
O75569	Interferon-inducible double-stranded RNA-dependent protein kinase activator A
P05023	Sodium/potassium-transporting ATPase subunit alpha-1
P61326	Protein mago nashi homolog
P84157	Isoform 2 of Matrix-remodeling-associated protein 7
Q07955	Serine/arginine-rich splicing factor 1
Q08170	Serine/arginine-rich splicing factor 4
Q15813	Tubulin-specific chaperone E
Q2NKX8	DNA excision repair protein ERCC-6-like
Q86VS8	Protein Hook homolog 3
Q8NEZ2	Vacuolar protein sorting-associated protein 37A
Q8WXD5	Gem-associated protein 6
Q9BT78	COP9 signalosome complex subunit 4
Q9HC38	Glyoxalase domain-containing protein 4
Q9HD26	Golgi-associated PDZ and coiled-coil motif-containing protein
Q9NQP4	Prefoldin subunit 4
Q9UKF6	Cleavage and polyadenylation specificity factor subunit 3
Q9UPY3	Endoribonuclease Dicer
A6NL28	Putative tropomyosin alpha-3 chain-like protein
O00170	AH receptor-interacting protein
O00264	Membrane-associated progesterone receptor component 1
O00512	B-cell CLL/lymphoma 9 protein
O14579	Coatomer subunit epsilon
O14639	Actin-binding LIM protein 1
O14744	Protein arginine N-methyltransferase 5

O60488	Long-chain-fatty-acid-CoA ligase 4
O75044	SLIT-ROBO Rho GTPase-activating protein 2
O75663	TIP41-like protein
O75683	Surfeit locus protein 6
O95453	Poly(A)-specific ribonuclease PARN
O95817	BAG family molecular chaperone regulator 3
P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2
P22681	E3 ubiquitin-protein ligase CBL
P24534	Elongation factor 1-beta
P25398	40S ribosomal protein S12
P27448	MAP/microtubule affinity-regulating kinase 3
P30307	M-phase inducer phosphatase 3
P35520	Cystathionine beta-synthase
P36543	V-type proton ATPase subunit E 1
P51003	Poly(A) polymerase alpha
P52756	RNA-binding protein 5
P60174	Triosephosphate isomerase
P61106	Ras-related protein Rab-14
P67936	Tropomyosin alpha-4 chain
Q07157	Tight junction protein ZO-1
Q10567	AP-1 complex subunit beta-1
Q12972	Nuclear inhibitor of protein phosphatase 1
Q13206	Probable ATP-dependent RNA helicase DDX10
Q13492	Phosphatidylinositol-binding clathrin assembly protein
Q13838	Spliceosome RNA helicase DDX39B
Q14155	Isoform 1 of Rho guanine nucleotide exchange factor 7
Q14315	Filamin-C
Q14684	Ribosomal RNA processing protein 1 homolog B
Q15287	RNA-binding protein with serine-rich domain 1
Q15369	Transcription elongation factor B polypeptide 1
Q15427	Splicing factor 3B subunit 4
Q15814	Tubulin-specific chaperone C
Q16531	DNA damage-binding protein 1
Q16718	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5
Q4VC05	B-cell CLL/lymphoma 7 protein family member A
Q5H9R7	Serine/threonine-protein phosphatase 6 regulatory subunit 3
Q5T5Y3	Calmodulin-regulated spectrin-associated protein 1
Q5TAX3	Terminal uridylyltransferase 4
Q69YQ0	Cytospin-A
Q8TAE8	Growth arrest and DNA damage-inducible proteins-interacting protein 1
Q96CT7	Coiled-coil domain-containing protein 124
Q96DF8	Protein DGCR14
Q96KB5	Lymphokine-activated killer T-cell-originated protein kinase
Q96MW1	Coiled-coil domain-containing protein 43
Q96NB3	Zinc finger protein 830

Q96RU3	Formin-binding protein 1
Q9BRX5	DNA replication complex GINS protein PSF3
Q9BVG4	Protein PBDC1
Q9BW85	Coiled-coil domain-containing protein 94
Q9BXB4	Oxysterol-binding protein-related protein 11
Q9H3Q1	Cdc42 effector protein 4
Q9HAU0	Pleckstrin homology domain-containing family A member 5
Q9NQ92	Coordinator of PRMT5 and differentiation stimulator
Q9NR50	Translation initiation factor eIF-2B subunit gamma
Q9NRF9	DNA polymerase epsilon subunit 3
Q9NX58	Cell growth-regulating nucleolar protein
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1
Q9NZT2	Opioid growth factor receptor
Q9P0L0	Vesicle-associated membrane protein-associated protein A
Q9UHD8	Septin-9
Q9ULV4	Coronin-1C
Q9Y2L1	Exosome complex exonuclease RRP44
Q9Y4W2	Ribosomal biogenesis protein LAS1L
Q9Y570	Protein phosphatase methylesterase 1
Q9Y5M8	Signal recognition particle receptor subunit beta
C9J7I0	Uncharacterized protein
Q14527	Helicase-like transcription factor
P08237	6-phosphofructokinase, muscle type
P61160	Actin-related protein 2
P00338	L-lactate dehydrogenase A chain
O95433	Activator of 90 kDa heat shock protein ATPase homolog 1
Q8NB90	Spermatogenesis-associated protein 5
O60547	GDP-mannose 4,6 dehydratase
P31939	Bifunctional purine biosynthesis protein PURH
P22307	Non-specific lipid-transfer protein
P18621	60S ribosomal protein L17
Q13885	Tubulin beta-2A chain
Q14011	Cold-inducible RNA-binding protein
Q15293	Reticulocalbin-1
Q96A73	Putative monooxygenase p33MONOX
Q86US8	Telomerase-binding protein EST1A
Q92973	Transportin-1
Q92900	Regulator of nonsense transcripts 1
Q96FZ2	Embryonic stem cell-specific 5-hydroxymethylcytosine-binding protein
Q13557	Calcium/calmodulin-dependent protein kinase type II subunit delta
Q9C0C2	182 kDa tankyrase-1-binding protein
Q8IX90	Spindle and kinetochore-associated protein 3
P31350	Ribonucleoside-diphosphate reductase subunit M2
Q9UGU0	Transcription factor 20
Q9HAV4	Exportin-5

Q9NPD8	Ubiquitin-conjugating enzyme E2 T
P16278	Beta-galactosidase
Q86TP1	Protein prune homolog
Q13242	Serine/arginine-rich splicing factor 9
Q9BVM2	Protein DPCD
Q9P2N5	RNA-binding protein 27
P49588	Alanine-tRNA ligase, cytoplasmic
Q99614	Tetratricopeptide repeat protein 1
Q6NSI4	Uncharacterized protein CXorf57
Q6P2H3	Centrosomal protein of 85 kDa
P42696	RNA-binding protein 34
Q9NR31	GTP-binding protein SAR1a
Q9C005	Protein dpy-30 homolog
Q5PRF9	Protein Smaug homolog 2
Q8NCA5	Protein FAM98A
Q9NR46	Endophilin-B2
Q9NZD8	Maspardin
Q6UVJ0	Spindle assembly abnormal protein 6 homolog
Q9UI26	Importin-11
P54619	5'-AMP-activated protein kinase subunit gamma-1
P62899	60S ribosomal protein L31
O14641	Segment polarity protein dishevelled homolog DVL-2
Q6ZSR9	Uncharacterized protein FLJ45252
Q96T21	Selenocysteine insertion sequence-binding protein 2
Q7Z4Q2	HEAT repeat-containing protein 3
O60361	Putative nucleoside diphosphate kinase
Q96T37	Putative RNA-binding protein 15
Q96SI9	Spermatid perinuclear RNA-binding protein
Q9NRA8	Eukaryotic translation initiation factor 4E transporter
P04049	RAF proto-oncogene serine/threonine-protein kinase
Q15418	Ribosomal protein S6 kinase alpha-1
P00558	Phosphoglycerate kinase 1
Q13625	Apoptosis-stimulating of p53 protein 2
Q6P1N0	Coiled-coil and C2 domain-containing protein 1A
P31483	Nucleolysin TIA-1 isoform p40
Q00013	55 kDa erythrocyte membrane protein
P25685	DnaJ homolog subfamily B member 1
Q99590	Protein SCAF11
P17858	6-phosphofructokinase, liver type
O95721	Synaptosomal-associated protein 29
Q9BW83	Intraflagellar transport protein 27 homolog
O94762	ATP-dependent DNA helicase Q5
P18124	60S ribosomal protein L7
Q16864	V-type proton ATPase subunit F
Q9BUF5	Tubulin beta-6 chain

Q9UKZ1	CCR4-NOT transcription complex subunit 11
O75792	Ribonuclease H2 subunit A
Q9UKN8	General transcription factor 3C polypeptide 4
Q96EP5	DAZ-associated protein 1
O94830	Phospholipase DDHD2
P51809	Vesicle-associated membrane protein 7
P57076	UPF0769 protein C21orf59
Q92769	Histone deacetylase 2
P62847	40S ribosomal protein S24
Q7Z7F0	UPF0469 protein KIAA0907
P38919	Eukaryotic initiation factor 4A-III
P30154	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform
Q9UHF7	Zinc finger transcription factor Trps1
Q96GX5	Serine/threonine-protein kinase greatwall
P10619	Lysosomal protective protein
Q01844	RNA-binding protein EWS
P11908	Ribose-phosphate pyrophosphokinase 2
Q15434	RNA-binding motif, single-stranded-interacting protein 2
O75506	Heat shock factor-binding protein 1
Q8IY67	Ribonucleoprotein PTB-binding 1
Q86XZ4	Spermatogenesis-associated serine-rich protein 2
Q9UBL3	Set1/Ash2 histone methyltransferase complex subunit ASH2
Q96KG9	N-terminal kinase-like protein
Q93034	Cullin-5
Q9Y237	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4
Q9BY89	Uncharacterized protein KIAA1671
O15397	Importin-8
O95819	Mitogen-activated protein kinase kinase kinase kinase 4
Q9H223	EH domain-containing protein 4
Q14498	RNA-binding protein 39
Q9P031	Thyroid transcription factor 1-associated protein 26
Q9Y580	RNA-binding protein 7
P40616	ADP-ribosylation factor-like protein 1
O75844	CAAX prenyl protease 1 homolog
Q9BVJ6	U3 small nucleolar RNA-associated protein 14 homolog A
O95218	Zinc finger Ran-binding domain-containing protein 2
P53675	Clathrin heavy chain 2
Q9Y2X9	Zinc finger protein 281
O15381	Nuclear valosin-containing protein-like
Q5VT06	Centrosome-associated protein 350
Q15596	Nuclear receptor coactivator 2
Q8TCG1	Protein CIP2A
Q99755	Isoform 2 of Phosphatidylinositol 4-phosphate 5-kinase type-1 alpha
Q14232	Translation initiation factor eIF-2B subunit alpha

Q92917	G patch domain and KOW motifs-containing protein
Q9GZS3	WD repeat-containing protein 61
P42766	60S ribosomal protein L35
Q9Y450	HBS1-like protein
P12814	Alpha-actinin-1
Q96L91	E1A-binding protein p400
Q6PK04	Coiled-coil domain-containing protein 137
P53350	Serine/threonine-protein kinase PLK1
Q9BTC0	Death-inducer obliterator 1
Q9BVC5	Ashwin
Q92997	Segment polarity protein dishevelled homolog DVL-3
Q8NG31	Protein CASC5
P99999	Cytochrome c
P41250	Glycine-tRNA ligase
Q9ULR0	Pre-mRNA-splicing factor ISY1 homolog
Q13555	Calcium/calmodulin-dependent protein kinase type II subunit gamma
P61981	14-3-3 protein gamma
O00567	Nucleolar protein 56
Q9Y3B4	Pre-mRNA branch site protein p14
Q15274	Nicotinate-nucleotide pyrophosphorylase [carboxylating]
Q969J3	Loss of heterozygosity 12 chromosomal region 1 protein
Q9BXK5	Bcl-2-like protein 13
P50750	Cyclin-dependent kinase 9
Q5VWZ2	Lysophospholipase-like protein 1
O14730	Serine/threonine-protein kinase RIO3
Q9BVC3	Sister chromatid cohesion protein DCC1
Q14469	Transcription factor HES-1
Q8WUH6	UPF0444 transmembrane protein C12orf23
P21108	Ribose-phosphate pyrophosphokinase 3
O94776	Metastasis-associated protein MTA2
Q9UBE0	SUMO-activating enzyme subunit 1
Q96J01	THO complex subunit 3
P15880	40S ribosomal protein S2
P15531	Nucleoside diphosphate kinase A
Q8IU60	m7GpppN-mRNA hydrolase
Q9NQW6	Actin-binding protein anillin
Q676U5	Autophagy-related protein 16-1
O75832	26S proteasome non-ATPase regulatory subunit 10
O95758	Polypyrimidine tract-binding protein 3
Q9BW61	DET1- and DDB1-associated protein 1
P61026	Ras-related protein Rab-10
Q8IWR0	Zinc finger CCCH domain-containing protein 7A
Q14257	Reticulocalbin-2
O75694	Nuclear pore complex protein Nup155

Table B.12 Proteins identified in all pull-down assays performed with ES27 and MDA-MB-231 cells (Group $I_{ES27,M}$).

Reference	Annotation
O75822	Eukaryotic translation initiation factor 3 subunit J
Q15637	Splicing factor 1
O95425	Supervillin
P18583	Protein SON
P53621	Coatomer subunit alpha
Q5SW79	Centrosomal protein of 170 kDa
P20290	Transcription factor BTF3

Table B.13 Proteins identified in half of the pull-down assays performed with ES27 and MDA-MB-231 cells (Group $II_{ES27,M}$).

