UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY
CHAIR OF MOLECULAR BIOLOGY

JEKATERINA AID

The role of S4-like RNA-binding and catalytic domains in the activity and specificity of bacterial pseudouridine synthases RluC and RluD *in vitro*

Master's Thesis

Supervisors: Prof. Jaanus Remme, PhD
Margus Leppik, PhD

TARTU 2015

Table of Contents

Abbreviations

30S subunit	prokaryotic small ribosomal subunit
50S subunit	prokaryotic large ribosomal subunit
70S	prokaryotic ribosome
Amp	ampicillin
AMV reverse transcriptase	Avian Myeloblastosis Virus (AMV) reverse transcriptase
A-site	acceptor site for aminoacyl tRNA on the ribosome
BipA	GTP-binding protein
CMCT	1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-p-toluene sulfonate
CP	central protuberance
D30S	30S subunits dissociated from the 70S ribosomes
D50S	50S subunits dissociated from the 70S ribosomes
DC	decoding center
DNase I	Deoxyribonuclease I
dNTP	deoxynucleotide triphosphate
DTT (DTE)	dithiothreitol (dithioerythritol)
EF-G	prokaryotic elongation factor G
EF-Tu	prokaryotic elongation factor thermo unstable
E-site	exit site for decylated tRNA on the ribosome
GTPase	GTP binding and hydrolyzing enzyme
H69	stem-loop 69 of the ribosomal large subunit RNA
helix 44	stem-loop 44 of the ribosomal small subunit RNA
HPLC	High-performance liquid chromatography
LepA	leader peptidase A or elongation factor 4
L-proteins	ribosomal large subunit proteins
LSU	ribosomal large subunit
mRNA	messenger RNA
NMR	nuclear magnetic resonance
P-site	acceptor site for peptidyl tRNA on the ribosome
PTC	peptidyl transferase center
Pus10	putative tRNA pseudouridine synthase Pus10
RF2	ribosomal release factor 2
RluA	ribosomal large subunit pseudouridine synthase A

RluC	ribosomal large subunit pseudouridine synthase C
RluCD	chimeric pseudouridine synthase with RluD catalytic domain and RluC S4-like domain
RluD	ribosomal large subunit pseudouridine synthase D
RluDC	chimeric pseudouridine synthase with RluC catalytic domain and RluD S4-like domain
r-proteins	ribosomal proteins
rRNA	ribosomal RNA
RsuA	ribosomal small subunit pseudouridine synthase A
S4-like domain	protein domain that resembles ribosomal small subunit protein S4
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
S-proteins	ribosomal small subunit proteins
SRL	sarcin-ricin loop of 23S ribosomal RNA
SSU	ribosomal small subunit
tmRNA	transfer-messenger RNA
tRNA	transfer RNA
TruA	tRNA oseudouridine synthase A
TruB	tRNA oseudouridine synthase B
TruD	tRNA oseudouridine synthase D
Ψ synthase	pseudouridine synthase
Ψ	pseudouridine

Introduction

Protein synthesis is a remarkably accurate process of translating three-letter words of the nucleic acid code into protein "language" using 20 amino acid letters. This translational dictionary is virtually universal. To carry out protein synthesis all organisms ranging from bacteria to eukarya use ribosomal machineries which are composed of a large and a small subunit. Both subunits consist of ribosomal RNA and ribosomal proteins whose coordinated action ensures the fidelity of protein synthesis. Ribosomal RNA (rRNA) constitutes the largest portion of the ribosome. Notably, rRNA, the most abundant noncoding RNA in the cell, undergoes numerous posttranscriptional site-specific nucleotide modifications. Although the precise function of distinct types of rRNA modifications are not fully understood, it is now becoming clear that modifications cluster within important regions of the ribosome. Findings that the majority of rRNA modifications are highly conserved and increase in number from archaea to eukarya suggest an important functional role for modifications within ribosomes (McMahon *et al*., 2013). Moreover, defects in the enzymes that carry out those modifications have been connected with inheritable human diseases and cancer (Scheper *et al*., 2007), pointing to the importance of these modifications for cellular function.

The most abundant modification in ribosomal RNA is pseudouridine – a 5-rybosyl isomer of uridine. Despite the fact that pseudouridines were discovered over 50 years ago (Davis and Allen, 1957), little is known about their synthesis and function. Pseudouridine synthases, the enzymes responsible for pseudouridylation, were first identified in *Escherichia coli*. They are classified into six families, five of which are named after the *E. coli* enzymes RluA, RsuA, TruA, TruB, and TruD (Del Campo *et al*., 2001; Gustafsson *et al*., 1996; Kaya and Ofengand 2003; Koonin, 1996), and the sixth family Pus10 is present only in archaea and eukarya (McCleverty *et al*., 2007). In bacteria, pseudouridine synthases are protein enzymes which possess catalytic activity, substrate recognition specificity and RNA binding activity. Most bacterial pseudouridine synthases modify only one uridine at a specific site, with the exception of four RNA pseudouridine synthases (TruA, RluA, RluD and RluC) which modify uridine nucleotides at several positions. Both RluD and RluC synthases isomerize uridines at functionally important regions of the ribosome. This is supported by the discovery that RluC deficiency is connected with increased susceptibility of bacteria to peptidyl transferase center inhibitors (Toh and Mankin, 2008). Also, it has been found that defects in eukaryotic pseudouridine synthases are connected with inherited diseases such as mitochondrial myopathy and sideroblastic anemia (MLASA), X-linked dyskeratosis congenita and cancer in human (Bykhovskaya *et al*., 2004; Scheper *et al*., 2007). Moreover, pseudouridines have been

found in eukaryotic mRNA (Schwartz *et al*., 2014) suggesting their direct role in the regulation of gene expression. Finally, pseudouridines have potential in therapeutic applications due to their ability to enhance stability and translational capacity of mRNA while diminishing its immunogenicity *in vivo* (Karikó *et al*., 2012). In the light of these latest discoveries, understanding pseudouridine synthase function and mechanism of action becomes an important goal in molecular biology.

This work focuses on bacterial pseudouridine synthases RluD and RluC. Getting insight into the substrate recognition and catalytic mechanism of these pseudouridine synthases as well as the role of their RNA binding domain is important for the basic understanding of the ribosome biogenesis and the control of gene expression at the translational level. For this study, chimeric proteins RluCD and RluDC were used with exchanged RNA binding S4-like domains, with RluCD carrying the catalytic domain of RluD synthase and S4-like domain of RluC synthase, and RluDC – *vice versa*. The aims of this thesis were a) to purify active chimeric pseudouridine synthase proteins, b) to test their activity on different substrates, c) to map the positions of pseudouridines in 23S rRNA produced by RluCD synthase and d) to study the dependence of their catalytic activity on the concentration of magnesium cations in solution during pseudouridine isomerization reaction *in vitro*. The results of the present study show that S4-like domain plays an important role in pseudouridine synthase specificity. They highlight the importance of coordinated action of the S4-like and catalytic domains to assure specificity of the RluD and RluC synthases.

Keywords: pseudouridine synthase; RluD; RluC; chimeric pseudouridine synthases; pseudouridine

1. Literature overview

1.1. The composition of bacterial ribosome

Gene expression through protein synthesis, a process named translation, is essential for all life forms from viruses and bacteria to mammals. Decoding genetic information stored in the nucleic acid, RNA or DNA sequence, into the amino acid sequence of the proteins generates macromolecules that support structure and function of the cell. The ribosome is the main component of the translational machinery that carries out this important task. The bacterial ribosome consists of two subunits: a large subunit and a small subunit and has molecular weight of about 2.5 MDa. The ribosomal subunits contain many small proteins with molecular masses of less than 20 kDa (Waller and Harris, 1961) and the core of each subunit is formed by a large untranslated ribosomal RNA (rRNA) molecule (Kurland, 1960). Thus, the ribosome is two-thirds ribosomal rRNA and one-third ribosomal proteins (Tissières *et al*., 1959; Ban *et al*., 2000). Bacterial ribosome sediments as a 70S particle: the small subunit sediments at 30S and the large subunit at 50S (Tissieres and Watson, 1958; Tissières *et al*., 1959; Ramakrishnan, 2002). The small 30S subunit contains 16S rRNA and about 21 proteins, and the large 50S subunit consists of two rRNAs, 23S rRNA and 5S rRNA, and over 30 proteins (Yusupov *et al*., 2001). Association of the 30S and 50S subunits though a network of intermolecular bridges produces the complete functional 70S ribosome (Yusupov *et al*., 2001).

The mechanism of the ribosome action in translation is mainly based on the catalytic properties of the rRNA, i.e., ribosome is ribozyme (Yusupov *et al*., 2001). During translation, transfer RNA (tRNA) occupies intersubunit space and its anticodons base-pair with the messenger RNA (mRNA) codons in the decoding center (DC) of 30S subunit, while tRNA 3'-CCA end with growing polypeptide chain or the incoming amino acid reach into the 50S subunit. There is the pepridyl transferase center located on 50S subunit, where peptide bond formation takes place (Yusupov *et al*., 2001).

1.1.1. The small subunit

The small ribosomal subunit, referred to as 30S subunit in prokaryotes, provides the decoding of genetic information during translation. The 30S subunit decodes mRNA by monitoring base-pairing between the codon on messenger RNA (mRNA) and the anticodon on transfer RNA (tRNA) (Schluenzen *et al*., 2000; Wimberly *et al*., 2000). Also, it initiates mRNA

engagement, regulates mRNA and tRNA translocation and controls fidelity of codon-anticodon interactions. The small subunit has the molecular weight of about 0.85 MDa and consists of 16S rRNA and about 21 ribosomal proteins or r-proteins (S1-S21) in *E. coli*. The shape of the 30S subunit is mainly determined by the RNA component. There are two sides distinguished in the small subunit, the interface side interacts with the large 50S subunit, and the opposite side – the back or the solvent side of the 30S subunit (Lake, 1985). The distribution of the r-proteins and RNA in the 30S subunit is asymmetric (Ramakrishnan, 1986; Wimberly *et al*., 2000; Yusupov *et al*., 2001). The interface side of the 30S subunit has few proteins, and they are mostly located on the periphery. Consequently, this suggests that the intersubunit contacting surface of the 30S ribosomal subunit is composed of ribosomal RNA. Most proteins of the 30S subunit are located on the exterior side of the subunit, S12 protein is the only exception located at the RNA-rich surface that interacts with the large subunit (Yusupov and Spirin, 1986; Yusupov *et al*., 2001; Wimberly *et al*., 2000; Schluenzen *et al*., 2000; Carter *et al*., 2000).