Reference	Annotation[d]
Q14204	Cytoplasmic dynein 1 heavy chain 1
Q9Y5K6	CD2-associated protein
Q96A49	Synapse-associated protein 1
Q00688	Peptidyl-prolyl cis-trans isomerase FKBP3
P52272	Heterogeneous nuclear ribonucleoprotein M
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha
Q96PK6	RNA-binding protein 14
Q9Y2D5	A-kinase anchor protein 2
O14818	Proteasome subunit alpha type-7
Q9Y696	Chloride intracellular channel protein 4
P11216	Glycogen phosphorylase, brain form
Q9Y6G9	Cytoplasmic dynein 1 light intermediate chain 1
P60174	Triosephosphate isomerase
P35580	Myosin-10
Q07866	Kinesin light chain 1
O95793	Double-stranded RNA-binding protein Staufen homolog 1
Q15084	Protein disulfide-isomerase A6
P62753	40S ribosomal protein S6
P22234	Multifunctional protein ADE2
P60228	Eukaryotic translation initiation factor 3 subunit E
Q99417	C-Myc-binding protein
Q99436	Proteasome subunit beta type-7
P18031	Tyrosine-protein phosphatase non-receptor type 1
Q9H9H4	Vacuolar protein sorting-associated protein 37B
O60506	Heterogeneous nuclear ribonucleoprotein Q
Q96HJ9	Isoform 2 of UPF0562 protein C7orf55
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2
P50552	Vasodilator-stimulated phosphoprotein

Q9BWU0	Kanadaptin
Q13148	TAR DNA-binding protein 43
Q99747	Gamma-soluble NSF attachment protein
Q71RC2	La-related protein 4
Q99729	Heterogeneous nuclear ribonucleoprotein A/B
Q9BRA2	Thioredoxin domain-containing protein 17
Q86UU0	B-cell CLL/lymphoma 9-like protein
P62241	40S ribosomal protein S8
Q14974	Importin subunit beta-1
Q5H9R7	Serine/threonine-protein phosphatase 6 regulatory subunit 3
Q6NZY4	Zinc finger CCHC domain-containing protein 8
P42025	Beta-centractin
P09914	Interferon-induced protein with tetratricopeptide repeats 1
P23284	Peptidyl-prolyl cis-trans isomerase B
Q16762	Thiosulfate sulfurtransferase
O60547	GDP-mannose 4,6 dehydratase
Q9Y570	Protein phosphatase methylesterase 1
O60841	Eukaryotic translation initiation factor 5B
P51114	Fragile X mental retardation syndrome-related protein 1
P06737	Glycogen phosphorylase, liver form
O60869	Endothelial differentiation-related factor 1
P06703	Protein S100-A6
Q9UHD8	Septin-9
P14866	Heterogeneous nuclear ribonucleoprotein L
P26583	High mobility group protein B2
Q969H8	UPF0556 protein C19orf10
P41250	Glycine-tRNA ligase
P09493	Isoform 2 of Tropomyosin alpha-1 chain
Q9BVP2	Guanine nucleotide-binding protein-like 3
O15511	Actin-related protein 2/3 complex subunit 5
P00568	Adenylate kinase isoenzyme 1
P84157	Isoform 2 of Matrix-remodeling-associated protein 7
Q15843	NEDD8
Q13442	28 kDa heat- and acid-stable phosphoprotein
Q9Y2V2	Calcium-regulated heat stable protein 1
Q8IWC1	MAP7 domain-containing protein 3
P08237	6-phosphofructokinase, muscle type
Q9NTJ3	Structural maintenance of chromosomes protein 4
Q16204	Coiled-coil domain-containing protein 6
P60983	Glia maturation factor beta
P29218	Inositol monophosphatase 1
A0MZ66	Shootin-1
P46013	Antigen KI-67
Q9HC38	Glyoxalase domain-containing protein 4
Q5JSH3	WD repeat-containing protein 44

Q5T5P2	Sickle tail protein homolog
Q13573	SNW domain-containing protein 1
P09211	Glutathione S-transferase P
Q9BY77	Polymerase delta-interacting protein 3
P55884	Eukaryotic translation initiation factor 3 subunit B
P49588	Alanine-tRNA ligase, cytoplasmic
P62847	40S ribosomal protein S24
O75223	Gamma-glutamylcyclotransferase
P20042	Eukaryotic translation initiation factor 2 subunit 2
P40925	Malate dehydrogenase, cytoplasmic
Q8NHV4	Protein NEDD1
Q9UMX5	Neudesin
O95757	Heat shock 70 kDa protein 4L
Q6IN85	Serine/threonine-protein phosphatase 4 regulatory subunit 3A
O75717	WD repeat and HMG-box DNA-binding protein 1
Q15678	Tyrosine-protein phosphatase non-receptor type 14
Q96RT1	Protein LAP2
Q9Y2Q3	Glutathione S-transferase kappa 1
O43815	Striatin
Q9NZU5	LIM and cysteine-rich domains protein 1
Q13564	NEDD8-activating enzyme E1 regulatory subunit
Q86UE4	Protein LYRIC
P53004	Biliverdin reductase A
Q9UPQ9	Trinucleotide repeat-containing gene 6B protein
O60524	Nuclear export mediator factor NEMF
Q99832	T-complex protein 1 subunit eta
Q15276	Rab GTPase-binding effector protein 1
P05556	Integrin beta-1
Q9NR45	Sialic acid synthase
Q96K76	Ubiquitin carboxyl-terminal hydrolase 47
P12081	Histidine-tRNA ligase, cytoplasmic
O95684	FGFR1 oncogene partner
Q86U42	Polyadenylate-binding protein 2
O75410	Transforming acidic coiled-coil-containing protein 1
O00762	Ubiquitin-conjugating enzyme E2 C
P61758	Prefoldin subunit 3
Q8NBF2	NHL repeat-containing protein 2
O15460	Prolyl 4-hydroxylase subunit alpha-2
Q99598	Translin-associated protein X
O75575	DNA-directed RNA polymerase III subunit RPC9
O14497	AT-rich interactive domain-containing protein 1A
O76070	Gamma-synuclein
O43583	Density-regulated protein
O00161	Synaptosomal-associated protein 23
O75369	Isoform 8 of Filamin-B

Q96PU8	Protein quaking
P05388	60S acidic ribosomal protein P0
Q14151	Scaffold attachment factor B2
Q14677	Clathrin interactor 1
Q07065	Cytoskeleton-associated protein 4
Q13542	Eukaryotic translation initiation factor 4E-binding protein 2
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1
Q9P0L0	Vesicle-associated membrane protein-associated protein A
P61289	Proteasome activator complex subunit 3
P49023	Paxillin
P08708	40S ribosomal protein S17
P47755	F-actin-capping protein subunit alpha-2
O43237	Cytoplasmic dynein 1 light intermediate chain 2
Q9Y5Z4	Heme-binding protein 2
Q9NRL3	Striatin-4
P07711	Cathepsin L1
Q09028	Histone-binding protein RBBP4
Q6FIF0	AN1-type zinc finger protein 6
Q9NVA2	Septin-11
O15305	Phosphomannomutase 2
Q15811	Intersectin-1
P07203	Glutathione peroxidase 1
Q96EP5	DAZ-associated protein 1
O15371	Eukaryotic translation initiation factor 3 subunit D
O75886	Signal transducing adapter molecule 2
Q9BTT0	Acidic leucine-rich nuclear phosphoprotein 32 family member E
Q12802	A-kinase anchor protein 13
P78344	Eukaryotic translation initiation factor 4 gamma 2
O95373	Importin-7
Q8IXM2	Chromatin complexes subunit BAP18
Q8WUF5	RelA-associated inhibitor
Q8WXX5	DnaJ homolog subfamily C member 9
Q15003	Condensin complex subunit 2
P43034	Platelet-activating factor acetylhydrolase IB subunit alpha
O75506	Heat shock factor-binding protein 1
Q99618	Cell division cycle-associated protein 3
Q9UBT2	SUMO-activating enzyme subunit 2
Q9Y2W2	WW domain-binding protein 11
A0AVT1	Ubiquitin-like modifier-activating enzyme 6
P60953	Cell division control protein 42 homolog
Q92541	RNA polymerase-associated protein RTF1 homolog
Q9Y4Z0	U6 snRNA-associated Sm-like protein LSm4
Q99615	DnaJ homolog subfamily C member 7
O60282	Kinesin heavy chain isoform 5C
P34910	Protein EVI2B

Q9Y263	Phospholipase A-2-activating protein
L0R819	Alternative protein ASNSD1
P61201	COP9 signalosome complex subunit 2
Q8TCU4	Alstrom syndrome protein 1
Q9HC35	Echinoderm microtubule-associated protein-like 4
P25787	Proteasome subunit alpha type-2
P55145	Mesencephalic astrocyte-derived neurotrophic factor
Q9UKK9	ADP-sugar pyrophosphatase
Q9BVG4	Protein PBDC1
Q13283	Ras GTPase-activating protein-binding protein 1
P99999	Cytochrome c
Q15020	Squamous cell carcinoma antigen recognized by T-cells 3
O43776	Asparagine-tRNA ligase, cytoplasmic
Q9H1H9	Kinesin-like protein KIF13A
O43491	Band 4.1-like protein 2
P53618	Coatomer subunit beta
P52292	Importin subunit alpha-1
P49368	T-complex protein 1 subunit gamma
P62266	40S ribosomal protein S23

Table B.14 Proteins identified in all pull-down assays performed with ES27b and HEK293T cells (Group $I_{ES27b,H}$).

Reference	Annotation[d]
P84085	ADP-ribosylation factor 5
Q5VTR2	E3 ubiquitin-protein ligase BRE1A
Q14011	Cold-inducible RNA-binding protein
Q9Y696	Chloride intracellular channel protein 4
P40925	Malate dehydrogenase, cytoplasmic
O43347	RNA-binding protein Musashi homolog 1
O14974	Protein phosphatase 1 regulatory subunit 12A
Q68EM7	Rho GTPase-activating protein 17
O75792	Ribonuclease H2 subunit A
Q96T21	Selenocysteine insertion sequence-binding protein 2
Q13242	Serine/arginine-rich splicing factor 9
Q13885	Tubulin beta-2A chain

Table B.15 Proteins identified in half of the pull-down assays performed with ES27b and HEK293T cells (Group $II_{ES27b,H}$).

Reference	Annotation[d]
O95757	Heat shock 70 kDa protein 4L
Q86U44	N6-adenosine-methyltransferase 70 kDa subunit

Q1KMD3	Heterogeneous nuclear ribonucleoprotein U-like protein 2
O43395	U4/U6 small nuclear ribonucleoprotein Prp3
Q15025	TNFAIP3-interacting protein 1
P22307	Non-specific lipid-transfer protein
P31949	Protein S100-A11
Q9NXV6	CDKN2A-interacting protein
O43765	Small glutamine-rich tetratricopeptide repeat-containing protein alpha
P61758	Prefoldin subunit 3
Q02878	60S ribosomal protein L6
Q9BXP5	Serrate RNA effector molecule homolog
Q86V81	THO complex subunit 4
Q9NR56	Muscleblind-like protein 1
P63208	S-phase kinase-associated protein 1
Q9NQG5	Regulation of nuclear pre-mRNA domain-containing protein 1B
Q92879	CUGBP Elav-like family member 1
Q92530	Proteasome inhibitor PI31 subunit
P35611	Alpha-adducin
Q96AT1	Uncharacterized protein KIAA1143
Q9UQN3	Charged multivesicular body protein 2b
Q9NPI6	mRNA-decapping enzyme 1A
P55884	Eukaryotic translation initiation factor 3 subunit B
P62263	40S ribosomal protein S14
Q92905	COP9 signalosome complex subunit 5
P07196	Neurofilament light polypeptide
Q2TAL8	Glutamine-rich protein 1
O00193	Small acidic protein
P04792	Heat shock protein beta-1
O95163	Elongator complex protein 1
O00231	26S proteasome non-ATPase regulatory subunit 11
Q9GZR7	ATP-dependent RNA helicase DDX24
P08621	U1 small nuclear ribonucleoprotein 70 kDa
Q9UJA5	tRNA (adenine(58)-N(1))-methyltransferase non-catalytic subunit TRM6
Q9ULV4	Coronin-1C
Q9UKG1	DCC-interacting protein 13-alpha
Q9UEE9	Craniofacial development protein 1
Q9H3U1	Protein unc-45 homolog A
O75312	Zinc finger protein ZPR1
P49458	Signal recognition particle 9 kDa protein
Q13153	Serine/threonine-protein kinase PAK 1
P62805	Histone H4
P63220	40S ribosomal protein S21
P61019	Ras-related protein Rab-2A
O43809	Cleavage and polyadenylation specificity factor subunit 5
O95218	Zinc finger Ran-binding domain-containing protein 2
P38919	Eukaryotic initiation factor 4A-III

Q9H2J4	Phosducin-like protein 3
O94888	UBX domain-containing protein 7
Q9NR31	GTP-binding protein SAR1a
O15371	Eukaryotic translation initiation factor 3 subunit D
P08133	Annexin A6
Q6NSW5	Protein FAM45B
Q99570	Phosphoinositide 3-kinase regulatory subunit 4
P18085	ADP-ribosylation factor 4
Q96FW1	Ubiquitin thioesterase OTUB1
Q96H79	Zinc finger CCCH-type antiviral protein 1-like
Q96IK1	Biorientation of chromosomes in cell division protein 1
O00429	Isoform 4 of Dynamin-1-like protein
P50750	Cyclin-dependent kinase 9
P61088	Ubiquitin-conjugating enzyme E2 N
Q9UKF6	Cleavage and polyadenylation specificity factor subunit 3
Q9BUL8	Programmed cell death protein 10
Q8NE71	ATP-binding cassette sub-family F member 1
O75575	DNA-directed RNA polymerase III subunit RPC9
Q92844	TRAF family member-associated NF-kappa-B activator
P11908	Ribose-phosphate pyrophosphokinase 2
P19623	Spermidine synthase
P31483	Nucleolysin TIA-1 isoform p40
Q9P258	Protein RCC2
Q7L5N1	COP9 signalosome complex subunit 6
O75886	Signal transducing adapter molecule 2
Q92973	Transportin-1
O95782	AP-2 complex subunit alpha-1
Q5JTH9	RRP12-like protein
Q8NEY8	Periphilin-1
Q15042	Rab3 GTPase-activating protein catalytic subunit
Q9Y5B6	PAX3- and PAX7-binding protein 1
Q8IWC1	MAP7 domain-containing protein 3
O75526	RNA-binding motif protein, X-linked-like-2
P22090	40S ribosomal protein S4, Y isoform 1
Q15057	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 2
Q9P289	Serine/threonine-protein kinase MST4
P39019	40S ribosomal protein S19
Q9UM54	Unconventional myosin-VI
O75844	CAAX prenyl protease 1 homolog
Q9NTZ6	RNA-binding protein 12
O94973	AP-2 complex subunit alpha-2
Q15257	Serine/threonine-protein phosphatase 2A activator
Q9UPN9	E3 ubiquitin-protein ligase TRIM33
Q969Z0	Protein TBRG4
Q7L2E3	Putative ATP-dependent RNA helicase DHX30

Q13310	Polyadenylate-binding protein 4
Q15370	Transcription elongation factor B polypeptide 2
O14964	Hepatocyte growth factor-regulated tyrosine kinase substrate
P14649	Myosin light chain 6B
Q15007	Pre-mRNA-splicing regulator WTAP
O75391	Sperm-associated antigen 7
Q9C0C9	Ubiquitin-conjugating enzyme E2 O
Q9H3R5	Centromere protein H
Q9Y6A5	Transforming acidic coiled-coil-containing protein 3
P49589	Cysteine-tRNA ligase, cytoplasmic
Q14527	Helicase-like transcription factor
P60983	Glia maturation factor beta
Q9UNE7	E3 ubiquitin-protein ligase CHIP
O95400	CD2 antigen cytoplasmic tail-binding protein 2
P18615	Negative elongation factor E
O95628	CCR4-NOT transcription complex subunit 4
O43148	mRNA cap guanine-N7 methyltransferase
Q9BXF6	Rab11 family-interacting protein 5
Q04917	14-3-3 protein eta
Q01085	Nucleolysin TIAR
Q9H6T3	RNA polymerase II-associated protein 3
Q8NDV7	Trinucleotide repeat-containing gene 6A protein
Q5VT25	Serine/threonine-protein kinase MRCK alpha
Q86TB9	Protein PAT1 homolog 1
A8MPP1	Putative ATP-dependent RNA helicase DDX11-like protein 8
Q9H7Z7	Prostaglandin E synthase 2
O76070	Gamma-synuclein
Q5TBB1	Ribonuclease H2 subunit B
Q9UBE0	SUMO-activating enzyme subunit 1
Q15121	Astrocytic phosphoprotein PEA-15
P26639	Threonine-tRNA ligase, cytoplasmic
P55072	Transitional endoplasmic reticulum ATPase
P49959	Double-strand break repair protein MRE11A
Q96C19	EF-hand domain-containing protein D2
Q9NSD9	Phenylalanine-tRNA ligase beta subunit
P62736	Actin, aortic smooth muscle
Q9H0L4	Cleavage stimulation factor subunit 2 tau variant
Q86UU0	B-cell CLL/lymphoma 9-like protein
Q5VSL9	Striatin-interacting protein 1
P19838	Nuclear factor NF-kappa-B p105 subunit
Q8N3F8	MICAL-like protein 1
O00763	Acetyl-CoA carboxylase 2
O43251	RNA binding protein fox-1 homolog 2
Q9NQP4	Prefoldin subunit 4
Q8IY63	Angiomotin-like protein 1

P52907	F-actin-capping protein subunit alpha-1
P61204	ADP-ribosylation factor 3
Q9NZM3	Intersectin-2
Q92793	CREB-binding protein
P21108	Ribose-phosphate pyrophosphokinase 3
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1
Q7Z2K8	G protein-regulated inducer of neurite outgrowth 1
P07858	Cathepsin B
P43246	DNA mismatch repair protein Msh2
Q8IUH3	RNA-binding protein 45
Q8IX90	Spindle and kinetochore-associated protein 3
Q96CS2	HAUS augmin-like complex subunit 1
P10275	Androgen receptor
Q13564	NEDD8-activating enzyme E1 regulatory subunit
Q9NVE7	Pantothenate kinase 4
P25789	Proteasome subunit alpha type-4
P52789	Hexokinase-2
O60739	Eukaryotic translation initiation factor 1b
Q9P0K7	Ankycorbin
Q9Y570	Protein phosphatase methylesterase 1
P30154	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform
Q9H9A7	RecQ-mediated genome instability protein 1
Q8N7H5	RNA polymerase II-associated factor 1 homolog
P45983	Mitogen-activated protein kinase 8
Q9BY43	Charged multivesicular body protein 4a
Q9Y5S9	RNA-binding protein 8A
Q9UBT2	SUMO-activating enzyme subunit 2
Q92820	Gamma-glutamyl hydrolase
Q8IWZ3	Ankyrin repeat and KH domain-containing protein 1
P31689	DnaJ homolog subfamily A member 1
O94788	Retinal dehydrogenase 2
Q00013	55 kDa erythrocyte membrane protein
Q9NZ63	Uncharacterized protein C9orf78
Q96K17	Transcription factor BTF3 homolog 4
P08195	4F2 cell-surface antigen heavy chain
Q9BX40	Protein LSM14 homolog B
Q9H3N1	Thioredoxin-related transmembrane protein 1
Q16352	Alpha-internexin
Q8NHV4	Protein NEDD1
P50542	Peroxisomal targeting signal 1 receptor
O00213	Amyloid beta A4 precursor protein-binding family B member 1
P57678	Gem-associated protein 4
Q8NCM8	Cytoplasmic dynein 2 heavy chain 1
Q8NEZ5	F-box only protein 22

Q16555	Dihydropyrimidinase-related protein 2
Q8TE77	Protein phosphatase Slingshot homolog 3
P80303	Nucleobindin-2
O75496	Geminin
Q99576	TSC22 domain family protein 3
Q8N684	Cleavage and polyadenylation specificity factor subunit 7
P06737	Glycogen phosphorylase, liver form
P52948	Nuclear pore complex protein Nup98-Nup96
Q8IW35	Centrosomal protein of 97 kDa
P53999	Activated RNA polymerase II transcriptional coactivator p15
L0R819	Alternative protein ASNSD1
O75351	Vacuolar protein sorting-associated protein 4B
Q8N1F7	Nuclear pore complex protein Nup93
P48507	Glutamate-cysteine ligase regulatory subunit
Q9H467	CUE domain-containing protein 2
P30566	Adenylosuccinate lyase
P06753	Isoform 2 of Tropomyosin alpha-3 chain
Q02543	60S ribosomal protein L18a
Q15185	Prostaglandin E synthase 3

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

C.1 MATERIALS AND METHODS

C.1.1 Sequence design of ES7 of *Candida albicans*

ES7 rRNA of *C. albicans* ($ES7_{CA}$) is composed of 209 nucleotides. Addition of the T7 promoter, restriction sites (EcoRI-5'end and HindIII-3'end) and stability bases provides a sequence with a total of 247 nucleotides. The full-length DNA sequence was divided into six partially complementary DNA oligonucleotides forming a tile and assembled together by recursive PCR (221).

Table C.1 DNA oligonucleotides used to construct the full-length DNA of $ES7_{CA}$ by recursive PCR.

Gene	Oligo	Sequence (5' to 3')[a]
C. albicans ES7	1	GTGGGAATTC*TAATACGACTCACTATAGGG*CTTGAGATCAGA CTTGGTATTTTG
	2	CAGCGGCCGCCCCCGAGAGAGCAGCATGCAAAATACCAAG TCTGATCTCAAGCC
	3	CGGGGGCGGCCGCTGCGGTTTACCGGGCCAGCATCGGTTT GGAGCGGCAGGATAATG
	4	CACACAGCAGAAGCCGTGCCACATTCCTCCGCCATTATCCT GCCGCTCCAAACC
	5	GGCACGGCTTCTGCTGTGTGTTATAGCCTCTGACGATGCTG CCAGCCTAGACCGAGGACTG
	6	CACCAAGCTTGACTTAAGATCATTATGCCAACATCCTAGGT TAAAAACCGCAGTCCTCGGTCTAGGCTGG

a) Restriction sites are underlined and the T7 promoter sequence is in italics.

Table C.2 Primers used to build $ES7_{CA}$ by recursive PCR.

Gene	Primer	Sequence (5' to 3')
C. albicans ES7	Forward	GTGGGAATTCTAATACGACTCACTATAG
	Reverse	CACCAAGCTTGACTTAAGATCATTATGC

C.1.2 Recursive PCR

Oligos and primers were dissolved in nuclease-free water to working stock solutions of 2.5 µM for oligos and 20 µM for primers (strand). Final PCR conditions consisted of 0.25 mM of each dNTP (New England BioLabs), 1x Complete Long PCR Buffer (Ambion), 500 nM each, of forward and reverse primers, 3U of *Pfu* DNA polymerase, and 50 nM of each DNA oligomer in 50 µL reactions. Reaction mixtures were heated to 95 °C for 5 min during an initial denaturation step, and cycled 30 times through denaturation (98 °C for 30 sec), annealing (52.0 °C for 1 min) and elongation (72 °C for 1 min). The final elongation step was performed at 72 °C for 10 min. The target amplification products were purified by preparative gel electrophoresis using standard protocols.

Table C.3 DNA sequence encoding $ES7_{CA}$

Gene	Sequence (5' to 3')[a]
C. albicans ES7	GAATTC*TAATACGACTCACTATAG*GGCTTGAGATCAGACTTGGT ATTTTGCATGCTGCTCTCTCGGGGGCGGCCGCTGCGGTTTACC GGGCCAGCATCGGTTTGGAGCGGCAGGATAATGGCGGAGGA ATGTGGCACGGCTTCTGCTGTGTGTTATAGCCTCTGACGATGC TGCCAGCCTAGACCGAGGACTGCGGTTTTTAACCTAGGATGT TGGCATAATGATCTTAAGTCAAGCTT

Table C.4 Primers used to insert the SHAPE structural cassette.

Gene	Primer	Sequence (5' to 3')
C. albicans ES7	Forward	AATCGGGCTTCGGTCCGGTTCAAGCTTGGCGTAAT CATG
	Reverse	TGGATCCGGCGAACCGGATCGAGACTTAAGATCAT TATGCCAAC

C.1.3 Docking of F-neo and Neo to $ES7_{CA}$

Docking of F-neo and Neo to $ES7_{CA}$ were performed using Autodock Vina 1.1.2 (279). Default parameters were used with the exception of 'binding modes', which was set to the maximum allowed value of 20. $ES7_{CA}$ was held rigid. Docking was performed

repeatedly, with different random seed numbers, to observe consensus docking sites. Approximately 12 distinct binding sites were observed. All rotatable bonds of F-neo and Neo were not restrained. AutoDock Tools 1.5.6 was used to prepare receptor and ligand molecules for docking with AutoDock Vina (280).

C.2 RESULTS AND DISCUSSION

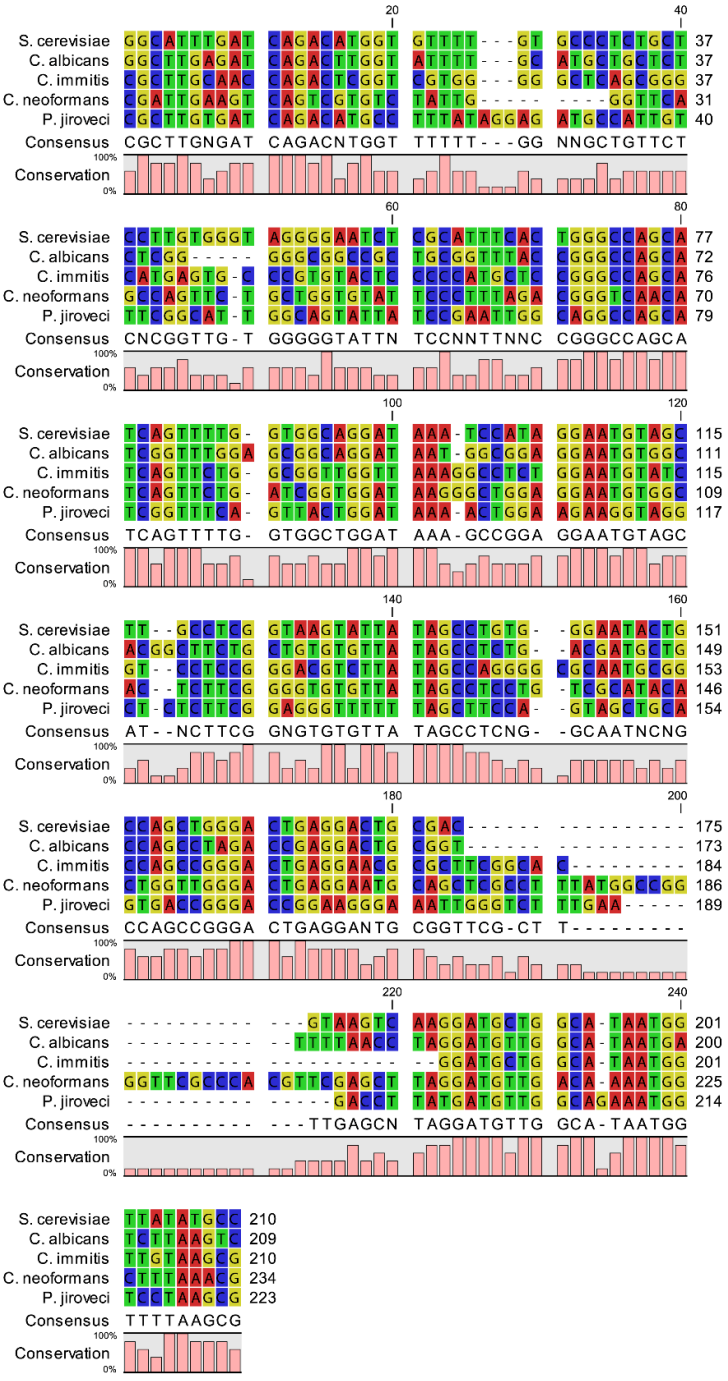


Figure C.1 Sequence alignment of ES7 rDNAs from *S. cerevisiae*, *C. albicans*, *C. immitis*, *C. neoformans* and *P. jiroveci*. Alignment performed with MAFFT using the L-INS-i refinement method (http://mafft.cbrc.jp/alignment/server/).

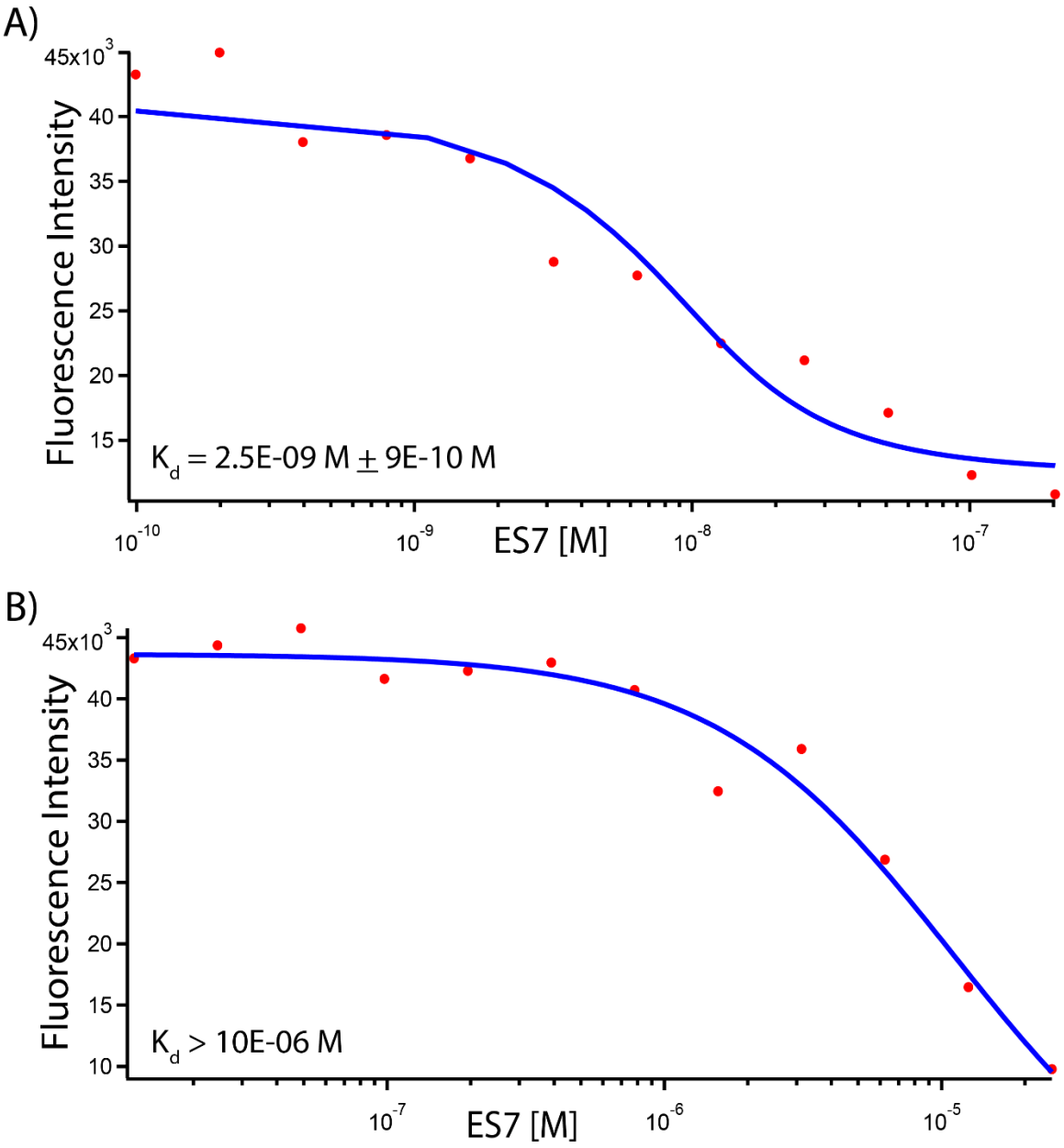


Figure C.2 Binding assays of ES7 rRNAs with F-neo. Fluorescent binding assay of A) $ES7_{CA}$ and B) $ES7_{HS}$ with F-neo were performed to determine the dissociation constant, K_d. The K_d for $ES7_{CA}$ and Neo was estimated by fitting as 2.5 x 10^{-9} M. The K_d for $ES7_{HS}$ and Neo was estimated to be greater than 10 µM.