The 30S subunit has features called the body, the neck and the head. The upper part of the body has the “shoulder” and on the opposite side - the “platform”. The bottom part of the body has a protuberance called the “toe” or the “spur”. The head has the “nose” with the “beak” (Figure 1) (Wimberly *et al*., 2000; Schluenzen *et al*., 2000). Regardless of species, this general structure of the small subunit seems to be universal (Lake, 1985). The 16S rRNA of 30S subunit is divided into four domains: 5’ domain, central domain, 3’ major domain and 3’ minor domain. Almost all domains of the 16S rRNA are located in different regions of the 30S subunit. The 5’ domain of the 16S rRNA located in the body of small subunit which contains S4, S5, S12, S16, S17 and S20 proteins. The central domain makes up most of the platform by interacting with proteins S1, S6, S8, S11, S15, and S18. The 3’ major domain forms the bulk of the head which is containing S2, S3, S7, S9, S10, S13, S14, and S19. The only one exception is the 3’ minor domain which is a part of the body at the subunit interface. The 3’ minor domain is made up of the two helices h44 and h45, where long helix h44 runs from the region between the head and the body down to the bottom of the 30S subunit on the surface that faces the 50S subunit (Figure 1) (Wimberly *et al*., 2000; Schluenzen *et al*., 2000; Wilson and Nierhaus 2005). The all four domains of 16S rRNA branch from center of the neck, which is functionally the most important region of the small subunit. This domain organization gives flexibility to the small subunit that is essential for its function (Wimberly *et al*., 2000; Ogle *et al*., 2003).

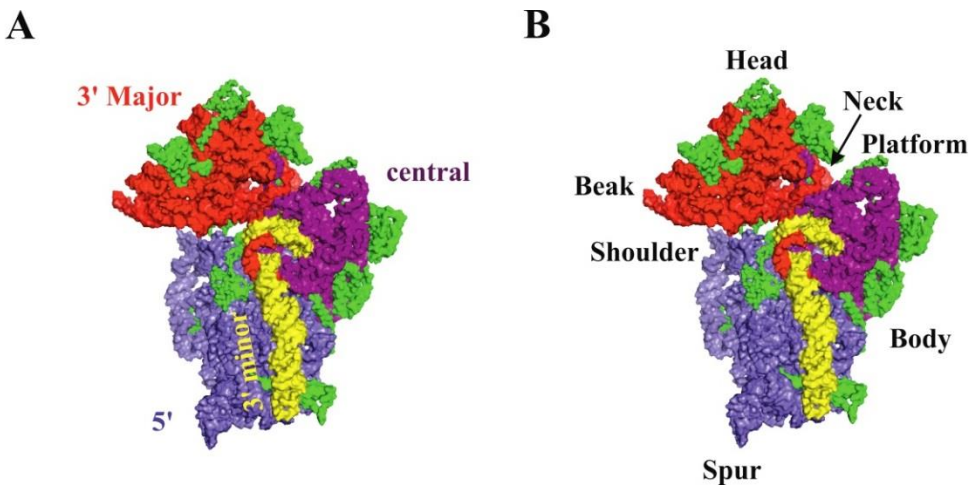

Figure 1. The tertiary structure of 30S ribosomal subunit of *Thermus thermophiles* (PDB entry 1FKA), showing the 50S of "front" view with indication of different 16S rRNA domains. A. The domains of 16S rRNA assembled into 30S ribosomal subunit. The 5' domain (blue), central domain (purple), 3' major domain (red), 3' minor domain (yellow). The small ribosomal proteins are colored green. **B.** The morphological features of the 30S ribosomal subunit: head, neck, beak, platform, shoulder, body and spur. Illustration was created using PyMOL DeLano Scientific software.

The main function of 30S is decoding the mRNA; therefore the most important parts of the 30S subunit are the substrate-binding A-, P- and E-sites. The A-site (also termed decoding center or DC) is much wider and shallower than the P- or E-sites, and has much lower affinity for tRNA (Carter *et al*., 2000). The decoding center provides mRNA and tRNA translocation and controls fidelity in codon-anticodon interactions (Green and Noller, 1997). The A-site is made up of four different domains: the head, shoulder, platform and helix 44 (Ogle *et al*., 2003). The P-site codon of mRNA threads through the major groove of the upper portion of the helix 44, in a universally conserved region of 16S RNA. Unlike the A- and P-sites, E-site consists mostly of proteins. The main challenge of decoding is to discriminate near-cognate from cognate tRNA. When the small subunit binds to the mRNA, 16S rRNA bases interact with the base pairing of tRNA and mRNA to distinguish cognate from near-cognate codons (Carter *et al*., 2000; Ogle *et al*., 2003; Ogle *et al*., 2002). The presense of the cognate tRNA in the A-site leads to the recognition of the base-pairing geometry and induces the closure of the domains of the small subunit around cognate tRNA. In closed conformation of 30S subunit the shoulder and the head domains are rotated towards the subunit interface and helix 44 (Ogle *et al*., 2003).

1.1.2. The large subunit

The large ribosomal subunit, which sediments at 50S and has the weight of 1.45 MDa in bacteria, catalyzes the peptide bond formation and provides a path for the nascent polypeptide chain. Thirty three r-proteins were found in the large subunit (L1-L36; L for large subunit) of *E. coli* (Stelzl *et al*., 2001). All proteins are present in only one copy per ribosome except L7/L12, where L7 is the N-acetylated form of L12 which exists only in *E. coli* but not in other species (Stelzl *et al*., 2001). Together with L10 this protein appears as a pentameric complex L10(L7/L12)4 that was once called L8 before its multimeric structure was known (Stelzl *et al*., 2001; Pettersson *et al*., 1976). Also, L26 was erroneously ascribed to the large subunit, but belongs to the small ribosomal subunit and is called S20 (Stelzl *et al*., 2001).

The large subunit has a crown-like structure when seen from the side of the subunit interface surface. The three projections that radiate from surface of the particle are called the central protuberance (CP), the L7/L12 stalk on the right hand side (Strycharz *et al*., 1978; Ban *et al*., 2000) and the L1 stalk on the left hand side (Figure 2) (Lake and Strycharz, 1981; Dabbs *et. al*., 1981). The L1 stalk includes helices H75-H78 of the 23S rRNA and protein L1. The L7/L12 stalk consists of the 23S rRNA helices H42-H44 and protein L10 and multiple copies of protein L7/L12 (Zhao *et al*., 2004). Highly mobile L7/L12 C-terminal domains promote recruitment of translation factors to the ribosome and stimulate GTP hydrolysis (Diaconu *et al*., 2005). The CP is composed of helices H80-H88 of the 23S rRNA as main contribution, 5S rRNA at the top, and L5, L18 and L25 as binding proteins (Zhao *et al*., 2004; Diaconu *et al*., 2005).

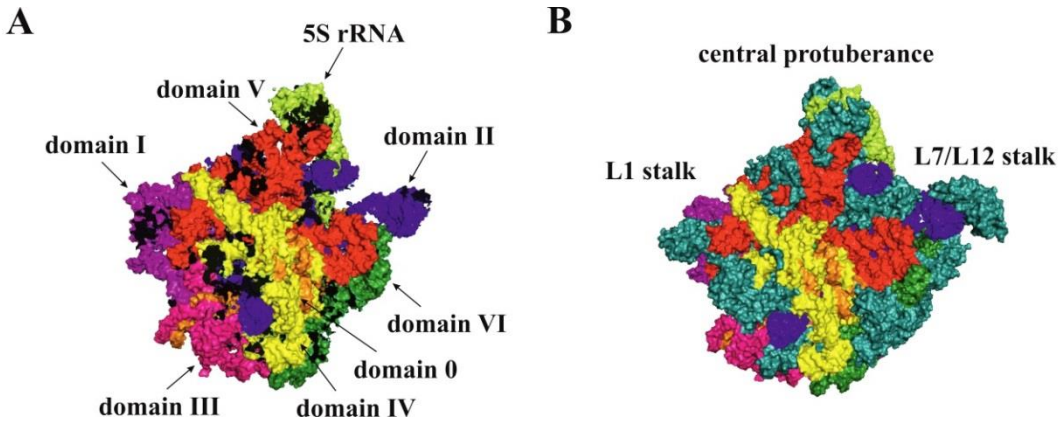

Figure 2. The *H. marismortui* large ribosomal subunit (PDN entry 2QA4). In this view, the surface of the subunit that interacts with the small subunit faces the reader. A. The tertiary structure of the rRNA in the *H. marismortui* large ribosomal subunit and its 0-VI domains. The

domain 0 (orange), the domain I (purple), the domain II (blue), the domain III (pink), the domain IV (yellow), the domain V (red), the domain VI (green). The 5S rRNA is rendered in light green. **B.** The L7/L12 stalk is to the right, the L1 stalk is to the left, and the central protuberance is at the top. The rRNA domains are color-coded as shown in the **A** panel. The large ribosomal proteins are colored cyan. Illustration was created using PyMOL DeLano Scientific software.

The two ribosomal subunits have different types of flexibility; the small subunit has interdomain flexibility, whereas in the large subunit only the protuberances are mobile. The seven domains of the 23S RNA, identified from the analysis of its secondary structures, are thoroughly interwoven (Ban *et al*., 2000; Harms *et al*., 2001; Petrov *et al*., 2013), making monolithic structure of the large subunit. Thus, in three dimensions the large subunit is a single, gigantic domain. Therefore the core of the large subunit is stable, whereas the small subunit has a flexible core. In the 50S subunit, the L1 stalk, the L7/L12 stalk, the central protuberance (CP), and the peptidyl transferase center (PTC) cleft are the most dynamic and flexible parts (Zhao *et al*., 2004).