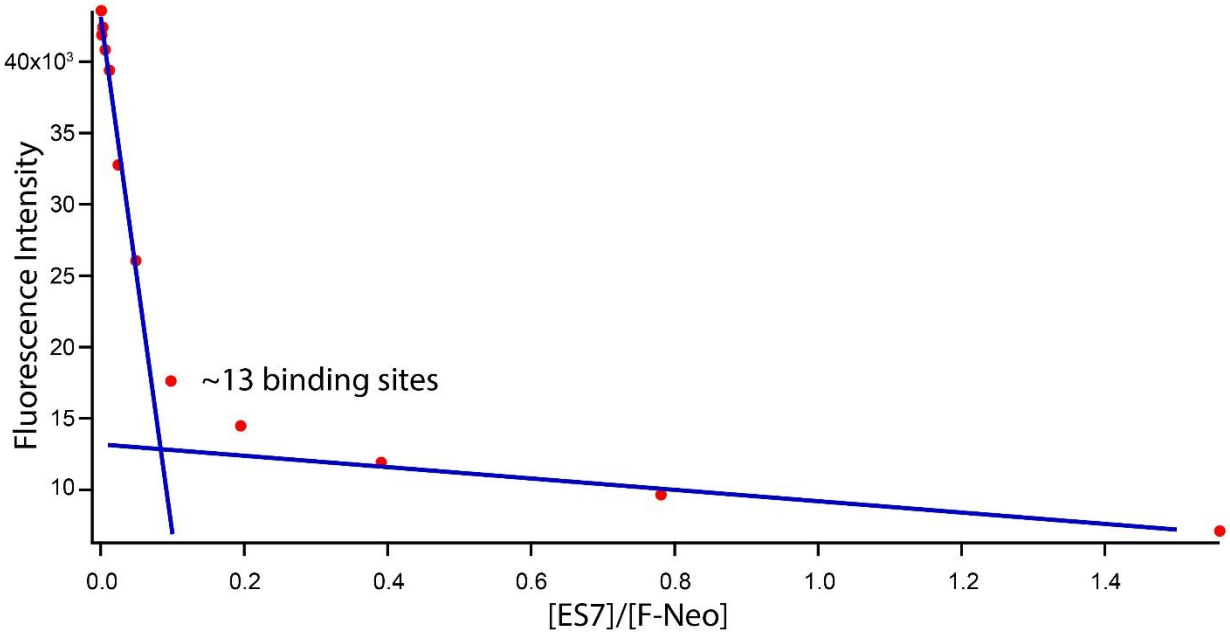


Figure C.3 Saturation binding plot of F-neo and $ES7_{CA}$. Fluorescence titration of F-neo with $ES7_{CA}$ was performed by maintaining F-neo constant at 10 nM and increasing the concentration of $ES7_{CA}$.

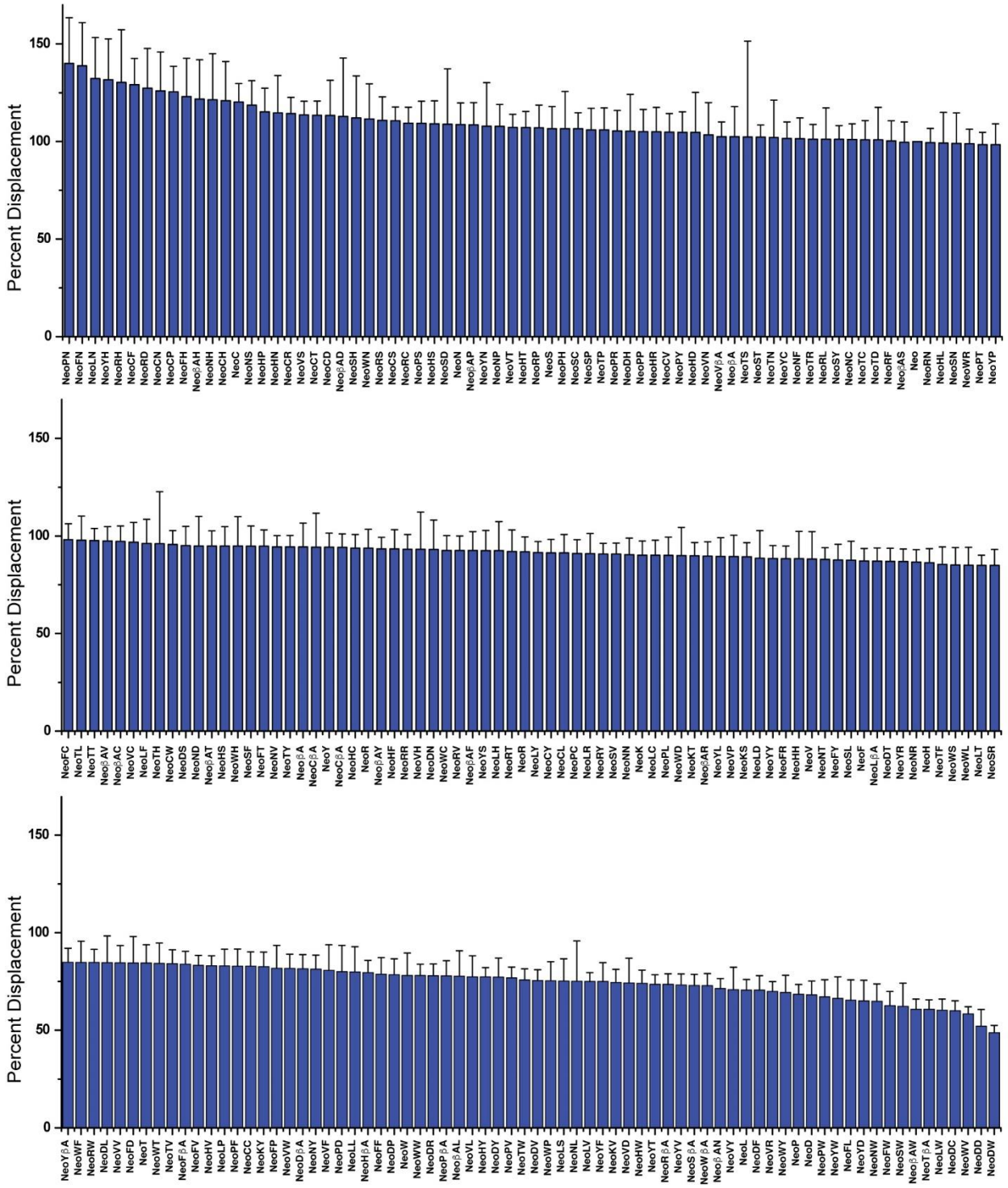


Figure C.4 High throughput screening results of the displacement of F-neo from $ES7_{CA}$ by members of a 215 peptidyl aminosugar (PA) library. The percent of displacement of F-neo by PAs was obtained as stated in the materials and methods. Every experiment was performed in duplicates. Standard deviation for each conjugate is shown with error bars.

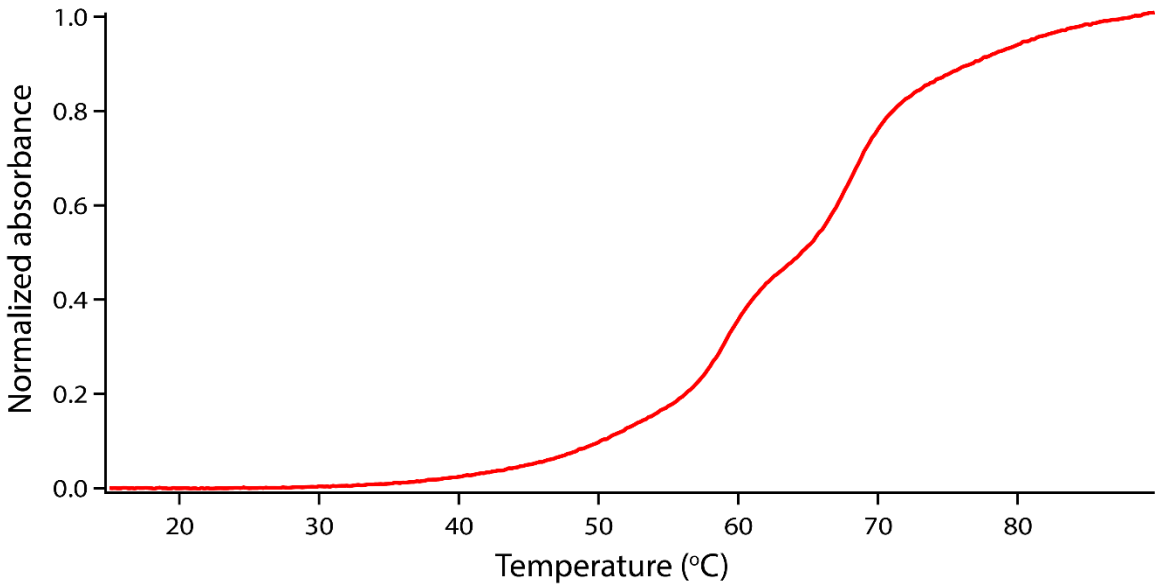


Figure C.5 Normalized, baseline corrected thermal melting of $ES7_{CA}$ rRNA at 260 nm. Multiple melting transitions are observed in the melting process.

Table C.5 Estimation of the melting temperatures of $ES7_{CA}$ in the presence of Neo. The melting profile, which is the derivative of the fraction unfolded with respect to temperature was computed. The profile was fit to a non-sequential, independent transition model with the program Igor Pro. The melting temperature ($T_{m,n}$) was estimated for each transition, n. Molar ratios of Neo to $ES7_{CA}$ are indicated.

Neo:ES7 Molar Ratio / **$T_{m,n}$ (°C)**	**0**	**10**	**100**	**500**	**1000**
1	52.3	51.6	53.2	60	61.5
2	59.3	59.4	61.7	66.9	68.4
3	68.1	68.0	70.4	75.4	76.8
4	77.6	78.4	80.7	85.1	84.6

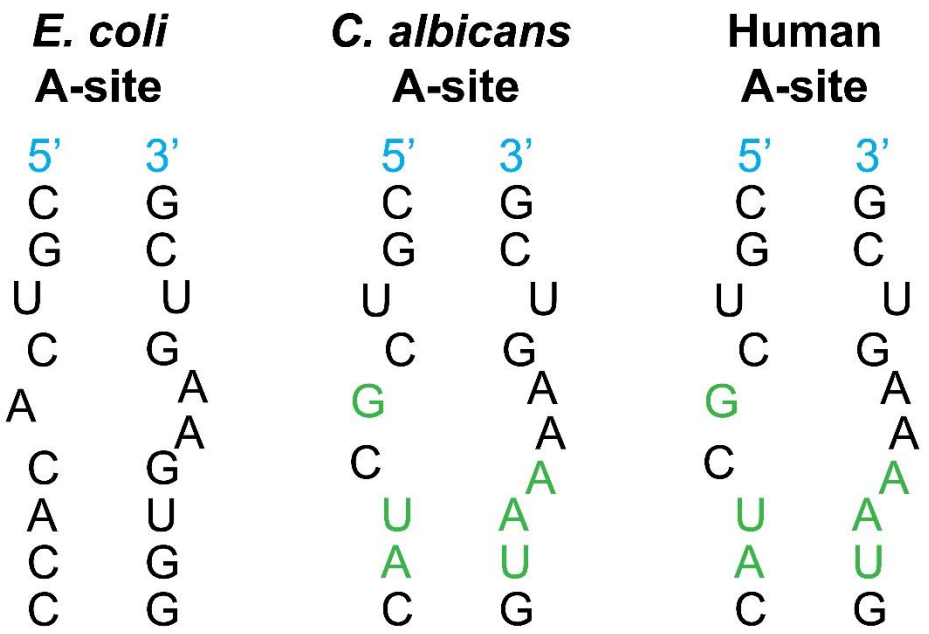


Figure C.6 Secondary structure of the A site of *E. coli*, *C. albicans* and human. Nucleotides in green highlight the difference in sequence between the A site of *E. coli* and the A site of *C. albicans* and human.

REFERENCES

[1] Doudna, J. A.; Rath, V. L.,"Structure and function of the eukaryotic ribosome: the next frontier" *Cell* **2002**, *109*, 153.

[2] Michot, B.; Bachellerie, J. P.,"Comparisons of large subunit rRNAs reveal some eukaryote-specific elements of secondary structure" *Biochimie* **1987**, *69*, 11.

[3] Gerbi, S. A., "Expansion segments: Regions of variable size that interrupt the universal core secondary structure of ribosomal RNA" In *Ribosomal RNA—Structure, evolution, processing, and function in protein synthesis*, Zimmermann, R. A.; Dahlberg, A. E., Eds. CRC Press: Boca Raton, FL, **1996**, 71.

[4] Petrov, A. S.; Bernier, C. R.; Hsiao, C.; Norris, A. M.; Kovacs, N. A.; Waterbury, C. C.; Stepanov, V. G.; Harvey, S. C.; Fox, G. E.; Wartell, R. M.; Hud, N. V.; Williams, L. D.,"Evolution of the ribosome at atomic resolution" *Proceedings of the National Academy of Sciences* **2014**, *111*, 10251.

[5] Agrawal, R. K.; Sharma, M. R.; Yassin, A.; Lahiri, I.; Spremulli, L. L., "Structure and function of organellar ribosomes as revealed by cryo-EM" In *Ribosomes Structure, Function and Dynamics*, Rodnina, M. V.; Wintermeyer, W.; Green, R., Eds. SpringerWien: New York, **2011**, 83.

[6] Noller, H. F.; Hoffarth, V.; Zimniak, L.,"Unusual resistance of peptidyl transferase to protein extraction procedures" *Science* **1992**, *256*, 1416.

[7] Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A.,"The structural basis of ribosome activity in peptide bond synthesis" *Science* **2000**, *289*, 920.

[8] Stark, H.; Mueller, F.; Orlova, E. V.; Schatz, M.; Dube, P.; Erdemir, T.; Zemlin, F.; Brimacombe, R.; van Heel, M.,"The 70S Escherichia coli ribosome at 23 A resolution: fitting the ribosomal RNA" *Structure* **1995**, *3*, 815.

[9] Gabashvili, I. S.; Agrawal, R. K.; Spahn, C. M.; Grassucci, R. A.; Svergun, D. I.; Frank, J.; Penczek, P.,"Solution structure of the E. coli 70S ribosome at 11.5 A resolution" *Cell* **2000**, *100*, 537.

[10] Gao, H.; Sengupta, J.; Valle, M.; Korostelev, A.; Eswar, N.; Stagg, S. M.; Van Roey, P.; Agrawal, R. K.; Harvey, S. C.; Sali, A.; Chapman, M. S.; Frank, J.,"Study of the structural dynamics of the E coli 70S ribosome using real-space refinement" *Cell* **2003**, *113*, 789.

[11] Mitra, K.; Schaffitzel, C.; Shaikh, T.; Tama, F.; Jenni, S.; Brooks, C. L., 3rd; Ban, N.; Frank, J.,"Structure of the E. coli protein-conducting channel bound to a translating ribosome" *Nature* **2005**, *438*, 318.

[12] Estrozi, L. F.; Boehringer, D.; Shan, S. O.; Ban, N.; Schaffitzel, C.,"Cryo-EM structure of the E. coli translating ribosome in complex with SRP and its receptor" *Nature structural & molecular biology* **2011**, *18*, 88.

[13] Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H.,"Structures of the bacterial ribosome at 3.5 Å resolution" *Science* **2005**, *310*, 827.

[14] Berk, V.; Zhang, W.; Pai, R. D.; Cate, J. H.,"Structural basis for mRNA and tRNA positioning on the ribosome" *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 15830.

[15] Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H.; Noller, H. F.,"Crystal structure of the ribosome at 5.5 Å resolution" *Science* **2001**, *292*, 883.
[16] Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A.,"The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution" *Science* **2000**, *289*, 905.
[17] Ben-Shem, A.; de Loubresse, N. G.; Melnikov, S.; Jenner, L.; Yusupova, G.; Yusupov, M.,"The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution" *Science* **2011**, *334*, 1524.
[18] Armache, J. P.; Jarasch, A.; Anger, A. M.; Villa, E.; Becker, T.; Bhushan, S.; Jossinet, F.; Habeck, M.; Dindar, G.; Franckenberg, S.; Marquez, V.; Mielke, T.; Thomm, M.; Berninghausen, O.; Beatrix, B.; Soding, J.; Westhof, E.; Wilson, D. N.; Beckmann, R.,"Cryo-EM structure and rRNA model of a translating eukaryotic 80S ribosome at 5.5 Å resolution" *Proc Natl Acad Sci U S A* **2010**, *107*, 19748.
[19] Anger, A. M.; Armache, J. P.; Berninghausen, O.; Habeck, M.; Subklewe, M.; Wilson, D. N.; Beckmann, R.,"Structures of the human and Drosophila 80S ribosome" *Nature* **2013**, *497*, 80.
[20] Khatter, H.; Myasnikov, A. G.; Natchiar, S. K.; Klaholz, B. P.,"Structure of the human 80S ribosome" *Nature* **2015**, *advance online publication*.
[21] Melnikov, S.; Ben-Shem, A.; Garreau de Loubresse, N.; Jenner, L.; Yusupova, G.; Yusupov, M.,"One core, two shells: bacterial and eukaryotic ribosomes" *Nature structural & molecular biology* **2012**, *19*, 560.
[22] Hsiao, C.; Mohan, S.; Kalahar, B. K.; Williams, L. D.,"Peeling the onion: ribosomes are ancient molecular fossils" *Mol Biol Evol* **2009**, *26*, 2415.
[23] Chan, Y. L.; Olvera, J.; Wool, I. G.,"The structure of rat 28S ribosomal ribonucleic acid inferred from the sequence of nucleotides in a gene" *Nucleic Acids Res* **1983**, *11*, 7819.
[24] Petrov, A. S.; Gulen, B.; Norris, A. M.; Kovacs, N. A.; Bernier, C. R.; Lanier, K. A.; Fox, G. E.; Harvey, S. C.; Wartell, R. M.; Hud, N. V.; Williams, L. D.,"History of the ribosome and the origin of translation" *Proc Natl Acad Sci U S A* **2015**, *112*, 15396.
[25] Ware, V. C.; Tague, B. W.; Clark, C. G.; Gourse, R. L.; Brand, R. C.; Gerbi, S. A.,"Sequence analysis of 28S ribosomal DNA from the amphibian Xenopus laevis" *Nucleic Acids Res* **1983**, *11*, 7795.
[26] Rubtsov, P. M.; Musakhanov, M. M.; Zakharyev, V. M.; Krayev, A. S.; Skryabin, K. G.; Bayev, A. A.,"The structure of the yeast ribosomal RNA genes. I. The complete nucleotide sequence of the 18S ribosomal RNA gene from Saccharomyces cerevisiae" *Nucleic Acids Res* **1980**, *8*, 5779.
[27] Salim, M.; Maden, B. E.,"Nucleotide sequence of Xenopus laevis 18S ribosomal RNA inferred from gene sequence" *Nature* **1981**, *291*, 205.
[28] Veldman, G. M.; Klootwijk, J.; de Regt, V. C.; Planta, R. J.; Branlant, C.; Krol, A.; Ebel, J. P.,"The primary and secondary structure of yeast 26S rRNA" *Nucleic Acids Res* **1981**, *9*, 6935.
[29] Branlant, C.; Krol, A.; Machatt, M. A.; Pouyet, J.; Ebel, J. P.; Edwards, K.; Kossel, H.,"Primary and secondary structures of escherichia-coli mre-600-23s

ribosomal-rna - comparison with models of secondary structure for maize chloroplast 23s ribosomal-rna and for large portions of mouse and human 16s mitochondrial ribosomal-rnas" *Nucleic Acids Research* **1981**, *9*, 4303.

[30] Glotz, C.; Zwieb, C.; Brimacombe, R.; Edwards, K.; Kossel, H.,"Secondary structure of the large subunit ribosomal-rna from escherichia-coli, zea-mays chloroplast, and human and mouse mitochondrial ribosomes" *Nucleic Acids Research* **1981**, *9*, 3287.

[31] Nei, M., *Molecular evolutionary genetics*. Columbia University Press: New York, **1987**.

[32] Hassouna, N.; Michot, B.; Bachellerie, J. P.,"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 Res* **1984**, *12*, 3563.

[33] Clark, C. G.; Tague, B. W.; Ware, V. C.; Gerbi, S. A.,"Xenopus laevis 28S ribosomal RNA: a secondary structure model and its evolutionary and functional implications" *Nucleic Acids Res* **1984**, *12*, 6197.

[34] Raue, H. A.; Klootwijk, J.; Musters, W.,"EVOLUTIONARY CONSERVATION OF STRUCTURE AND FUNCTION OF HIGH MOLECULAR-WEIGHT RIBOSOMAL-RNA" *Progress in Biophysics & Molecular Biology* **1988**, *51*, 77.

[35] Wakeman, J. A.; Maden, B. E.,"28 S ribosomal RNA in vertebrates. Locations of large-scale features revealed by electron microscopy in relation to other features of the sequences" *The Biochemical journal* **1989**, *258*, 49.

[36] Galtier, N.; Lobry, J. R.,"Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes" *J Mol Evol* **1997**, *44*, 632.

[37] Galtier, N.; Tourasse, N.; Gouy, M.,"A nonhyperthermophilic common ancestor to extant life forms" *Science* **1999**, *283*, 220.

[38] Gonzalez, I. L.; Gorski, J. L.; Campen, T. J.; Dorney, D. J.; Erickson, J. M.; Sylvester, J. E.; Schmickel, R. D.,"Variation among human 28S ribosomal RNA genes" *Proc Natl Acad Sci U S A* **1985**, *82*, 7666.

[39] Gorski, J. L.; Gonzalez, I. L.; Schmickel, R. D.,"The secondary structure of human 28S rRNA: the structure and evolution of a mosaic rRNA gene" *J Mol Evol* **1987**, *24*, 236.

[40] Gonzalez, I. L.; Chambers, C.; Gorski, J. L.; Stambolian, D.; Schmickel, R. D.; Sylvester, J. E.,"Sequence and structure correlation of human ribosomal transcribed spacers" *J Mol Biol* **1990**, *212*, 27.

[41] Hancock, J. M.; Dover, G. A.,"'Compensatory slippage' in the evolution of ribosomal RNA genes" *Nucleic Acids Res* **1990**, *18*, 5949.

[42] Tautz, D.; Trick, M.; Dover, G. A.,"Cryptic simplicity in DNA is a major source of genetic variation" *Nature* **1986**, *322*, 652.

[43] Tautz, D.; Hancock, J. M.; Webb, D. A.; Tautz, C.; Dover, G. A.,"Complete sequences of the rRNA genes of Drosophila melanogaster" *Mol Biol Evol* **1988**, *5*, 366.

[44] Nunn, G. B.; Theisen, B. F.; Christensen, B.; Arctander, P.,"Simplicity-correlated size growth of the nuclear 28S ribosomal RNA D3 expansion segment in the crustacean order Isopoda" *J Mol Evol* **1996**, *42*, 211.

[45] Hancock, J. M.; Dover, G. A.,"Molecular coevolution among cryptically simple expansion segments of eukaryotic 26S/28S rRNAs" *Mol Biol Evol* **1988**, *5*, 377.
[46] Hancock, J. M.; Tautz, D.; Dover, G. A.,"Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of Drosophila melanogaster" *Mol Biol Evol* **1988**, *5*, 393.
[47] Gorab, E.; Garcia de Lacoba, M.; Botella, L. M.,"Structural constraints in expansion segments from a midge 26S rDNA" *J Mol Evol* **1995**, *41*, 1016.
[48] Schnare, M. N.; Heinonen, T. Y.; Young, P. G.; Gray, M. W.,"A discontinuous small subunit ribosomal RNA in Tetrahymena pyriformis mitochondria" *J Biol Chem* **1986**, *261*, 5187.
[49] Gray, M. W.; Boer, P. H.; Collings, J. C.; Heinonen, T. Y. K.; Spencer, D. F.; Schnare, M. N., "Ribosomal RNA genes in pieces" In *Highlights of Modern Biochemistry. Volume 1: Proceedings of the 14th International Congress of Biochemistry*, Kotyk, A.; Skoda, J.; Paces, V.; Kostka, V., Eds. CRC Press: **1989**, 521.
[50] Han, H.; Schepartz, A.; Pellegrini, M.; Dervan, P. B.,"Mapping RNA regions in eukaryotic ribosomes that are accessible to methidiumpropyl-EDTA.Fe(II) and EDTA.Fe(II)" *Biochemistry* **1994**, *33*, 9831.
[51] Schnare, M. N.; Damberger, S. H.; Gray, M. W.; Gutell, R. R.,"Comprehensive comparison of structural characteristics in eukaryotic cytoplasmic large subunit (23 S-like) ribosomal RNA" *J Mol Biol* **1996**, *256*, 701.
[52] Dube, P.; Bacher, G.; Stark, H.; Mueller, F.; Zemlin, F.; van Heel, M.; Brimacombe, R.,"Correlation of the expansion segments in mammalian rRNA with the fine structure of the 80 S ribosome; a cryoelectron microscopic reconstruction of the rabbit reticulocyte ribosome at 21 A resolution" *J Mol Biol* **1998**, *279*, 403.
[53] Yokoyama, T.; Suzuki, T.,"Ribosomal RNAs are tolerant toward genetic insertions: evolutionary origin of the expansion segments" *Nucleic Acids Res* **2008**, *36*, 3539.
[54] Morgan, D. G.; Menetret, J. F.; Radermacher, M.; Neuhof, A.; Akey, I. V.; Rapoport, T. A.; Akey, C. W.,"A comparison of the yeast and rabbit 80 S ribosome reveals the topology of the nascent chain exit tunnel, inter-subunit bridges and mammalian rRNA expansion segments" *J Mol Biol* **2000**, *301*, 301.
[55] Rabl, J.; Leibundgut, M.; Ataide, S. F.; Haag, A.; Ban, N.,"Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1" *Science* **2011**, *331*, 730.
[56] Klinge, S.; Voigts-Hoffmann, F.; Leibundgut, M.; Arpagaus, S.; Ban, N.,"Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6" *Science* **2011**, *334*, 941.
[57] Ban, N.; Beckmann, R.; Cate, J. H.; Dinman, J. D.; Dragon, F.; Ellis, S. R.; Lafontaine, D. L.; Lindahl, L.; Liljas, A.; Lipton, J. M.; McAlear, M. A.; Moore, P. B.; Noller, H. F.; Ortega, J.; Panse, V. G.; Ramakrishnan, V.; Spahn, C. M.; Steitz, T. A.; Tchorzewski, M.; Tollervey, D.; Warren, A. J.; Williamson, J. R.; Wilson, D.; Yonath, A.; Yusupov, M.,"A new system for naming ribosomal proteins" *Curr Opin Struct Biol* **2014**, *24*, 165.

[58] Sweeney, R.; Yao, M. C.,"Identifying functional regions of rRNA by insertion mutagenesis and complete gene replacement in Tetrahymena thermophila" *EMBO J* **1989**, *8*, 933.
[59] Yao, M. C.; Yao, C. H.,"Accurate processing and amplification of cloned germ line copies of ribosomal DNA injected into developing nuclei of Tetrahymena thermophila" *Mol Cell Biol* **1989**, *9*, 1092.
[60] Sweeney, R.; Chen, L.; Yao, M.-C.,"An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence" *Molecular and cellular biology* **1994**, *14*, 4203.
[61] Musters, W.; Venema, J.; van der Linden, G.; van Heerikhuizen, H.; Klootwijk, J.; Planta, R. J.,"A system for the analysis of yeast ribosomal DNA mutations" *Mol Cell Biol* **1989**, *9*, 551.
[62] Musters, W.; Boon, K.; van der Sande, C. A.; van Heerikhuizen, H.; Planta, R. J.,"Functional analysis of transcribed spacers of yeast ribosomal DNA" *EMBO J* **1990**, *9*, 3989.
[63] Venema, J.; Dirks-Mulder, A.; Faber, A. W.; Raue, H. A.,"Development and application of an in vivo system to study yeast ribosomal RNA biogenesis and function" *Yeast* **1995**, *11*, 145.
[64] Musters, W.; Concalves, P. M.; Boon, K.; Raue, H. A.; van Heerikhuizen, H.; Planta, R. J.,"The conserved GTPase center and variable region V9 from Saccharomyces cerevisiae 26S rRNA can be replaced by their equivalents from other prokaryotes or eukaryotes without detectable loss of ribosomal function" *Proc Natl Acad Sci U S A* **1991**, *88*, 1469.
[65] Kooi, E. A. "The role and nature of the interaction between ribosomal protein L25 and 26S rRNA in yeast ribosomes" PhD Thesis, Vrije Universitat, Amsterdam **1995**.
[66] van Nues, R. W.; Venema, J.; Planta, R. J.; Raue, H. A.,"Variable region V1 of Saccharomyces cerevisiae 18S rRNA participates in biogenesis and function of the small ribosomal subunit" *Chromosoma* **1997**, *105*, 523.
[67] Jeeninga, R. E.; Van Delft, Y.; de Graaff-Vincent, M.; Dirks-Mulder, A.; Venema, J.; Raue, H. A.,"Variable regions V13 and V3 of Saccharomyces cerevisiae contain structural features essential for normal biogenesis and stability of 5.8S and 25S rRNA" *RNA-Publ. RNA Soc.* **1997**, *3*, 476.
[68] Gao, H.; Ayub, M. J.; Levin, M. J.; Frank, J.,"The structure of the 80S ribosome from Trypanosoma cruzi reveals unique rRNA components" *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 10206.
[69] Beckmann, R.; Spahn, C. M.; Eswar, N.; Helmers, J.; Penczek, P. a.; Sali, a.; Frank, J.; Blobel, G.,"Architecture of the protein-conducting channel associated with the translating 80S ribosome" *Cell* **2001**, *107*, 361.
[70] Nilsson, J.; Sengupta, J.; Gursky, R.; Nissen, P.; Frank, J.,"Comparison of fungal 80 S ribosomes by cryo-EM reveals diversity in structure and conformation of rRNA expansion segments" *J Mol Biol* **2007**, *369*, 429.
[71] Diaconu, M.; Kothe, U.; Schlunzen, F.; Fischer, N.; Harms, J. M.; Tonevitsky, A. G.; Stark, H.; Rodnina, M. V.; Wahl, M. C.,"Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation" *Cell* **2005**, *121*, 991.