The proteins of the large subunit are dispersed throughout the structure and mostly concentrated on its surface, except for the regions of the subunit that are of primary functional significance to protein synthesis: the peptidyl transferase active site and the flat surface that interacts with the 30S subunit (Ban *et al*., 2000; Harms *et al*., 2001). The interior of the particle is not protein-free, but it is protein-poor compared with the surface of the particle. The proteins of the 50S subunit do not extend significantly beyond the envelope defined by the RNA, except for proteins L1, L7, L10 and L11, which form the tips of the two lateral protuberances (Ban *et al*., 2000). The most surprising feature of many of these proteins is the extended, irregular structure of their loops and termini, which penetrate between RNA helices and filling the gaps between neighboring elements of RNA secondary structure (Ban *et al*., 2000). Two regions of the large subunit are particularly rich in proteins: the region binding the translational GTPase factors (L3, L6, L11, L10, L12, L13, and L14) and the external side of the polypeptide exit tunnel (L22, L23, L24, L29) (Klein *et al*., 2004). The primary role of the most proteins in the subunit appears to be stabilization of the 3D structure of its rRNA (Ban *et al*., 2000). All of the proteins in the particle except L12 interact directly with RNA. Proteins L1, L10 and L11 participate directly in the protein synthesis process (Ban *et al*., 2000).

The major function of the large subunit is to catalyze peptidyl transfer during protein elongation. This is done in the peptidyl transfer center (PTC). Here the acceptor ends of the tRNAs are stably bound close to each other, with the nascent peptide on the P-site tRNA and the incoming amino acid on the A-site tRNA. Nucleotides of the 23S RNA known to be

important for binding of the A- and P-site tRNAs were identified in a groove across the interface side of the subunit (Ban *et al*., 2000; Harms *et al*., 2001; Nissen *et al*., 2000; Schlünzen *et al*., 2001). A remarkable finding is that the N-terminus of protein L27 is very close to the acceptor ends of the A- and P-site tRNAs, suggesting a role for this protein in the peptidyl transfer reaction (Voorhees *et al*., 2009) previously suggested to be catalyzed by RNA alone (Nissen *et al*., 2000; Schmeing *et al*., 2005).

1.1.3. Ribosomal RNA

The rRNAs form the core of the ribosome and provide binding sites for the ribosomal proteins. rRNA is the main catalytic molecule of the ribosome. This is supported by the fact that catalytic sites of both ribosomal subunits are formed by rRNA, whereas ribosomal proteins are located more peripherally from the catalytic sites of the ribosome (Yusupov *et al*., 2001; Ban *et al*., 2000).

The first complete rRNA sequences were obtained from *E. coli* (Brosius *et al*., 1978; Brosius *et al*., 1980). In bacteria, the small ribosomal subunit has one rRNA molecule, and it is called 16S rRNA. In the large ribosomal subunit of bacteria, there is one small RNA molecule, called the 5S RNA, and a large RNA molecule, called the 23S RNA. The size of the corresponding RNA molecules varies among different organisms. *E. coli* 16S rRNA contains 1542 nucleotides, the 5S rRNA is 120 nucleotides, and the 23S rRNA molecule is 2904 nucleotides long (Wimberly *et al*., 2000; Ban *et al*., 2000; Noller and Woese 1981). The structure of rRNA molecules showed that the base-pairing pattern of secondary structure is generally conserved, which gave a good consensus model of rRNAs secondary structures (Glotz and Brimacombe, 1980; Glotz *et al*., 1981; Noller and Woese, 1981; Noller *et al*., 1981). These studies of rRNAs' secondary structure identified the arrangement of the rRNAs into helices and domains. The base-pared regions showed less sequence conservation than single-stranded regions, suggesting that single-stranded rRNA could carry out essential functions of ribosomal RNA. The 16S rRNA of the small ribosomal subunit has 45 helices which are denoted as h1-h45 (Figure 3). These helices fall into four different domains: the 5'-domain, the central domain, the 3'-major domain and the 3'-minor domain. These four domains extend from a central part of the small subunit – at the sites of subunit functional interactions with mRNA and tRNA – and are expected to move relative to one another during protein synthesis (Yusupov *et al*., 2001; Wimberly *et al*., 2000). The 3' end of the 16S rRNA is known to be highly flexible and contains the anti-Shine-Dalgarno sequence that base-pairs

with the Shine-Dalgarno sequence present at 5' end in many messenger RNAs (Shine and Dalgarno, 1974). The Shine-Dalgarno and anti-Shine-Dalgarno sequences interaction is critical for initiation of protein synthesis in bacteria (Schluenzen *et al*., 2000).

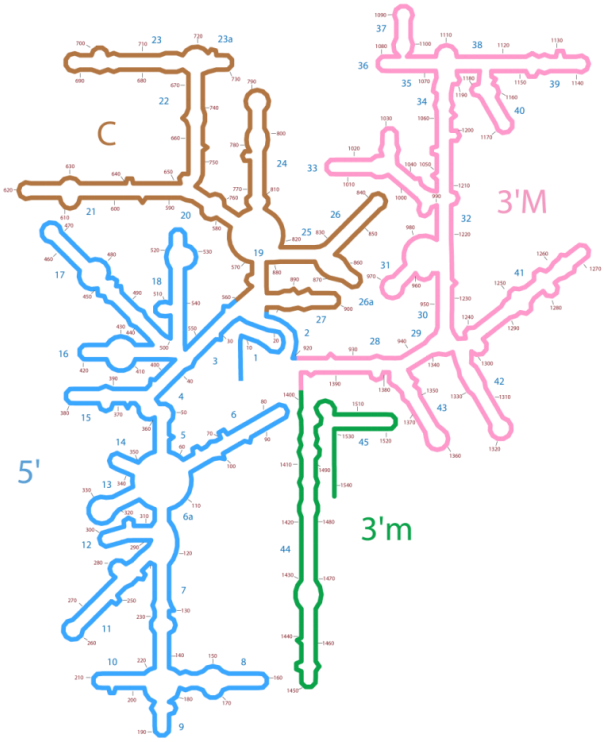

Figure 3. The secondary structure of the 16S rRNA of *E. coli*. The 5' domain (blue), the central domain (brown), the 3' major (3'M) domain (pink), the 3' minor (3'm) domain (green). The numbers of nucleotides and helices are indicated. The *E. coli* secondary structure was adapted from the Center for ribosomal origins and evolution site (http://apollo.chemistry.gatech.edu/RibosomeGallery/).

The 23S rRNA of the large subunit has 105 helices which are marked as H1-H101 (Figure 4). The secondary structure of 5S rRNA is a Y-shaped with 5 short helices. The 23S rRNA and 5S rRNA together form seven secondary structure domains of the large subunit: a central domain (Domain 0) forms the essential core of the 23S rRNA to which other six domains of the 23S rRNA are rooted. The 5S rRNA is positioned and oriented additionally to Domain 2 (Petrov *et al*., 2013). In contrast to the 16S rRNA, the domains of 23S rRNA and 5S rRNA are largely intertwined with each other, producing a compact, monolithic RNA mass (Yusupov *et al*., 2001; Ban *et al*., 2000). As a result, in spite of the complex secondary structure of the 23S rRNA, three-dimensionally the large subunit is a single, gigantic domain (Ban *et al*., 2000).

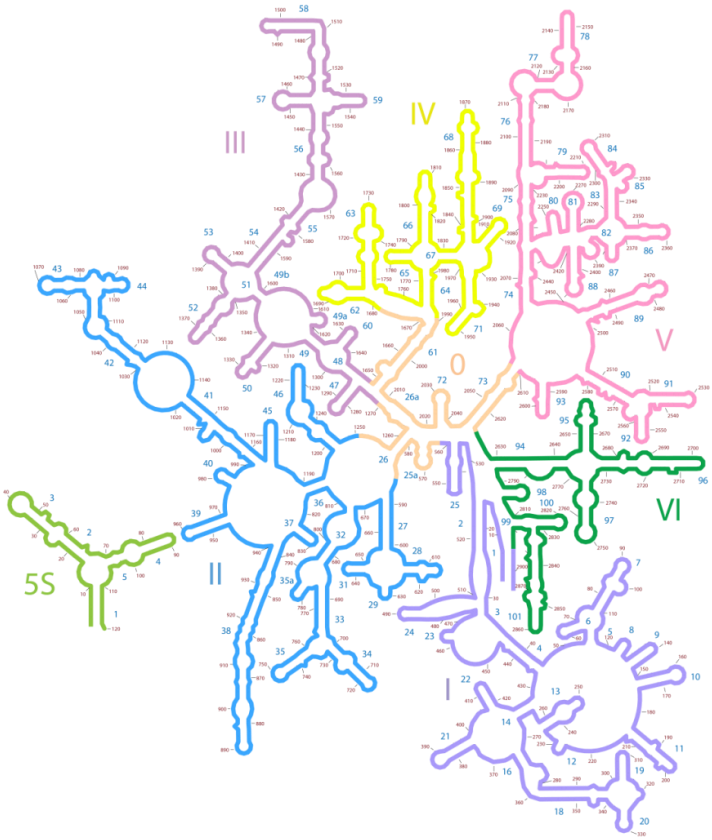

Figure 4. The secondary structure of the 23S rRNA and 5S rRNA of *E. coli*. The domains 0 (orange), I (purple), II (blue), III (violet), IV (yellow), V (pink), VI (green), the 5S rRNA (light green). The numbers of nucleotides and helices are indicated. The *E. coli* secondary structure was adapted from the Center for ribosomal origins and evolution site (http://apollo.chemistry.gatech.edu/RibosomeGallery/).

50S subunit has a number of molecular stalks made up of rRNA elements from domains II, IV, V and VI of 23S rRNA. Some of the stalks form bridges with 30S subunit, while others connect with tRNA and different elongation factors. The stalks seem to be dynamic elements of the 50S subunit (Yusupov *et al*., 2001). Divalent and monovalent cations binding the rRNA stabilize the tertiary structure of the 23S rRNA by mediating interactions between its structural domains. Bound metal ions are particularly abundant in the region surrounding the peptidyl transferase center of domain V and the conserved regions of domains II and IV of the 23S rRNA. Magnesium is essential for neutralizing the negative charge associated with the RNA phosphate backbone (Klein *et al*., 2004b).

1.1.3.1.Modifications of ribosomal RNA

Modified RNA nucleotides are chemically altered versions of the standard A, U, G and C

nucleotides (Ofengand and Del Campo, 2004). There are 36 modifications found in *E. coli* rRNAs. The most abundant being methyl group added to heterocyclic bases and ribose molecules and the conversion of uridines to pseudouridines. In total, 16S rRNA contains 11 modified nucleotides: 10 methylations and one pseudouridine; 23S rRNA contains 25 modified nucleotides: 13 methylations, 9 pseudouridines, one methylated pseudouridine ($m^3\Psi$), one dihydrouridine (D), and one 5-hydroxycytidine (ho5C) (Ofengand and Del Campo, 2004; Decatur and Fournier, 2002).

Modifications of rRNA nucleotides in bacteria are made by protein-only enzymes such as pseudouridine synthases and methyltransferases that contain both the catalytic activity for a particular modification reaction and the specificity for a cognate rRNA substrate. In *E. coli* there are 32 rRNA modification enzymes in total, 25 of them are methyltransferases and seven pseudouridine synthases. In bacteria, all pseudouridine synthases and most of the rRNA methyltransferases have been identified (Ofengand and Del Campo, 2004; Purta *et al*., 2009). It is interesting, that nucleotide alterations in eukaryotes are mediated by small nucleolar RNA-protein complexes (snoRNPs) where RNA is responsible for the site-specificity and catalysis is mediated by the protein component (Kiss-László *et al*., 1996; Tycowski *et al*., 1996; Bousquet-Antonelli *et al*., 1997). The distribution of modifications in rRNA are not random, since they are primarily concentrated at the functional center of the ribosome, such as the decoding center of the small subunit and the peptidyl transferase center of the large subunit, the peptide exit tunnel and intersubunit bridges (Figure 5) (Brimacombe *et al*., 1993; Decatur and Fournier, 2002). Three-dimensional modification maps provide additional information that most modifications correlate with regions known to be functionally important (Decatur and Fournier, 2002). This may denote that modifications influence both the function and the structure of ribosome (Brimacombe *et al*., 1993; Bakin *et al*., 1994). This clustering is conserved in organisms ranging from *E. coli* to humans, and the number of the modifications increases with the complexity of an organism (Ofengand and Bakin, 1997). Thus, it may be suggested that such post-transcriptional rRNA nucleotide modifications were required early in the evolution of the translational machinery. Three-dimensional maps show that for the most part, the modifications occur in the interior of the RNA mass, oriented towards the faces of the subunits. They are predominantly absent from areas abundantly covered by ribosomal proteins: the external surfaces and the periphery of the interface regions (Decatur and Fournier, 2002). In *E. coli*, the sites of modification in the SSU are concentrated in the area where the head, neck and upper body regions converge. The pseudouridine residues are highly concentrated in two areas, in the head and neck area, and at the bottom of the body (Decatur and Fournier, 2002). In the LSU, domains II, IV and V contain almost all

the modifications in *E. coli* and yeast. Domain V lies at the center of the subunit interface and encompasses the PTC, domains II and IV surround it (Decatur and Fournier, 2002). Another functionally important region with several modifications is the terminal stem-loop of helix 69 which contacts with helix 44 of 16S rRNA in 70S ribosome, forming the intersubunit bridge B2a (Yusupov *et al*., 2001; Schuwirth *et al*., 2005; Hirabayashi *et al*., 2006). Modifications are absent from the sarcin-ricin loop (SRL) of 23S rRNA and the lower end of the polypeptide exit tunnel of LSU (Decatur and Fournier, 2002). The SRL interacts with both initiation and elongation factors (Wriggers *et al*., 2000; La Teana *et al*., 2001). The fact that modifications are absent in this loop and in the most of the RNA surfaces peripheral to the A site suggests no positive influence of modifications on factors that binding to these rRNA elements (Decatur and Fournier, 2002). So far, no single rRNA modification has been found to be essential for ribosome function (Liang *et al*., 2009; Lövgren and Wikström, 2001; Khaitovich *et al*., 1999). Thus, it is probable that individual modifications contribute to the benefit of the ribosome function and the optimal functioning is supported by the full set of modifications.

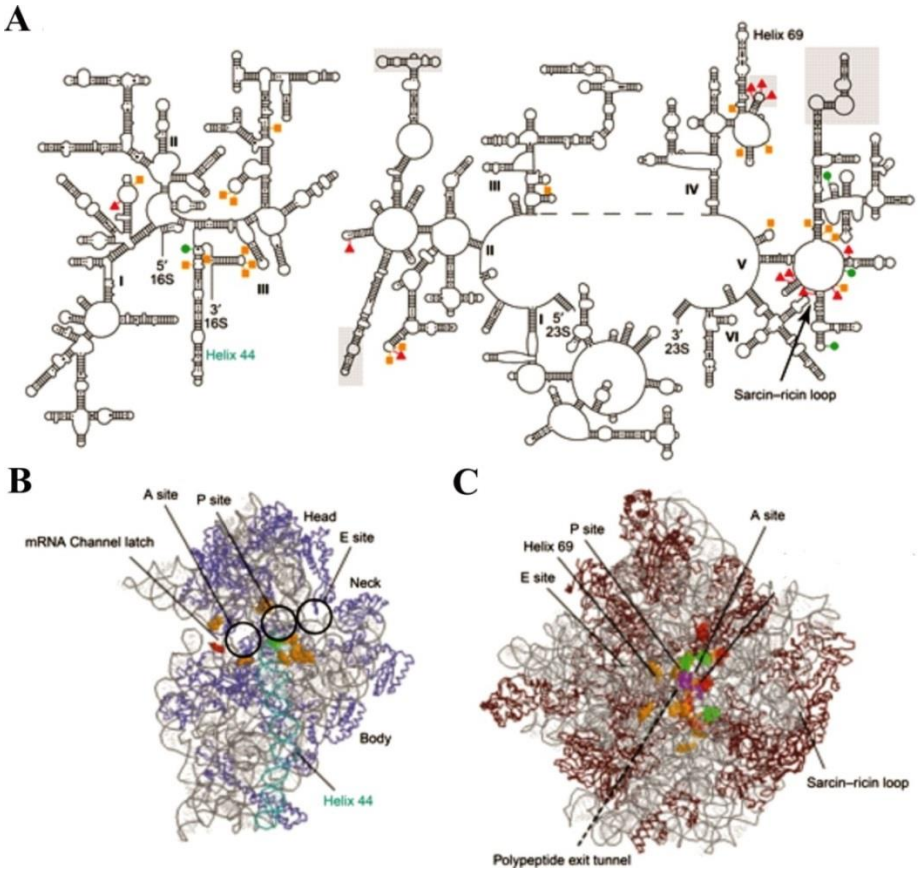

Figure 5. Distribution of *Escherichia coli* rRNA modifications in the ribosome. Nucleotides known to be modified in *E. coli* are highlighted on the *E. coli* secondary structure and in crystal structures of the small and large ribosomal subunits (SSU and LSU), derived from *Thermus*

thermophiles and *Haloarcula marismortui*, respectively. **A.** *E. coli* secondary structure was adapted from the Gutell Laboratory Comparative RNA site (http://www.rna.icmb.utexas.edu/). There is a total of 35 modified nucleotides, here divided into three groups and highlighted with different colours. The groupings include: (1) 11 pseudouridines and modified pseudouridines (Ψ, red triangles); (2) four 2'-O-methylations (Nm, green circles); and (3) "other", consisting of 19 base methylations (Nm, different types) plus one LSU dihydrouridine (orange squares). Their distributions are (SSU/LSU): Ψ, 1/10; Nm, 1/13; "other", 9/11. **B.** SSU (*T. thermophilus*; PDB entry 1FJF). The E. coli SSU contains 11 modified sites. Helix 44 is indicated in cyan. Morphological features (head, neck and body) are shown. **C.** LSU (*H. marismortui*; PDB entries 1FFK and 1FFZ).The *E. coli* LSU contains 24 modified sites. Three Ψs in helix 69 are not shown because of disorder in parts of the current crystal structure [shaded areas in panel **A**]. Functional regions are indicated for each subunit. (B, C) The modified nucleotides are distinguished by showing full atomic volume (van der Waals radii), whereas a backbone representation is used for the rRNA (grey) and protein chains (blue for SSU, maroon for LSU), and a skeleton representation for unmodified nucleotides (grey). Illustration adapted from (Decatur and Fournier, 2002).

Substrate specificity of the rRNA modification enzymes has been studied mostly *in vitro* using purified enzymes. Some modification enzymes demonstrated dependence on the presence of ribosomal proteins, while other modifications can be synthesized using the protein-free rRNA or even rRNA fragments as substrates (Ofengand and Del Campo, 2004; Siibak and Remme, 2010). With respect to ribosome *in vivo* assembly the rRNA modification enzymes can be divided into three major groups: early, intermediate, and late assembly stage-specific modifications (Siibak and Remme, 2010) (see Table 1).