[72] Srivastava, S.; Verschoor, A.; Frank, J.,"Eukaryotic initiation factor 3 does not prevent association through physical blockage of the ribosomal subunit-subunit interface" *J Mol Biol* **1992**, *226*, 301.
[73] Siridechadilok, B.; Fraser, C. S.; Hall, R. J.; Doudna, J. A.; Nogales, E.,"Structural roles for human translation factor eIF3 in initiation of protein synthesis" *Science* **2005**, *310*, 1513.
[74] Yu, Y.; Abaeva, I. S.; Marintchev, A.; Pestova, T. V.; Hellen, C. U.,"Common conformational changes induced in type 2 picornavirus IRESs by cognate trans-acting factors" *Nucleic Acids Res* **2011**, *39*, 4851.
[75] Fleming, A.,"On the bacterial action of a culture of a penicillium, with special reference to their use in the isolation of *B. influenzae*" *British Journal of Experimental Pathology* **1929**, *10*, 226.
[76] Abraham, E. P.; Chain, E.,"An Enzyme from Bacteria able to Destroy Penicillin" *Nature* **1940**, *146*, 837.
[77] Hampton, T.,"Report reveals scope of US antibiotic resistance threat" *JAMA* **2013**, *310*, 1661.
[78] Center for Disease Control and Prevention. "Antibiotic / Antimicrobial Resistance"http://www.cdc.gov/drugresistance/ (accessed November 16, 2016).
[79] Walsh, C.,"Molecular mechanisms that confer antibacterial drug resistance" *Nature* **2000**, *406*, 775.
[80] Melnyk, A. H.; Wong, A.; Kassen, R.,"The fitness costs of antibiotic resistance mutations" *Evol Appl* **2015**, *8*, 273.
[81] Andersson, D. I.,"Persistence of antibiotic resistant bacteria" *Curr Opin Microbiol* **2003**, *6*, 452.
[82] Levin, B. R.; Lipsitch, M.; Perrot, V.; Schrag, S.; Antia, R.; Simonsen, L.; Walker, N. M.; Stewart, F. M.,"The population genetics of antibiotic resistance" *Clin Infect Dis* **1997**, *24 Suppl 1*, S9.
[83] Andersson, D. I.; Levin, B. R.,"The biological cost of antibiotic resistance" *Curr Opin Microbiol* **1999**, *2*, 489.
[84] Andersson, D. I.; Hughes, D.,"Antibiotic resistance and its cost: is it possible to reverse resistance?" *Nat Rev Microbiol* **2010**, *8*, 260.
[85] Blair, J. M.; Webber, M. A.; Baylay, A. J.; Ogbolu, D. O.; Piddock, L. J.,"Molecular mechanisms of antibiotic resistance" *Nat Rev Microbiol* **2015**, *13*, 42.
[86] Powers, J. H.,"Antimicrobial drug development--the past, the present, and the future" *Clin Microbiol Infect* **2004**, *10 Suppl 4*, 23.
[87] Lewis, K.,"Platforms for antibiotic discovery" *Nat Rev Drug Discov* **2013**, *12*, 371.
[88] Coates, A. R.; Hu, Y.,"Novel approaches to developing new antibiotics for bacterial infections" *British journal of pharmacology* **2007**, *152*, 1147.
[89] Payne, D. J.; Gwynn, M. N.; Holmes, D. J.; Pompliano, D. L.,"Drugs for bad bugs: confronting the challenges of antibacterial discovery" *Nat Rev Drug Discov* **2007**, *6*, 29.
[90] Sánchez, S. A.; Demain, A. L., *Antibiotics : current innovations and future trends*. Caister Academic Press: Norfolk, U.K., **2015**.

[91] Fernandes, P.,"Antibacterial discovery and development--the failure of success?" *Nat Biotechnol* **2006**, *24*, 1497.
[92] Fischbach, M. A.,"Combination therapies for combating antimicrobial resistance" *Curr Opin Microbiol* **2011**, *14*, 519.
[93] Harvey, A. L.,"Natural products in drug discovery" *Drug Discov Today* **2008**, *13*, 894.
[94] Marris, E.,"Marine natural products: drugs from the deep" *Nature* **2006**, *443*, 904.
[95] Kwon, H. C.; Kauffman, C. A.; Jensen, P. R.; Fenical, W.,"Marinomycins A-D, antitumor-antibiotics of a new structure class from a marine actinomycete of the recently discovered genus "marinispora"" *J Am Chem Soc* **2006**, *128*, 1622.
[96] Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.; Murgolo, N.; Barbieri, C. M.; Mann, P. A.; Carr, D.; Xia, E.; Zuck, P.; Riley, D.; Painter, R. E.; Walker, S. S.; Sherborne, B.; de Jesus, R.; Pan, W.; Plotkin, M. A.; Wu, J.; Rindgen, D.; Cummings, J.; Garlisi, C. G.; Zhang, R.; Sheth, P. R.; Gill, C. J.; Tang, H.; Roemer, T.,"Selective small-molecule inhibition of an RNA structural element" *Nature* **2015**, *526*, 672.
[97] Oxford, A. E.; Raistrick, H.; Simonart, P.,"Studies in the biochemistry of micro-organisms: Griseofulvin, C(17)H(17)O(6)Cl, a metabolic product of Penicillium griseo-fulvum Dierckx" *The Biochemical journal* **1939**, *33*, 240.
[98] White, T. C.; Marr, K. A.; Bowden, R. A.,"Clinical, cellular, and molecular factors that contribute to antifungal drug resistance" *Clin Microbiol Rev* **1998**, *11*, 382.
[99] Kontoyiannis, D. P.; Lewis, R. E.,"Antifungal drug resistance of pathogenic fungi" *Lancet* **2002**, *359*, 1135.
[100] Kanafani, Z. A.; Perfect, J. R.,"Antimicrobial resistance: resistance to antifungal agents: mechanisms and clinical impact" *Clin Infect Dis* **2008**, *46*, 120.
[101] Ghannoum, M. A.; Rice, L. B.,"Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance" *Clin Microbiol Rev* **1999**, *12*, 501.
[102] Odds, F. C.; Brown, A. J.; Gow, N. A.,"Antifungal agents: mechanisms of action" *Trends Microbiol* **2003**, *11*, 272.
[103] Dominguez, J. M.; Kelly, V. A.; Kinsman, O. S.; Marriott, M. S.; Gomez de las Heras, F.; Martin, J. J.,"Sordarins: A new class of antifungals with selective inhibition of the protein synthesis elongation cycle in yeasts" *Antimicrob Agents Chemother* **1998**, *42*, 2274.
[104] Roemer, T.; Xu, D.; Singh, S. B.; Parish, C. A.; Harris, G.; Wang, H.; Davies, J. E.; Bills, G. F.,"Confronting the challenges of natural product-based antifungal discovery" *Chem Biol* **2011**, *18*, 148.
[105] De Backer, M. D.; Nelissen, B.; Logghe, M.; Viaene, J.; Loonen, I.; Vandoninck, S.; de Hoogt, R.; Dewaele, S.; Simons, F. A.; Verhasselt, P.; Vanhoof, G.; Contreras, R.; Luyten, W. H.,"An antisense-based functional genomics approach for identification of genes critical for growth of Candida albicans" *Nat Biotechnol* **2001**, *19*, 235.

[106] Uhl, M. A.; Biery, M.; Craig, N.; Johnson, A. D.,"Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans" *EMBO J* **2003**, *22*, 2668.
[107] Roemer, T.; Jiang, B.; Davison, J.; Ketela, T.; Veillette, K.; Breton, A.; Tandia, F.; Linteau, A.; Sillaots, S.; Marta, C.; Martel, N.; Veronneau, S.; Lemieux, S.; Kauffman, S.; Becker, J.; Storms, R.; Boone, C.; Bussey, H.,"Large-scale essential gene identification in Candida albicans and applications to antifungal drug discovery" *Mol Microbiol* **2003**, *50*, 167.
[108] Becker, J. M.; Kauffman, S. J.; Hauser, M.; Huang, L.; Lin, M.; Sillaots, S.; Jiang, B.; Xu, D.; Roemer, T.,"Pathway analysis of Candida albicans survival and virulence determinants in a murine infection model" *Proc Natl Acad Sci U S A* **2010**, *107*, 22044.
[109] Brown, E. D.; Wright, G. D.,"New targets and screening approaches in antimicrobial drug discovery" *Chem Rev* **2005**, *105*, 759.
[110] Alexander, B. D.; Perfect, J. R.,"Antifungal resistance trends towards the year 2000. Implications for therapy and new approaches" *Drugs* **1997**, *54*, 657.
[111] Shoemaker, C. J.; Green, R.,"Translation drives mRNA quality control" *Nature structural & molecular biology* **2012**, *19*, 594.
[112] Sherman, M. Y.; Qian, S.-B.,"Less is more: improving proteostasis by translation slow down" *Trends in Biochemical Sciences* **2013**, *38*, 585.
[113] Pechmann, S.; Willmund, F.; Frydman, J.,"The ribosome as a hub for protein quality control" *Molecular cell* **2013**, *49*, 411.
[114] Ben-Shem, A.; Jenner, L.; Yusupova, G.; Yusupov, M.,"Crystal structure of the eukaryotic ribosome" *Science* **2010**, *330*, 1203.
[115] Jeeninga, R. E.; Van Delft, Y.; de Graaff-Vincent, M.; Dirks-Mulder, A.; Venema, J.; Raué, H. A.,"Variable regions V13 and V3 of Saccharomyces cerevisiae contain structural features essential for normal biogenesis and stability of 5.8 S and 25S rRNA" *RNA-Publ. RNA Soc.* **1997**, *3*, 476.
[116] Ramesh, M.; Woolford, J. L., Jr.,"Eukaryote-specific rRNA expansion segments function in ribosome biogenesis" *RNA-Publ. RNA Soc.* **2016**.
[117] Hashem, Y.; Des Georges, A.; Fu, J.; Buss, S. N.; Jossinet, F.; Jobe, A.; Zhang, Q.; Liao, H. Y.; Grassucci, R. A.; Bajaj, C.,"High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome" *Nature* **2013**, *494*, 385.
[118] Wilkinson, K. A.; Merino, E. J.; Weeks, K. M.,"Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution" *Nature protocols* **2006**, *1*, 1610.
[119] Hsiao, C.; Lenz, T. K.; Peters, J. K.; Fang, P. Y.; Schneider, D. M.; Anderson, E. J.; Preeprem, T.; Bowman, J. C.; O'Neill, E. B.; Lie, L.; Athavale, S. S.; Gossett, J. J.; Trippe, C.; Murray, J.; Petrov, A. S.; Wartell, R. M.; Harvey, S. C.; Hud, N. V.; Williams, L. D.,"Molecular paleontology: a biochemical model of the ancestral ribosome" *Nucleic Acids Res* **2013**, *41*, 3373.
[120] Brion, P.; Westhof, E.,"Hierarchy and dynamics of RNA folding" *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113.

[121] Colmenarejo, G.; Tinoco, I.,"Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme" *Journal of Molecular Biology* **1999**, *290*, 119.
[122] Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D.,"Cations in charge: magnesium ions in RNA folding and catalysis" *Current Opinion in Structural Biology* **2012**, *22*, 262.
[123] Mortimer, S. A.; Weeks, K. M.,"Time-Resolved RNA SHAPE Chemistry" *J Am Chem Soc* **2008**, *130*, 16178.
[124] Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M.,"RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE)" *J Am Chem Soc* **2005**, *127*, 4223.
[125] Athavale, S. S.; Gossett, J. J.; Hsiao, C.; Bowman, J. C.; O'Neill, E.; Hershkovitz, E.; Preeprem, T.; Hud, N. V.; Wartell, R. M.; Harvey, S. C.; Williams, L. D.,"Domain III of the *T. thermophilus* 23S rRNA folds independently to a near-native state" *RNA-Publ. RNA Soc.* **2012**, *18*, 752.
[126] Bernier, C.; Petrov, A. S.; Waterbury, C.; Jett, J.; Li, F.; Freil, L. E.; Xiong, b.; Wang, L.; Le, A.; Milhouse, B. L.; Hershkovitz, E.; Grover, M.; Xue, Y.; Hsiao, C.; Bowman, J. C.; Harvey, S. C.; Wartel, J. Z.; Williams, L. D.,"RiboVision: Visualization and Analysis of Ribosomes" *Faraday Discuss* **2014**, *169*, 195.
[127] Puglisi, J. D.; Tinoco, I., Jr.,"Absorbance melting curves of RNA" *Methods in Enzymology* **1989**, *180*, 304.
[128] Rappsilber, J.; Mann, M.; Ishihama, Y.,"Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips" *Nature protocols* **2007**, *2*, 1896.
[129] Seiler, C. Y.; Park, J. G.; Sharma, A.; Hunter, P.; Surapaneni, P.; Sedillo, C.; Field, J.; Algar, R.; Price, A.; Steel, J.; Throop, A.; Fiacco, M.; LaBaer, J.,"DNASU plasmid and PSI:Biology-Materials repositories: resources to accelerate biological research" *Nucleic Acids Res* **2014**, *42*, D1253.
[130] Smith, P. K.; Krohn, R. I.; Hermanson, G.; Mallia, A.; Gartner, F.; Provenzano, M.; Fujimoto, E.; Goeke, N.; Olson, B.; Klenk, D.,"Measurement of protein using bicinchoninic acid" *Analytical biochemistry* **1985**, *150*, 76.
[131] Shcherbakov, D.; Piendl, W.,"A novel view of gel-shifts: analysis of RNA-protein complexes using a two-color fluorescence dye procedure" *Electrophoresis* **2007**, *28*, 749.
[132] Fleischer, T. C.; Weaver, C. M.; McAfee, K. J.; Jennings, J. L.; Link, A. J.,"Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes" *Genes Dev* **2006**, *20*, 1294.
[133] Beckmann, B. M.; Horos, R.; Fischer, B.; Castello, A.; Eichelbaum, K.; Alleaume, A. M.; Schwarzl, T.; Curk, T.; Foehr, S.; Huber, W.; Krijgsveld, J.; Hentze, M. W.,"The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs" *Nat Commun* **2015**, *6*, 10127.
[134] Huh, W. K.; Falvo, J. V.; Gerke, L. C.; Carroll, A. S.; Howson, R. W.; Weissman, J. S.; O'Shea, E. K.,"Global analysis of protein localization in budding yeast" *Nature* **2003**, *425*, 686.