Table 1. Modified nucleosides in *E. coli* rRNAs[1]

Position	Modification[2]	Enzyme[3]	Alternative name(s)	*In vitro* substrate[4]	*In vivo* assambly stage[5]
16S rRNA					
516	Ψ	RsuA	YejD	pre-SSU	early, intermediate
527	m^7G	RsmG	GidB	SSU	intermediate
966	m^2G	RsmD	YhhF	SSU	late
967	m^5C	RsmB	YhdB, Fmu, RrmB	16S rRNA	early
1207	m^2G	RsmC	YjjT	SSU	late
1402	m^4Cm	RsmH / RsmI	MraW / YraL	SSU	stochastic
1407	m^5C	RsmF	YebU	SSU	late
1498	m^3U	RsmE	YggJ	SSU	late

1516	m^2G	RsmJ	YhiQ	SSU	late
1518	$m^6{}_2A$	RsmA	KsgA	SSU	late
1519	$m^6{}_2A$	RsmA	KsgA	SSU	late
23S rRNA					
745	m^1G	RlmA	RrmA, YebH	23S rRNA	early
746	Ψ	RluA	YabO	23S rRNA	early
747	m^5U	RlmC	RumB, YbjF		early
955	Ψ	RluC	YceC	23S rRNA	early
1618	m^6A	RlmF	YbiN	pre-LSU	early, intermediate
1835	m^2G	RlmG	YgjO	23S rRNA	early
1911	Ψ	RluD	YfiI, SfhB	LSU	late
1915	$m^3Ψ$	RluD / RlmH	YfiI, SfhB / YbeA	LSU / 70S	late / very late
1917	Ψ	RluD	YfiI, SfhB	LSU	late
1939	m^5U	RlmD	RumA, YgcA	23S rRNA	intermediate
1962	m^5C	RlmI	YccW	23S rRNA	early
2030	m^6A	RlmJ	YhiR	23S rRNA	early
2069	m^7G	RlmKL	YcbY		early, intermediate
2251	Gm	RlmB	YjfH		intermediate
2445	m^2G	RlmKL	YcbY	23S rRNA	early
2449	D	*RldA*			
2457	Ψ	RluE	YmfC	23S rRNA	early
2498	Cm	RlmM	YgdE	23S rRNA	intermediate
2501	ho^5C	*RltA*			
2503	m^2A	RlmN	YfgB		early
2504	Ψ	RluC	YceC	23S rRNA	early
2552	Um	RlmE	RrmJ, FtsJ, MrsF	LSU, 70S	late
2580	Ψ	RluC	YceC	23S rRNA	early
2604	Ψ	RluF	YjbC	23S rRNA, LSU	early
2605	Ψ	RluB	YciL	23S rRNA	early

[1]Data taken from the RNA Modification Database (Cantara *et al*., 2011), Modomics – A Database of RNA Modifications (Czerwoniec *et al*., 2009), and 3D Ribosomal Modification Maps Database (Piekna-Przybylska *et al*., 2008), unless otherwise indicated.

[2] $m^x{}_yN$ refers to a methylation (m) of the rRNA nucleotide N at the x of the base position (y is the number of methylations), whereas Nm indicates a methylation of the ribose at the 2' position of nucleotide N. Ψ, D, and ho^5C are pseudouridine, dihydrouridine, and 5-hydroxycytidine, respectively.

[3] According to unified nomenclature (Ofengand and Del Campo, 2004; Andersen and Douthwaite, 2006). Modification enzymes whose genes have not been identified yet are in *Italic*.

[4] Reviewed in (Ofengand *et al*., 2001a) and (Siibak and Remme, 2010).

[5] The *in vivo* assembly stage of the modification synthesis according to (Siibak and Remme, 2010).

Consequently, seven out of 11 modified nucleotides of 16S rRNA during the assembly of ribosome seems to be a late event, in contrast, 16 out of 25 modified nucleotides of 23S rRNA are made during early steps of ribosome assembly (Siibak and Remme, 2010).

The function of rRNA modifications remains largely unknown. Sergiev and coworkers performed phylogenetic profile, genome neighborhood, co-expression, phenotype profile and co-purification data to predict functionally linked factors involved in rRNA modification (Sergiev *et al*., 2012). These analysis showed, that all rRNA pseudouridine synthases, several tRNA pseudouridine synthases and a translation termination protein RF2, formed a separate cluster (Sergiev *et al*., 2012). Earlier, it was experimentally revealed that the requirement of the *rluD* gene presence correlates with RF2 activity (O'Connor and Gregory, 2011; Gutgsell *et al*., 2005; Ejby *et al*., 2007). A surprisingly high proportion of the rRNA modification genes are co-expressed with various genes coding for transmembrane proteins, proteins involved in cell wall synthesis, transmembrane transport, etc. (Sergiev *et al*., 2012). Co-expression in a number of growth conditions might reflect similar mechanisms of gene expression regulation, which, in turn speaks in favor of co-involvement in the same functional pathway. It is believed, that rRNA modification enzymes could be involved in the assembly of ribosome and also in post-assembly regulation of ribosomal function. Thus, the role of rRNA modification enzymes is more complex than has been suggested before (Sergiev *et al*., 2012).

1.1.3.2.Pseudouridines in the ribosomal RNA

Pseudouridine (Ψ), a so-called “fifth nucleotide”, is the 5-rybosyl isomer of uridine (U). It was first isolated from yeast and its physical and chemical properties were described over fifty years ago (Davis and Allen, 1957; Cohn, 1960). Pseudouridine was the first modified nucleotide discovered in RNA and it turned out to be the most prevailing single nucleoside modification in RNA molecules. To date, it has been found in ribosomal RNA (rRNA), transfer RNA (tRNA), transfer-messenger RNA (tmRNA), and, in eukaryotes, in small nuclear (snRNA), nucleolar (snoRNA) RNAs (Del Campo *et al*., 2004; Ofengand *et al*., 2001a) and mRNA (Lovejoy *et al*., 2014). Pseudouridine does not have C-N base-ribosyl linkage, but carries unusual C-C glycoside bond, which differentiates it from other modified

nucleotides (Cohn, 1960) and gives the greater conformational flexibility due to the enhanced rotational freedom (Charette and Gray, 2000). Ψ residues are synthesized from uridine residues only after latter have been incorporated into RNA by enzymatic cleavage of the N-glycosyl bond (N1-C1'), rotation of the uracil ring 180° along the N3-C6 axis while still enzyme-bound, so that C5 replace N1 at its position, and formation C-C bond (C5-C1') (Figure 6).

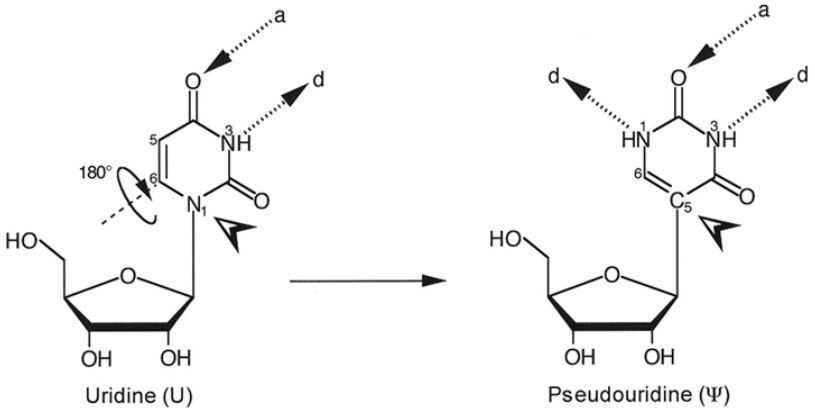

Figure 6. Chemical differences between uridine and pseudouridine. Pseudouridine synthase hydrolyses N1-C1' glycoside bond of uridine, rotates nitrogen base around C6-N3 axel and forms new C5-C1' glycoside bond. No external energy or factors are required. Pseudouridine contains one extra hydrogen bond donor and new C-C glycoside bond exhibiting higher conformational flexibility. Illustration adapted from (Charette and Gray, 2000).

The reaction is energetically favored as it goes to completion and requires no cofactor or external energy source (Ofengand *et al*., 2001a; Ofengand, 2002; Ge and Yu, 2013). In contrast to uridine, pseudouridine provides two NH imino protons which serve as hydrogen bond donors (Davis, 1995). Within the double-stranded regions, N3-H of Ψ interacts with adenosine partner through a hydrogen bond, whereas, N1-H of Ψ participates in a water-mediated hydrogen-bonding with the phosphate backbone (Desaulniers *et al*., 2008; Noeske *et al*. 2015). N1-H proton of Ψ in pyrimidine ring is involved in a stable hydrogen bond even within putative single-stranded regions. In accordance with NMR and crystallographic data, the local structure of Ψ remains similar even in different sequence contexts (Davis, 1995). Desaulniers and coworkers showed that the addition of imino protons in the loop region of the Ψ-containing helix 69 in 23S rRNA lead to increased base stacking and decreased accessibility to the solvent. Consequently, the presence of Ψ residues in helix 69 of 23S rRNA is important for the formation of a specific tertiary structure (Desaulniers *et al*., 2008). However, depending on Ψ specific location and sequence context, in single-stranded regions

it shows either destabilizing or stabilizing effect on the RNA structure (Desaulniers *et al*., 2008). It is possible that one of the functions of pseudouridine modification is a fine tuning of the RNA structure through stabilization of local ribosomal structure by locking the nucleobase in particular position with respect to the rRNA backbone (Davis, 1995; Noeske *et al*., 2015). Another important function of the substitution of Ψ for uridine is decreasing the hydrophobicity of RNA structure (Davis, 1995). Noteworthy, modified nucleosides cluster mainly around functionally important regions of the rRNA. Such clustering is not limited to *E. coli*, but is found in yeast and in human large subunit's RNA (Ofengand and Del Campo, 2004). Thus, it is speculated that Ψ contributes to the proper functioning of the mature ribosome (Ofengand, 2002). It should be mentioned, that Ψ and methyl groups have opposite influence on the molecular structure of rRNA. In most cases, methylation increases local hydrophobicity by adding a positive charge. On the other hand, as mentioned above, Ψ introduces an additional hydrophilic H-bond donor from its N1 position (Ofengand and Del Campo, 2004). Supposedly, the exact positioning of both hydrophobic and hydrophilic regions on rRNA provides specific molecular means of stabilizing the ribosome structure in the proximity of functionally important regions and/or improving interaction with the ligands of ribosome such as tRNA, mRNA and translational factors (Ofengand and Del Campo, 2004). Also, Ψ may reduce flexibility by increasing RNA-RNA contacts or by improving RNA-protein interaction (Ofengand and Del Campo, 2004). When incorporated into RNA, Ψ can alter RNA structure, increase base stacking, improve base-pairing, and rigidify the sugar-phosphate backbone (Ge and Yu, 2013).

Interestingly, Ψ artificially introduced into mRNA by eukaryotic box H/ACA RNPs can mediate nonsense-to-sense codon conversion. Uridine appears in all three stop/nonsense codons (UAA, UAG, UGA) and each uridine contacts the release factor (RF) during translation termination, probably that the uridine in stop codons is crucial for translation termination. As results show, during translation, pseudouridylated stop codons are no longer recognized by RFs. Instead, they are recognized by specific aminoacylated tRNAs (Ge and Yu, 2013). Some studies have also linked Ψ, either directly or indirectly, to human diseases, such as Alzheimer’s and Parkinson’s diseases (Ge and Yu, 2013). Kariko and coworkers show that when *in vitro*-transcribed mRNA with substitution of Ψ for uridine is introduced into mammalian cells, the translation capacity is enhanced and stability of mRNA is increased significantly. Moreover, in contrast to unmodified mRNA, mRNA containing Ψ-s does not activate cellular RNA immune sensing mechanisms (Karikó *et al*., 2008). This effect on mRNA properties has not been found for any other naturally occurring modified nucleosides. Therefore, since mRNAs containing Ψ do not activate immune system after transfection, such

mRNAs are potentially useful for clinical applications (Karikó *et al*., 2008). Although Ψ are found in virtually all ribosomes, their function still remains unknown.

1.2. Pseudouridine synthases

U to Ψ conversion is a post-transcriptional isomerization reaction performed by pseudouridine synthases which does not require ATP or any other energy source and cofactors (Koonin, 1996; Hamma and Ferré-D'Amaré, 2006), one exception being pseudouridine synthase Pus1 in Saccharomyces cerevisiae which requires zinc to maintain protein structure and catalytic activity (Arluison *et al*., 1998).

1.2.1. Families of pseudouridine synthases

All known Ψ synthase sequences from archaea, bacteria, and eukarya can be classified into six families, five of them are named after the *E. coli* enzymes RluA, RsuA, TruA, TruB, and TruD (see Table 2) (Koonin, 1996; Kaya and Ofengand, 2003; Del Campo *et al*., 2001; Gustafsson *et al*., 1996), and sixth family Pus10 present only in archaea and eukarya (McCleverty *et al*., 2007). Proteins belonging to each family seem evolved from a common ancestor (Koonin, 1996; Hamma and Ferré-D'Amaré, 2006).

Table 2. Five families of Pseudouridine Synthases in *E. coli*

Name	Substrate RNA	Modification Site	Catalytic Aspartate	N-Terminal Extension
TruD family				
TruD[1]	tRNA	13	Asp80	_
TruA family				
TruA[2]	tRNA	38, 39, 40	Asp60	_
TruB family				
TruB[3]	tRNA	55	Asp48	_
RsuA family				
RsuA[4]	16S rRNA	516	Asp102	S4-like domain
RluB[5]	23S rRNA	2605	Asp110	S4-like domain
RluE[6]	23S rRNA	2457	Asp69	S4-like domain

RluF[7]	23S rRNA	2604	Asp107	S4-like domain
RluA family				
RluA[8]	23S rRNA	746	Asp64	–
	tRNA	32		
RluC[9]	23S rRNA	955, 2504, 2580	Asp144	S4-like domain
RluD[10]	23S rRNA	1911, 1915, 1917	Asp139	S4-like domain
TruC[11]	tRNA	65	Asp54	–

References: [1] (Kaya and Ofengand, 2003); [2] (Huang *et al*., 1998b; Kammen *et al*., 1988; Arps *et al*., 1985); [3] (Gutgsell *et al*., 2000; Nurse *et al*., 1995); [4] (Conrad *et al*., 1999); [5] (Del Campo *et al*., 2001); [6] (Del Campo *et al*., 2001); [7] (Del Campo *et al*., 2001); [8] (Raychaudhuri *et al*., 1999; Ramamurthy *et al*., 1999; Wrzesinski *et al*., 1995); [9] (Huang *et al*., 1998b; Conrad *et al*., 1999); [10] (Huang *et al*., 1998b; Gutgsell *et al*., 2001); [11] (Del Campo *et al*., 2001);

In eukaryotes, uridine-to-pseudouridine isomerization is performed by site-specific Ψ synthases or by box H/ACA RNPs, which are ribonucleoproteins consisting of one unique guide RNA (H/ACA sno- or sca-RNA) and four common core proteins, Nhp2 (L7Ae in archaea), Gar1, Nop10, and Cbf5 (Ge and Yu, 2013). The uridine selection process is performed by guide RNAs which base-pair with rRNA nucleotides surrounding the target uridine, while isomerization is carried out by the Cbf5 protein (Ofengand, 2002). In contrast, uridine-to-pseudouridine isomerization reaction in bacteria is performed only by protein enzymes with both catalytic and substrate recognition activity. Eukaryotic Cbf5 protein is ~35% identical in sequence to the bacterial TruB synthase (Watkins *et al*., 1998; Hamma and Ferré-D'Amaré, 2006). Archaeal Cbf5-Nop10 complexes show that Cbf5 adopts a structure that is very similar to that of TruB (Manival *et al*., 2006; Hamma *et al*., 2005). Cbf5 and Nop10 are minimally required for Ψ synthase activity, and enzymatic activity is enhanced when Gar1 and L7Ae are added (Charpentier *et al*., 2005).

One of the largest Ψ synthase families is named after the RluA protein which is encoded by the *rluA* gene; the RluA family includes four proteins in *E. coli* (Koonin, 1996; Ofengand, 2002). RsuA is the enzyme that forms the only Ψ residue in the *E. coli* 16S rRNA, and is a prototype of another large pseudouridine synthase family named RsuA (Koonin, 1996). RluA and RsuA family enzymes are the most closely related (Koonin, 1996; Hamma and Ferré-D'Amaré, 2006). TruD synthase has little sequence homology to other Ψ synthases (Kaya and Ofengand, 2003), while structure determination of *E. coli* TruD revealed that the order of the secondary structure elements of the core domain of TruD Ψ synthases is a circular

permutation of the order in which they are present in Ψ synthases of the five other families (Hoang and Ferre-D'Amare, 2004). The most probable evolutionary scenario is that TruD diverged first from all other synthases (Hamma and Ferré-D'Amaré, 2006).

1.2.2. Structure of pseudouridine synthases

The protein sequences within each of the six families of the characterized Ψ synthases are highly conserved (Koonin, 1996; Kaya and Ofengand, 2003). Comparison of the crystal structures of the pseudouridine synthases and amino acid sequences alignment reveal that all six Ψ synthase families contain five conserved motifs: I, II, IIa, III, and IIIa (Del Campo *et al*., 2004; Kaya *et al*., 2004; Ofengand and Del Campo, 2004; McCleverty *et al*., 2007). Also, pseudouridine synthases share a core with a common fold and a conserved active-site cleft (Hamma and Ferré-D'Amaré, 2006; McCleverty *et al*., 2007). RluA and RsuA family enzymes consist of two domains, N-terminal S4-like domain and C-terminal domain, which are connected by a flexible linker, the only exception being RluA and TruC proteins which do not have S4-like domain (Mizutani *et al*., 2004). N-terminal S4-like domain is similar to the RNA-binding domain of ribosomal protein S4, a small ~ 60-amino-acid modular domain that is found in many proteins that interact with RNA (Aravind and Koonin, 1999). The larger C-terminal domain is Ψ synthase catalytic domain and contains the aforementioned five sequence motifs (Ofengand and Del Campo, 2004).

The universally conserved residue in the active site loop is an aspartate residue that essential for catalysis in members of all six Ψ synthase families (Koonin, 1996; Huang *et al*., 1998b; Ramamurthy *et al*., 1999; Del Campo *et al*., 2001; McCleverty *et al*., 2007). This aspartate may contribute to catalysis either by forming a covalent bond to C1' of the ribose, to C6 of the uracil base (Huang *et al*., 1998b) or by abstracting a proton from C2' of the ribose (Miracco and Mueller, 2011). Currently, the two mechanisms are favored where the aspartate acts on the ribose rather than the uracil (Miracco and Mueller, 2011).

Within the active site, there are also two conserved polar basic residues such as lysine or arginine, whose side chain makes a buried salt bridge with the catalytic aspartate, and an aromatic residue tyrosine (exception is the TruD family members where tyrosine is replaced by phenylalanine) (Hamma and Ferré-D'Amaré, 2006). An aromatic residue tyrosine for most pseudouridine synthases or phenylalanine for TruD in the active site is found to stack against the target uracil base likely stabilizing the conformation of the base within the active site (Hoang and Ferré-D'Amaré, 2001; Phannachet and Huang, 2004; Pan *et al*., 2003). The

tyrosine has also been proposed to act as a general base abstracting a proton from C5 to complete the isomerization process (Phannachet *et al*., 2005). Also, absolutely conserved is a hydrophobic amino acid isoleucine or valine from motif III and leucine from motif IIa, which function is probably to hold the active site in a particular conformation (Ofengand and Del Campo, 2004). Friedt and coworkers suggested that conserved arginine or lysine in catalytic site must interact with catalytic aspartate, but the strength of the interaction needs to be modulated by additional contacts such as second-shell negatively charged residues (Friedt *et al*., 2014). The catalytic aspartate, the conserved basic residue and the second-shell negatively charged residue form an electrostatic interaction network that is critical for catalysis of pseudouridylation (Friedt *et al*., 2014).