[135] Huang, D. W.; Sherman, B. T.; Lempicki, R. A.,"Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists" *Nucleic Acids Research* **2009**, *37*, 1.
[136] Huang, D. W.; Sherman, B. T.; Lempicki, R. A.,"Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources" *Nature protocols* **2009**, *4*, 44.
[137] Kohrer, C.; Mayer, C.; Neumair, O.; Grobner, P.; Piendl, W.,"Interaction of ribosomal L1 proteins from mesophilic and thermophilic Archaea and Bacteria with specific L1-binding sites on 23S rRNA and mRNA" *Eur J Biochem* **1998**, *256*, 97.
[138] Uchiumi, T.; Sato, N.; Wada, A.; Hachimori, A.,"Interaction of the sarcin/ricin domain of 23 S ribosomal RNA with proteins L3 and L6" *J Biol Chem* **1999**, *274*, 681.
[139] Wilkinson, K. A.; Vasa, S. M.; Deigan, K. E.; Mortimer, S. A.; Giddings, M. C.; Weeks, K. M.,"Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA" *RNA-Publ. RNA Soc.* **2009**, *15*, 1314.
[140] Weeks, K. M.; Mauger, D. M.,"Exploring RNA Structural Codes with SHAPE Chemistry" *Accounts of Chemical Research* **2011**, *44*, 1280.
[141] Zuker, M.,"Mfold web server for nucleic acid folding and hybridization prediction" *Nucleic Acids Research* **2003**, *31*, 3406.
[142] Leshin, J. A.; Heselpoth, R.; Belew, A. T.; Dinman, J. D.,"High throughput structural analysis of yeast ribosomes using hSHAPE" *RNA biology* **2011**, *8*, 478.
[143] Draper, D. E.; Bukhman, Y. V.; Gluick, T. C., "Thermal methods for the analysis of RNA folding pathways" In *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]*, 2008/04/23 ed.; **2001**, Vol. Chapter 11, 11.3.1.
[144] Forconi, M.; Herschlag, D.,"Metal ion-based RNA cleavage as a structural probe" *Methods Enzymol* **2009**, *468*, 91.
[145] Ohmayer, U.; Gil-Hernández, Á.; Sauert, M.; Martín-Marcos, P.; Tamame, M.; Tschochner, H.; Griesenbeck, J.; Milkereit, P.,"Studies on the Coordination of Ribosomal Protein Assembly Events Involved in Processing and Stabilization of Yeast Early Large Ribosomal Subunit Precursors" *PloS one* **2015**, *10*, e0143768.
[146] Havrylenko, S.; Mirande, M.,"Aminoacyl-tRNA Synthetase Complexes in Evolution" *Int J Mol Sci* **2015**, *16*, 6571.
[147] Lee, S. W.; Cho, B. H.; Park, S. G.; Kim, S.,"Aminoacyl-tRNA synthetase complexes: beyond translation" *Journal of Cell Science* **2004**, *117*, 3725.
[148] Hausmann, C. D.; Praetorius-Ibba, M.; Ibba, M.,"An aminoacyl-tRNA synthetase:elongation factor complex for substrate channeling in archaeal translation" *Nucleic Acids Res* **2007**, *35*, 6094.
[149] Warner, J. R.,"The economics of ribosome biosynthesis in yeast" *Trends Biochem Sci* **1999**, *24*, 437.
[150] Schimmel, P. R.; Soll, D.,"Aminoacyl-tRNA synthetases: general features and recognition of transfer RNAs" *Annu Rev Biochem* **1979**, *48*, 601.
[151] Kozak, M.,"Regulation of translation in eukaryotic systems" *Annu Rev Cell Biol* **1992**, *8*, 197.
[152] Kozak, M.,"Initiation of translation in prokaryotes and eukaryotes" *Gene* **1999**, *234*, 187.

[153] Ibba, M.; Soll, D.,"Quality control mechanisms during translation" *Science* **1999**, *286*, 1893.
[154] Raught, B.; Gingras, A.-C.; Sonenberg, N., "Regulation of Ribosomal Recruitment in Eukaryotes" In *Translational Control of Gene Expression*, Cold Spring Harbor Laboratory Press: **2000**, Vol. 39, 245.
[155] Kisselev, L.; Ehrenberg, M.; Frolova, L.,"Termination of translation: interplay of mRNA, rRNAs and release factors?" *EMBO J* **2003**, *22*, 175.
[156] Nazar, R. N.,"Ribosomal RNA processing and ribosome biogenesis in eukaryotes" *IUBMB Life* **2004**, *56*, 457.
[157] Jackson, R. J.; Hellen, C. U.; Pestova, T. V.,"The mechanism of eukaryotic translation initiation and principles of its regulation" *Nat Rev Mol Cell Biol* **2010**, *11*, 113.
[158] Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V.,"Structure of the 70S ribosome complexed with mRNA and tRNA" *Science* **2006**, *313*, 1935.
[159] Wilson, D. N.; Doudna Cate, J. H.,"The structure and function of the eukaryotic ribosome" *Cold Spring Harb Perspect Biol* **2012**, *4*.
[160] Klinge, S.; Voigts-Hoffmann, F.; Leibundgut, M.; Ban, N.,"Atomic structures of the eukaryotic ribosome" *Trends Biochem Sci* **2012**, *37*, 189.
[161] Gómez Ramos, L. M.; Smeekens, J. M.; Kovacs, N. A.; Bowman, J. C.; Wartell, R. M.; Wu, R.; Williams, L. D.,"Yeast rRNA Expansion Segments: Folding and Function" *J Mol Biol* **2016**, *428*, 4048.
[162] Behrmann, E.; Loerke, J.; Budkevich, T. V.; Yamamoto, K.; Schmidt, A.; Penczek, P. A.; Vos, M. R.; Bürger, J.; Mielke, T.; Scheerer, P.,"Structural snapshots of actively translating human ribosomes" *Cell* **2015**, *161*, 845.
[163] Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W.,"Defining the human deubiquitinating enzyme interaction landscape" *Cell* **2009**, *138*, 389.
[164] Kofler, M.; Motzny, K.; Beyermann, M.; Freund, C.,"Novel interaction partners of the CD2BP2-GYF domain" *J Biol Chem* **2005**, *280*, 33397.
[165] Goehler, H.; Lalowski, M.; Stelzl, U.; Waelter, S.; Stroedicke, M.; Worm, U.; Droege, A.; Lindenberg, K. S.; Knoblich, M.; Haenig, C.; Herbst, M.; Suopanki, J.; Scherzinger, E.; Abraham, C.; Bauer, B.; Hasenbank, R.; Fritzsche, A.; Ludewig, A. H.; Bussow, K.; Coleman, S. H.; Gutekunst, C. A.; Landwehrmeyer, B. G.; Lehrach, H.; Wanker, E. E.,"A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease" *Mol Cell* **2004**, *15*, 853.
[166] Akaishi, J.; Onda, M.; Okamoto, J.; Miyamoto, S.; Nagahama, M.; Ito, K.; Yoshida, A.; Shimizu, K.,"Down-regulation of transcription elogation factor A (SII) like 4 (TCEAL4) in anaplastic thyroid cancer" *BMC Cancer* **2006**, *6*, 260.
[167] Kofler, M.; Heuer, K.; Zech, T.; Freund, C.,"Recognition sequences for the GYF domain reveal a possible spliceosomal function of CD2BP2" *J Biol Chem* **2004**, *279*, 28292.
[168] Taylor, D.; Unbehaun, A.; Li, W.; Das, S.; Lei, J.; Liao, H. Y.; Grassucci, R. A.; Pestova, T. V.; Frank, J.,"Cryo-EM structure of the mammalian eukaryotic release factor eRF1-eRF3-associated termination complex" *Proc Natl Acad Sci U S A* **2012**, *109*, 18413.

[169] Napoli, I.; Mercaldo, V.; Boyl, P. P.; Eleuteri, B.; Zalfa, F.; De Rubeis, S.; Di Marino, D.; Mohr, E.; Massimi, M.; Falconi, M.; Witke, W.; Costa-Mattioli, M.; Sonenberg, N.; Achsel, T.; Bagni, C.,"The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP" *Cell* **2008**, *134*, 1042.
[170] Zanotti, K. J.; Lackey, P. E.; Evans, G. L.; Mihailescu, M. R.,"Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA" *Biochemistry* **2006**, *45*, 8319.
[171] Darnell, J. C.; Jensen, K. B.; Jin, P.; Brown, V.; Warren, S. T.; Darnell, R. B.,"Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function" *Cell* **2001**, *107*, 489.
[172] Allton, K.; Jain, A. K.; Herz, H. M.; Tsai, W. W.; Jung, S. Y.; Qin, J.; Bergmann, A.; Johnson, R. L.; Barton, M. C.,"Trim24 targets endogenous p53 for degradation" *Proc Natl Acad Sci U S A* **2009**, *106*, 11612.
[173] Le Douarin, B.; Nielsen, A. L.; Garnier, J. M.; Ichinose, H.; Jeanmougin, F.; Losson, R.; Chambon, P.,"A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors" *EMBO J* **1996**, *15*, 6701.
[174] Le Douarin, B.; You, J.; Nielsen, A. L.; Chambon, P.; Losson, R.,"TIF1alpha: a possible link between KRAB zinc finger proteins and nuclear receptors" *J Steroid Biochem Mol Biol* **1998**, *65*, 43.
[175] Kartberg, F.; Asp, L.; Dejgaard, S. Y.; Smedh, M.; Fernandez-Rodriguez, J.; Nilsson, T.; Presley, J. F.,"ARFGAP2 and ARFGAP3 are essential for COPI coat assembly on the Golgi membrane of living cells" *J Biol Chem* **2010**, *285*, 36709.
[176] Goldfarb, D. S.; Corbett, A. H.; Mason, D. A.; Harreman, M. T.; Adam, S. A.,"Importin alpha: a multipurpose nuclear-transport receptor" *Trends Cell Biol* **2004**, *14*, 505.
[177] Nigg, E. A.,"Nucleocytoplasmic transport: signals, mechanisms and regulation" *Nature* **1997**, *386*, 779.
[178] Wang, C. C.; Schimmel, P.,"Species barrier to RNA recognition overcome with nonspecific RNA binding domains" *J Biol Chem* **1999**, *274*, 16508.
[179] Ahn, D.-G.; Lee, W.; Choi, J.-K.; Kim, S.-J.; Plant, E. P.; Almazán, F.; Taylor, D. R.; Enjuanes, L.; Oh, J.-W.,"Interference of ribosomal frameshifting by antisense peptide nucleic acids suppresses SARS coronavirus replication" *Antiviral research* **2011**, *91*, 1.
[180] Zheng, X. M.; Moncollin, V.; Egly, J. M.; Chambon, P.,"A general transcription factor forms a stable complex with RNA polymerase B (II)" *Cell* **1987**, *50*, 361.
[181] Sun, C.; Todorovic, A.; Querol-Audi, J.; Bai, Y.; Villa, N.; Snyder, M.; Ashchyan, J.; Lewis, C. S.; Hartland, A.; Gradia, S.; Fraser, C. S.; Doudna, J. A.; Nogales, E.; Cate, J. H.,"Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3)" *Proc Natl Acad Sci U S A* **2011**, *108*, 20473.
[182] Gingras, A. C.; Raught, B.; Sonenberg, N.,"eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation" *Annu Rev Biochem* **1999**, *68*, 913.

[183] Siomi, M. C.; Siomi, H.; Sauer, W. H.; Srinivasan, S.; Nussbaum, R. L.; Dreyfuss, G.,"FXR1, an autosomal homolog of the fragile X mental retardation gene" *EMBO J* **1995**, *14*, 2401.
[184] Hofmann, K.; Bucher, P.,"The PCI domain: a common theme in three multiprotein complexes" *Trends Biochem Sci* **1998**, *23*, 204.
[185] Hsiao, C.; Tannenbaum, M.; VanDeusen, H.; Hershkovitz, E.; Perng, G.; Tannenbaum, A.; Williams, L. D., "Complexes of Nucleic Acids with Group I and II Cations" In *Nucleic Acid Metal Ion Interactions*, Hud, N., Ed. The Royal Society of Chemistry: London, **2008**, 1.
[186] Gulen, B.; Petrov, A. S.; Okafor, C. D.; Vander Wood, D.; O'Neill, E. B.; Hud, N. V.; Williams, L. D.,"Ribosomal Small Subunit Domains Radiate from a Central Core" *Scientific Reports* **2016**, *6*, 20885.
[187] Lanier, K. A.; Athavale, S. S.; Petrov, A. S.; Wartell, R.; Williams, L. D.,"Imprint of Ancient Evolution on rRNA Folding" *Biochemistry* **2016**, *55*, 4603.
[188] Kikin, O.; D'Antonio, L.; Bagga, P. S.,"QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences" *Nucleic Acids Res* **2006**, *34*, W676.
[189] von Hacht, A.; Seifert, O.; Menger, M.; Schutze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J.,"Identification and characterization of RNA guanine-quadruplex binding proteins" *Nucleic Acids Res* **2014**, *42*, 6630.
[190] Fratta, P.; Mizielinska, S.; Nicoll, A. J.; Zloh, M.; Fisher, E. M.; Parkinson, G.; Isaacs, A. M.,"C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes" *Sci Rep* **2012**, *2*, 1016.
[191] Halder, K.; Wieland, M.; Hartig, J. S.,"Predictable suppression of gene expression by 5'-UTR-based RNA quadruplexes" *Nucleic Acids Res* **2009**, *37*, 6811.
[192] Arora, A.; Dutkiewicz, M.; Scaria, V.; Hariharan, M.; Maiti, S.; Kurreck, J.,"Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif" *RNA-Publ. RNA Soc.* **2008**, *14*, 1290.
[193] Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S.,"An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation" *Nat Chem Biol* **2007**, *3*, 218.
[194] Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S.,"Quadruplex DNA: sequence, topology and structure" *Nucleic Acids Research* **2006**, *34*, 5402.
[195] Wang, F.; Durfee, L. A.; Huibregtse, J. M.,"A cotranslational ubiquitination pathway for quality control of misfolded proteins" *Molecular cell* **2013**, *50*, 368.
[196] Schubert, U.; Anton, L. C.; Gibbs, J.; Norbury, C. C.; Yewdell, J. W.; Bennink, J. R.,"Rapid degradation of a large fraction of newly synthesized proteins by proteasomes" *Nature* **2000**, *404*, 770.
[197] Qian, S. B.; Princiotta, M. F.; Bennink, J. R.; Yewdell, J. W.,"Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control" *J Biol Chem* **2006**, *281*, 392.
[198] Ciechanover, A.,"The ubiquitin-proteasome pathway: on protein death and cell life" *EMBO J* **1998**, *17*, 7151.