The binding of the substrate to the pseuduridine synthase causes changes in secondary structure of the substrate, following by flipping out of three nucleotides including the site of pseudouridylation, so that these nucleotides are involved in forming interactions between the Ψ synthase and substrate. The conserved arginine in Ψ synthases of the RluA, RsuA, and TruA families and conserved histidine in the TruB family probably play a key role in substrate base-flipping (Alian *et al*., 2009).

1.2.3. Substrate specificity of pseudouridine synthases

Johnson and Söll showed that the activity of pseudouridine synthase is specific for particular RNAs, being inactive on poly-uridylic acid or λ RNA (Johnson and Söll, 1970). Four of the prokaryotic Ψ synthases make Ψ only in tRNA, these are TruD, TruA, TruB and TruC pseudouridine synthases (Ofengand, 2002; Hamma and Ferré-D'Amaré, 2006). RsuA makes the single Ψ in SSU rRNA, and six pseudouridine synthases make the 10 Ψ in LSU rRNA of *E. coli* which are RluB, RluE, RluF, RluA, RluC and RluD (Ofengand, 2002). In small subunit (SSU) RNA, there is only a single Ψ in *E. coli* and *B. subtilis*, whereas there are numerous in the eukaryotes *S. cerevisiae*, *M. musculus*, and *H. sapiens* (Ofengand, 2002). Ψ predominantly are positioned in the 3'-half of the LSU of the rRNA, and only two Ψ are in 5'-half in *E. coli*, while none of the eubacteria-like RNAs have any Ψ in the 5'-half of the LSU rRNA (Ofengand, 2002). In *E. coli* RluA is specific for U746 in LSU rRNA and also specific for U32 in the four tRNAs which have Ψ at that position. There is a common sequence context in LSU rRNA and tRNA at the site of modification by RluA, both substrates share consensus sequence U/ΨUXXAAA (X can be any nucleotide) (Wrzesinski *et al*., 1995). RluB and RluF select adjacent U residues for modification with no cross-reactivity (Ofengand,

2002). TruA and RluD are capable to modify several nearby sites on the one specific RNA (Hamma and Ferré-D'Amaré, 2006). RluC and RluD pseudouridine synthases make three Ψ on 23S rRNA each (Ofengand, 2002). RluD recognize its substrate uridines in or near the loop of the helix 69, residues U1911, U1915 and U1917 (Ofengand, 2002). RluC recognition seems to be more complex since the three Ψ it makes, U955, U2504, and U2580, share neither a common sequence context nor a secondary or tertiary structure that shows common features or are close in three-dimensional space (Ofengand, 2002). The exact recognition mechanism used by RluC and RluD pseudouridine synthases is not known by now.

1.2.4. The role of pseudouridine synthases

In experiments with eukaryotic cells, ribosomes containing unpseudouridylated (or hypo-pseudouridylated) rRNAs show decreased affinity for tRNA compared to the wild type ribosomes, which results in decreased translational fidelity (Ge and Yu, 2013). Also, it was recently discovered that pseudouridylation can be induced by stress, such as heat shock and nutrient depriviation and create Ψ on positions where pseudouridine synthase do not make isomerization under normal condition (Ge and Yu, 2013). In most cases, depletion of individual Ψ in bacteria has no consequence, but rather each Ψ contributes to a cumulative effect, such as stabilizing a particular RNA conformation (Ofengand, 2002).

Deletion of six of the seven rRNA Ψ synthases individually in *E. coli* had no effect on growth over range of temperatures (Conrad *et al*., 1999; Del Campo *et al*., 2001; Raychaudhuri *et al*., 1999; Conrad *et al*., 1998), and only RluD-deficient cells show a dramatic decrease in growth rate (Huang *et al*., 1998a). Similar situation exists in yeast, where a single or even multiple deletions of series of guide snoRNAs have no effect on growth. However, when all six of the Ψ are removed simultaneously, cells display reduced growth rate and reduced protein synthetic rate, as well as hypersensitivity to antibiotics that act on the LSU, and subtle disturbances in RNA structure in the LSU (Ofengand, 2002; Ge and Yu, 2013).

In summary, depletion of pseudouridine synthases and unpseudouridylation or hypo-pseudouridylation of rRNA affects cellular growth and function in bacteria and eukaryotes.

1.2.5. Pseudouridine synthase RluD

Ribosome large subunit pseudouridine synthase D (or RluD), formerly YfiI, is a 326 amino

acid protein enzyme. It is a member of the RluA family and is responsible for uridine-to-pseudouridine isomerization in 23S rRNA at positions 1911, 1915, and 1917 in a stem-loop structure of domain IV, helix 69 (Koonin, 1996; Huang *et al*., 1998a; Raychaudhuri *et al*., 1998; Wrzesinski *et al*., 2000). Two of three pseudouridines, Ψ1915 and Ψ1917, are found in the equivalent location in the LSU rRNA of all organisms examined, which include representatives from the Prokarya, Eukarya, Archaea, mitochondria, and chloroplasts (Ofengand and Bakin, 1997). It is notable that domain IV and helix 69 of 23S rRNA interact with mRNA, tRNAs, 16S rRNA, and Ribosomal Release Factor, consequently H69 pseudouridines may be involved in proper tRNA positioning, in translocation, and in release of mRNA from the post-termination complex (Agrawal *et al*., 2004). Also, helix 69 of 23S rRNA associates with helix 44 of 16S rRNA to form bridge B2a, which plays a vital role in bridging the two ribosomal subunits and stabilizing the ribosome (Yusupov *et al*., 2001; Schuwirth *et al*., 2005; Korostelev *et al*., 2006). Moreover it directly interacts with tRNA at A and P site (Korostelev *et al*., 2006; Selmer *et al*., 2006) and deletion of H69 is dominantly lethal in *E. coli* (Ali *et al*., 2006).

The experiments showed that in solution RluD is a monomeric enzyme (Mizutani *et al*., 2004) and contains two major domains, the C-terminal catalytic domain and N-terminal S4-like domain, which are joined by a flexible linker (Figure 7) (Sivaraman *et al*., 2004). The S4-like domain is a small, modular domain found in many proteins either known or predicted to bind RNA and named after the rRNA-binding domain of ribosomal protein S4 (Aravind and Koonin, 1999; Staker *et al*., 2000). This domain is also found at the N-terminus of some RsuA and RluA family members of pseudouridine synthases that modify rRNA. However pseudouridine synthases RluA, the pseudouridine synthase family is named after, does not have an S4 domain (Del Campo *et al*., 2004). As S4 domains typically recognize helical junctions (Powers and Noller, 1995), it has been suggested that the N-terminal S4-like domain of the RluD protein binds the junction of three helices of 23S rRNA, namely H68, H69 and H70 (Vaidyanathan *et al*., 2007). Interestingly, RluD protein with truncated S4 domain is weakly active either on 50S or on free 23S rRNA, without producing any preudouridines in helix 69 (Vaidyanathan *et al*., 2007). Mizutani and coworkers showed that during examination of the RluD electron density maps, N-terminal S4-like domain appears to be disordered, although no proteolysis had occurred (Mizutani *et al*,. 2004). Absence of the S4-like domain in electron density maps was suggested to be due to the highly flexible nature of the domain.

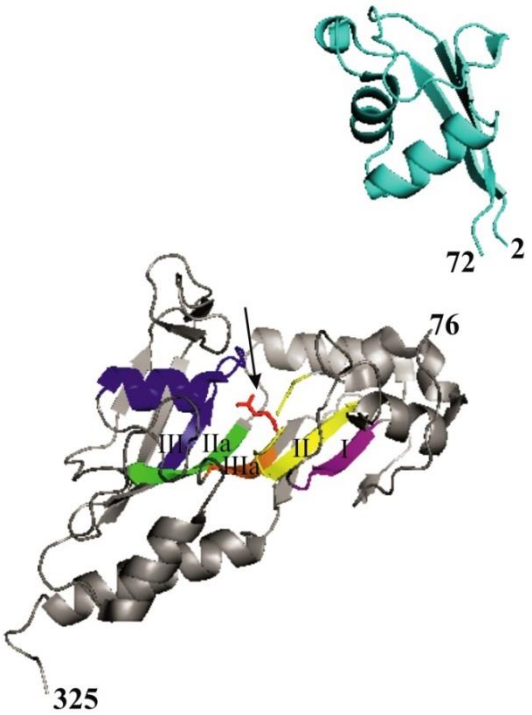

Figure 7. Tertiary structure of RluD with five Ψ synthase motifs and catalytic aspartate. Catalytic domain (grey), S4-like domain (cyan), catalytic aspartate 139 (red) indicated by arrow, motif I (purple), motif II (yellow), motif IIa (green), motif III (blue), motif IIIa (orange). Numbers are referring amino acids corresponding to S4-like and catalytic domains (without first and last amino acid in protein). Linker region is not visible due to the highly flexible nature. Illustration was created using PyMOL DeLano Scientific software.

The catalytic domain of RluD has a deep, central catalytic cleft with average dimensions 25Å long by 10Å wide by 14Å deep (Del Campo *et al*., 2004). The catalytic aspartate acid (Asp 139) which is conserved in all known pseudouridine synthases and is essential for their catalytic activity, located at the base of this catalytic cleft (Del Campo *et al*., 2004). The walls of the cleft have a positive charge and this cleft could be used to bind and position negatively charged substrate RNA (Del Campo *et al*., 2004). The active site of RluD has an overall positive charge, but the opposite face of the molecular surface is strongly negatively charged. The C-terminal subdomain of RluD which is termed the tail region is long, distinctive and highly negatively charged extension (Del Campo *et al*., 2004; Mizutani *et al*., 2004). The five C-terminal residues of tail region appear to be highly flexible (Mizutani *et al*,. 2004). In crystals, symmetry-related RluD molecules pack in such a way that the negatively charged tail region of one symmetry mate is bound in the positively charged cleft of another. The tail region of a symmetry mate covers most of the cleft and catalytic pocket of

the RluD (Del Campo *et al*., 2004). But it is suggested to be a crystal packing artifact, because in natural conditions inside the cell the cleft is needed for binding its substrate, the rRNA. The active site can accommodate only one uridine at a time, thus uridines 1911, 1915 and 1917 cannot get into the active site cavity without base flipping from their positions in the stem loop 69 (Del Campo *et al*., 2004; Sivaraman *et al*., 2004). As further experiments showed RluD synthesizes all three pseudouridines at positions 1911, 1915 and 1917 in the stem-loop 69 of 23S rRNA at a similar rate, independently of each other and without any specific order (Leppik *et al*., 2007b; Ero *et al*., 2010).