[199] Hershko, A., "The ubiquitin system" In *Ubiquitin and the Biology of the Cell*, Springer: **1998**, 1.
[200] Glickman, M. H.; Ciechanover, A.,"The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction" *Physiol Rev* **2002**, *82*, 373.
[201] van der Veen, A. G.; Ploegh, H. L.,"Ubiquitin-like proteins" *Annu Rev Biochem* **2012**, *81*, 323.
[202] Hochstrasser, M.,"Origin and function of ubiquitin-like proteins" *Nature* **2009**, *458*, 422.
[203] Kerscher, O.; Felberbaum, R.; Hochstrasser, M.,"Modification of proteins by ubiquitin and ubiquitin-like proteins" *Annu Rev Cell Dev Biol* **2006**, *22*, 159.
[204] Baltz, A. G.; Munschauer, M.; Schwanhausser, B.; Vasile, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; Wyler, E.; Bonneau, R.; Selbach, M.; Dieterich, C.; Landthaler, M.,"The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts" *Mol Cell* **2012**, *46*, 674.
[205] Stavreva, D. A.; Kawasaki, M.; Dundr, M.; Koberna, K.; Muller, W. G.; Tsujimura-Takahashi, T.; Komatsu, W.; Hayano, T.; Isobe, T.; Raska, I.; Misteli, T.; Takahashi, N.; McNally, J. G.,"Potential roles for ubiquitin and the proteasome during ribosome biogenesis" *Mol Cell Biol* **2006**, *26*, 5131.
[206] Glickman, M. H.; Raveh, D.,"Proteasome plasticity" *FEBS Lett* **2005**, *579*, 3214.
[207] Auffinger, P.; Grover, N.; Westhof, E.,"Metal ion binding to RNA" *Met Ions Life Sci* **2011**, *9*, 1.
[208] Draper, D. E.,"RNA Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions" *Biophysical Journal* **2008**, *95*, 5489.
[209] Siomi, M. C.; Zhang, Y.; Siomi, H.; Dreyfuss, G.,"Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them" *Mol Cell Biol* **1996**, *16*, 3825.
[210] Brazda, V.; Haronikova, L.; Liao, J. C.; Fojta, M.,"DNA and RNA quadruplex-binding proteins" *Int J Mol Sci* **2014**, *15*, 17493.
[211] Sutcliffe, J. A.,"Improving on nature: antibiotics that target the ribosome" *Current opinion in microbiology* **2005**, *8*, 534.
[212] Tenover, F. C.,"Mechanisms of antimicrobial resistance in bacteria" *The American journal of medicine* **2006**, *119*, S3.
[213] "Antimicrobial Resistance: Global Report on Surveillance " World Health Organization **2014**.
[214] Silver, L. L.,"Challenges of antibacterial discovery" *Clinical microbiology reviews* **2011**, *24*, 71.
[215] McNeil, M. M.; Nash, S. L.; Hajjeh, R. A.; Phelan, M. A.; Conn, L. A.; Plikaytis, B. D.; Warnock, D. W.,"Trends in mortality due to invasive mycotic diseases in the United States, 1980–1997" *Clinical Infectious Diseases* **2001**, *33*, 641.
[216] Nucci, M.; Marr, K. A.,"Emerging fungal diseases" *Clinical Infectious Diseases* **2005**, *41*, 521.
[217] Menzin, J.; Meyers, J. L.; Friedman, M.; Perfect, J. R.; Langston, A. A.; Danna, R. P.; Papadopoulos, G.,"Mortality, length of hospitalization, and costs associated

with invasive fungal infections in high-risk patients" *Am J Health Syst Pharm* **2009**, *66*, 1711.
[218] Edmond, M. B.; Wallace, S. E.; McClish, D. K.; Pfaller, M. A.; Jones, R. N.; Wenzel, R. P.,"Nosocomial bloodstream infections in United States hospitals: a three-year analysis" *Clinical infectious diseases* **1999**, *29*, 239.
[219] Rex, J. H.; Rinaldi, M.; Pfaller, M.,"Resistance of Candida species to fluconazole" *Antimicrobial agents and chemotherapy* **1995**, *39*, 1.
[220] "Antibiotic Resistance Threats in the United States" Center for Disease Control and Prevention: **2013**.
[221] Bowman, J. C.; Azizi, B.; Lenz, T. K.; Roy, P.; Williams, L. D., "Preparation of long templates for RNA in vitro transcription by Recursive PCR" In *Recombinant and In Vitro RNA Synthesis: Methods and Protocols, Methods in Molecular Biology*, Conn, G. L., Ed. Springer Science, LLC: **2012**, Vol. 941, 19.
[222] Jiang, L.; Watkins, D.; Jin, Y.; Gong, C.; King, A.; Washington, A. Z.; Green, K. D.; Garneau-Tsodikova, S.; Oyelere, A. K.; Arya, D. P.,"Rapid synthesis, RNA binding, and antibacterial screening of a peptidic-aminosugar (PA) library" *ACS Chem Biol* **2015**, *10*, 1278.
[223] Watkins, D.; Norris, F. A.; Kumar, S.; Arya, D. P.,"A fluorescence-based screen for ribosome binding antibiotics" *Anal Biochem* **2013**, in press.
[224] Popenda, M.; Szachniuk, M.; Antczak, M.; Purzycka, K. J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R. W.,"Automated 3D structure composition for large RNAs" *Nucleic Acids Research* **2012**.
[225] Biesiada, M.; Purzycka, K. J.; Szachniuk, M.; Blazewicz, J.; Adamiak, R. W.,"Automated RNA 3D Structure Prediction with RNAComposer" *Methods in molecular biology (Clifton, N.J.)* **2016**, *1490*, 199.
[226] Antczak, M.; Zok, T.; Popenda, M.; Lukasiak, P.; Adamiak, R. W.; Blazewicz, J.; Szachniuk, M.,"RNApdbee--a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs" *Nucleic Acids Res* **2014**, *42*, W368.
[227] Ryu, D. H.; Rando, R. R.,"Aminoglycoside binding to human and bacterial A-Site rRNA decoding region constructs" *Bioorg Med Chem* **2001**, *9*, 2601.
[228] Kumar, S.; Kellish, P.; Robinson, W. E., Jr.; Wang, D.; Appella, D. H.; Arya, D. P.,"Click dimers to target HIV TAR RNA conformation" *Biochemistry* **2012**, *51*, 2331.
[229] Vasicek, E. M.; Berkow, E. L.; Bruno, V. M.; Mitchell, A. P.; Wiederhold, N. P.; Barker, K. S.; Rogers, P. D.,"Disruption of the transcriptional regulator Cas5 results in enhanced killing of Candida albicans by Fluconazole" *Antimicrob Agents Chemother* **2014**, *58*, 6807.
[230] Famulok, M.; Huttenhofer, A.,"In vitro selection analysis of neomycin binding RNAs with a mutagenized pool of variants of the 16S rRNA decoding region" *Biochemistry* **1996**, *35*, 4265.
[231] Wallis, M. G.; von Ahsen, U.; Schroeder, R.; Famulok, M.,"A novel RNA motif for neomycin recognition" *Chem Biol* **1995**, *2*, 543.
[232] Xi, H.; Gray, D.; Kumar, S.; Arya, D. P.,"Molecular recognition of single-stranded RNA: neomycin binding to poly(A)" *FEBS Lett* **2009**, *583*, 2269.
[233] Arya, D. P.; Coffee, R. L., Jr.; Charles, I.,"Neomycin-induced hybrid triplex formation" *J Am Chem Soc* **2001**, *123*, 11093.

[234] Xue, L.; Charles, I.; Arya, D. P.,"Pyrene-neomycin conjugate: dual recognition of a DNA triple helix" *Chem Commun (Camb)* **2002**, 70.
[235] Wilson, D. N.,"Ribosome-targeting antibiotics and mechanisms of bacterial resistance" *Nat Rev Microbiol* **2014**, *12*, 35.
[236] Ren, B.; Huang, P.; Zhang, J.; He, W.; Han, J.; Liu, X.; Zhang, L., "Main Applications of Antibiotics" In *Antibiotics: Current Innovations and Future Trends*, Sanchez, S.; Demain, A. L., Eds. Caister Academic Press: Poole, UK, **2015**, 19.
[237] Thomas, J. R.; Hergenrother, P. J.,"Targeting RNA with small molecules" *Chemical reviews* **2008**, *108*, 1171.
[238] Howe, J. A.; Wang, H.; Fischmann, T. O.; Balibar, C. J.; Xiao, L.; Galgoci, A. M.; Malinverni, J. C.; Mayhood, T.; Villafania, A.; Nahvi, A.,"Selective small-molecule inhibition of an RNA structural element" *Nature* **2015**.
[239] Shortridge, M. D.; Varani, G.,"Structure based approaches for targeting non-coding RNAs with small molecules" *Current opinion in structural biology* **2015**, *30*, 79.
[240] Lee, M.-K.; Bottini, A.; Kim, M.; Bardaro, M. F.; Zhang, Z.; Pellecchia, M.; Choi, B.-S.; Varani, G.,"A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication" *Chem. Commun.* **2014**, *50*, 368.
[241] Davies, J.; Wright, G. D.,"Bacterial resistance to aminoglycoside antibiotics" *Trends Microbiol* **1997**, *5*, 234.
[242] Fourmy, D.; Recht, M. I.; Puglisi, J. D.,"Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 S rRNA" *Journal of Molecular Biology* **1998**, *277*, 347.
[243] Kohanski, M. A.; Dwyer, D. J.; Collins, J. J.,"How antibiotics kill bacteria: from targets to networks" *Nature Reviews Microbiology* **2010**, *8*, 423.
[244] Kohanski, M. A.; Dwyer, D. J.; Wierzbowski, J.; Cottarel, G.; Collins, J. J.,"Mistranslation of membrane proteins and two-component system activation trigger antibiotic-mediated cell death" *Cell* **2008**, *135*, 679.
[245] Liu, Y.; Imlay, J. A.,"Cell death from antibiotics without the involvement of reactive oxygen species" *Science* **2013**, *339*, 1210.
[246] Kaul, M.; Pilch, D. S.,"Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16S rRNA" *Biochemistry* **2002**, *41*, 7695.
[247] Ranjan, N.; Andreasen, K. F.; Kumar, S.; Hyde-Volpe, D.; Arya, D. P.,"Aminoglycoside binding to Oxytricha nova telomeric DNA" *Biochemistry* **2010**, *49*, 9891.
[248] Arya, D. P.; Coffee, R. L., Jr.,"DNA triple helix stabilization by aminoglycoside antibiotics" *Bioorganic & medicinal chemistry letters* **2000**, *10*, 1897.
[249] Arya, D. P.; Coffee, R. L., Jr.; Willis, B.; Abramovitch, A. I.,"Aminoglycoside-nucleic acid interactions: remarkable stabilization of DNA and RNA triple helices by neomycin" *J Am Chem Soc* **2001**, *123*, 5385.
[250] Arya, D. P.; Micovic, L.; Charles, I.; Coffee, R. L., Jr.; Willis, B.; Xue, L.,"Neomycin binding to Watson-Hoogsteen (W-H) DNA triplex groove: a model" *J Am Chem Soc* **2003**, *125*, 3733.

[251] Paige, J. S.; Wu, K. Y.; Jaffrey, S. R.,"RNA mimics of green fluorescent protein" *Science* **2011**, *333*, 642.
[252] Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R.,"Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution" *J Am Chem Soc* **2014**, *136*, 16299.
[253] Shaner, N. C.; Campbell, R. E.; Steinbach, P. A.; Giepmans, B. N.; Palmer, A. E.; Tsien, R. Y.,"Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein" *Nat Biotechnol* **2004**, *22*, 1567.
[254] Chow, C. S.; Lamichhane, T. N.; Mahto, S. K.,"Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications" *ACS Chem Biol* **2007**, *2*, 610.
[255] Cantara, W. A.; Crain, P. F.; Rozenski, J.; McCloskey, J. A.; Harris, K. A.; Zhang, X.; Vendeix, F. A.; Fabris, D.; Agris, P. F.,"The RNA Modification Database, RNAMDB: 2011 update" *Nucleic Acids Res* **2011**, *39*, D195.
[256] Andersen, T. E.; Porse, B. T.; Kirpekar, F.,"A novel partial modification at C2501 in Escherichia coli 23S ribosomal RNA" *RNA-Publ. RNA Soc.* **2004**, *10*, 907.
[257] Mengel-Jørgensen, J.; Jensen, S. S.; Rasmussen, A.; Poehlsgaard, J.; Iversen, J. J. L.; Kirpekar, F.,"Modifications in Thermus thermophilus 23 S Ribosomal RNA Are Centered in Regions of RNA-RNA Contact" *Journal of Biological Chemistry* **2006**, *281*, 22108.
[258] Maden, B. E.,"The numerous modified nucleotides in eukaryotic ribosomal RNA" *Prog Nucleic Acid Res Mol Biol* **1990**, *39*, 241.
[259] Sen, D.; Gilbert, W.,"Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis" *Nature* **1988**, *334*, 364.
[260] Sundquist, W. I.; Klug, A.,"Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops" *Nature* **1989**, *342*, 825.
[261] Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H.,"Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription" *Proc Natl Acad Sci U S A* **2002**, *99*, 11593.
[262] Cogoi, S.; Xodo, L. E.,"G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription" *Nucleic Acids Res* **2006**, *34*, 2536.
[263] Subramanian, M.; Rage, F.; Tabet, R.; Flatter, E.; Mandel, J. L.; Moine, H.,"G-quadruplex RNA structure as a signal for neurite mRNA targeting" *EMBO Rep* **2011**, *12*, 697.
[264] Huppert, J. L.; Balasubramanian, S.,"G-quadruplexes in promoters throughout the human genome" *Nucleic Acids Res* **2007**, *35*, 406.
[265] Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S.,"High-throughput sequencing of DNA G-quadruplex structures in the human genome" *Nat Biotechnol* **2015**, *33*, 877.
[266] Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S.,"Quantitative visualization of DNA G-quadruplex structures in human cells" *Nat Chem* **2013**, *5*, 182.

[267] Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S.,"Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells" *Nat Chem* **2014**, *6*, 75.
[268] Kwok, C. K.; Sahakyan, A. B.; Balasubramanian, S.,"Structural Analysis using SHALiPE to Reveal RNA G-Quadruplex Formation in Human Precursor MicroRNA" *Angew Chem Int Ed Engl* **2016**, *55*, 8958.
[269] Weldon, C.; Behm-Ansmant, I.; Hurley, L. H.; Burley, G. A.; Branlant, C.; Eperon, I. C.; Dominguez, C.,"Identification of G-quadruplexes in long functional RNAs using 7-deazaguanine RNA" *Nat Chem Biol* **2016**.
[270] Silver, L. L.,"Challenges of antibacterial discovery" *Clin Microbiol Rev* **2011**, *24*, 71.
[271] Eng, J. K.; McCormack, A. L.; Yates, J. R.,"An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database" *J Am Soc Mass Spectrom* **1994**, *5*, 976.
[272] Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P.,"Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome" *J Proteome Res* **2003**, *2*, 43.
[273] Elias, J. E.; Gygi, S. P.,"Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry" *Nat Methods* **2007**, *4*, 207.
[274] Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villen, J.; Haas, W.; Sowa, M. E.; Gygi, S. P.,"A tissue-specific atlas of mouse protein phosphorylation and expression" *Cell* **2010**, *143*, 1174.
[275] Wu, R.; Dephoure, N.; Haas, W.; Huttlin, E. L.; Zhai, B.; Sowa, M. E.; Gygi, S. P.,"Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes" *Mol Cell Proteomics* **2011**, *10*, M111 009654.
[276] Du, X.; Callister, S. J.; Manes, N. P.; Adkins, J. N.; Alexandridis, R. A.; Zeng, X.; Roh, J. H.; Smith, W. E.; Donohue, T. J.; Kaplan, S.; Smith, R. D.; Lipton, M. S.,"A computational strategy to analyze label-free temporal bottom-up proteomics data" *J Proteome Res* **2008**, *7*, 2595.
[277] Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J.,"Semi-supervised learning for peptide identification from shotgun proteomics datasets" *Nat. Methods* **2007**, *4*, 923.
[278] Marky, L. A.; Breslauer, K. J.,"Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves" *Biopolymers* **1987**, *26*, 1601.
[279] Trott, O.; Olson, A. J.,"AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading" *Journal of computational chemistry* **2010**, *31*, 455.
[280] Sanner, M. F.,"Python: a programming language for software integration and development" *J Mol Graph Model* **1999**, *17*, 57.