In vitro experiments with RluD have shown that although RluD synthase loses its specificity *in vitro* and modifies both free 16S and 23S rRNA at positions that are not its natural substrates, it does not recognize tRNA as a substrate (Huang *et al*., 1998a; Wrzesinski *et al*., 1995; Vaidyanathan *et al*., 2007). Thus, it still retains specificity for rRNA. The situation changes when 50S subunit is used as substrate for RluD. RluD exhibits fast and specific activity *in vitro* towards the H69 of 23S rRNA (Vaidyanathan *et al*., 2007; Ero *et al*., 2010). The reason for such low specificity *in vitro* can be the rRNA which is not correctly folded in the absence of r-proteins. Since RluD modifies uridines in H69 during late stages of the assembly of 23S rRNA into mature 50S subunits but before the 50S subunit enters the 70S pool (Vaidyanathan *et al*., 2007; Siibak and Remme, 2010; Leppik *et al*., 2007b), it may depend on some ribosomal proteins or rRNA structural motifs (Siibak and Remme, 2010; Ofengand and Del Campo, 2004; Kaczanowska and Rydén-Aulin, 2007). The specificity of RluD *in vitro* was found to correlate with the concentration of Mg^{2+}: a significant increase in activity of RluD at lower Mg^{2+} concentrations was noticed and the modification sites were "nonspecific" to RluD (Wrzesinski *et al*., 2000). Also, in contrast to the endogenous RluD purified from the cell extract (Wrzesinski *et al*., 2000), cloned and overexpressed *in vivo* His-tagged RluD was not able to modify uridine at position 1911, but could isomerize highly conserved uridines at positions 1915 and 1917 (Raychaudhuri *et al*., 1998).

In order to study the role of individual pseudouridine synthases in the cell and find Ψ residues synthesized by each of synthases, the respective synthase genes were deleted (Conrad *et al*., 1999; Raychaudhuri *et al*., 1999; Conrad *et al*., 1998; Gutgsell *et al*., 2000; Del Campo *et al*., 2001; Kaya and Ofengand, 2003; Gutgsell *et al*., 2005). As the results showed, RluD was the only synthase whose gene disruption and deletion not only blocked the synthesis of the Ψs, but also caused a dramatic decrease in growth rate, defects in ribosome assembly, biogenesis, and function in *E. coli* (Huang *et al*., 1998a; Raychaudhuri *et al*., 1998; Gutgsell *et al*., 2001; Gutgsell *et al*., 2005; Vaidyanathan *et al*., 2007). The following studies revealed that the slow growth and other defects associated with inactivation of *rluD* in *E. coli* were

restricted only to K-12 strain, and were due to a defective RF2 protein, with threonine at position 246. Inactivation of *rluD* in wild-type bacteria carrying a fully active RF2 with alanine at position 246 has negligible effects on growth, translation, termination, or ribosome subunit association (O'Connor and Gregory, 2011). Schaub and Hayes also showed that originally observed Δ*rluD* phenotypes resulted from synthetic interactions with *rpsG* (encoding ribosomal protein S7) and *prfB* (encoding RF2) alleles found within *E. coli* K-12 strains. Also, cells lacking RluD exhibit increased stop codon read-through, particularly at UGA stop codons (Schaub and Hayes, 2011). The studies of mutations in helix 69 at different positions showed that bases A1912 and U1917 are absolutely essential while a U1915C mutation results in a severe growth-defective phenotype (Liiv *et al*., 2005; Hirabayashi *et al*., 2006). However, the majority of those mutations had little influence on the activity of the pseudouridine synthase RluD at native positions. Uridines introduced by point mutations into the H69 at positions 1912, 1914 and 1919 were not isomerized to pseudouridine both *in vivo* or *in vitro* (Leppik *et al*., 2007b; Leppik *et al*., 2012), the only exception is uridine introduced in position 1916 which was modified into pseudouridine by RluD *in vitro* but not *in vivo*, and caused two-fold reduction of pseudouridine formation at all three native positions (Leppik *et al*., 2012). The base substitution A1916C did not have a significant effect on the RluD activity *in vitro* (Leppik *et al*., 2012), but the A1916U and A1916G mutations had a noticeable negative effect on the RluD activity *in vivo*, inhibiting formation of H69 pseudouridines (Leppik *et al*., 2007b). Thus, position A1916 is suggested to be an important specificity determinant for RluD in cells. Ability of RluD to convert uridines into pseudouridines is largely limited to the positions 1911, 1915, and 1917 in the H69 suggests that the uridines at those positions share structural features allowing RluD to distinguish them from uridines introduced into other positions of the H69. Crystal structure of the 50S subunit shows that all the native substrate uridines are located on the side of H69 that faces the peptidyltransferase cleft, whereas residues at the positions where uridines were artificially inserted tend to be located on the opposite side (Leppik *et al*., 2012). Also, extending the stem region of H69 by one base-pair strongly inhibits isomerization of U1911, while pseudouridines at positions 1915 and 1917 are still made by RluD, albeit at reduced level (Leppik *et al*., 2012). Thus, specificity determinants of RluD seem to be distinct for uridines at different positions of H69 and the docking site of RluD lies at least partially outside of the loop region of H69 (Leppik *et al*., 2012).

1.2.6. Pseudouridine synthase RluC

Ribosome large subunit pseudouridine synthase C or RluC, previously named as YceC, is a 319 amino-acid enzyme that belongs to the RluA family (Conrad *et al*., 1998; Huang *et al*., 1998a; Ofengand *et al*., 2001a). Pseudouridine synthase RluC converts three uridine bases in the 23S rRNA to pseudouridine and displays high homology with RluD. In contrast to the RluD, RluC is less studied. RluC modifies uridines at positions 955, 2504, and 2580 in 23S rRNA, which are widely spread along the sequence and do not appear to share any common features either on the primary or on the secondary structure level (Conrad *et al*., 1998; Ofengand *et al*., 2001a). The only visible common structural element is that all three target uridines are followed by a G residue (Conrad *et al*., 1998). In contrast to RluD, disruption or deletion of *rluC* does not affect significantly the growth of *E. coli* (Huang *et al*., 1998a; Conrad *et al*., 1998). In *in vitro* system RluC, like the RluD loses its specificity and hypermodifies both free 16S and 23S rRNA, converting an average of 1 of 18 of the uridine residues to Ψ in 23S rRNA, and ~1 of every 22 of the uridine residues of 16S rRNA to Ψ (Huang *et al*., 1998a). It may suggest that other factors in addition to ribonucleotide sequence must contribute to the *in vivo* specificity of this enzyme (Huang *et al*., 1998a). The exact mechanisms of RluC specificity still remain unknown.

In solution, RluC is a monomeric enzyme (Figure 8) (Mizutani *et al*., 2004). It contains two major domains, C-terminal catalytic domain and N-terminal S4-like domain. These domains are connected by flexible linker (Mizutani *et al*., 2004). Both crystallization studies were performed with truncated RluC missing its N-terminal domain. In the first study the full-length histidine-tagged RluC was purified, but crystallization revealed that the N-terminal residues were spontaneously cleaved by *E. coli* protease that supposedly remained associated with the His-tagged enzyme during purification (Corollo *et al*., 1999). Therefore, in the next crystallization studies Mizutani and co-workers chose only RluC catalytic domain, without S4 domain for crystallization (Mizutani *et al*., 2004).

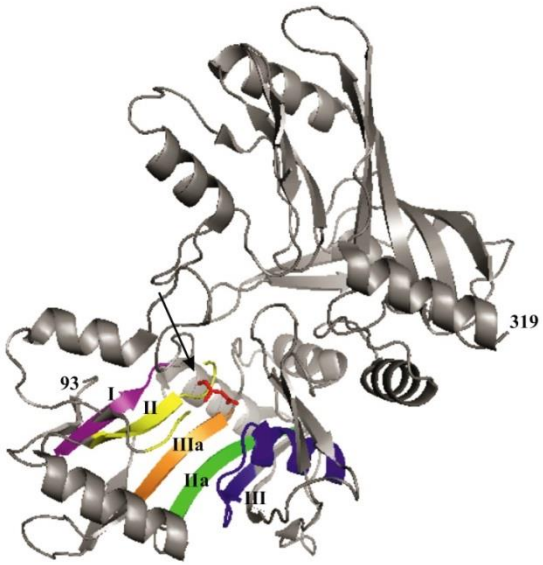

Figure 8. Tertiary structure of RluC catalytic domain with five Ψ synthase motifs and catalytic aspartate. Catalytic domain (grey), catalytic aspartate 144 (red) indicated by arrow, motif I (purple), motif II (yellow), motif IIa (green), motif III (blue), motif IIIa (orange). Numbers are referring amino acids corresponding to start and end of catalytic domain. Illustration was created using PyMOL DeLano Scientific software.

The active site and opposite face of RluC has an overall positive charge, what is in contrast with RluD which has negatively charged opposite face of the molecular surface (Mizutani *et al*., 2004). The catalytic aspartate acid 144 in the active site has an arginine on either side and forms bonds to the side chains of both arginine 142 and arginine 245 (Mizutani *et al*., 2004).

Since nucleotide modifications have been closely associated with both antibiotic sensitivity and antibiotic resistance, loss of some of these posttranscriptional modifications may affect the susceptibility of bacteria to antibiotics. Comparison of susceptibility of the bacterial strains to seven PTC inhibitors showed that the lack of most posttranscriptional modifications had a small effect on the minimal inhibitory concentrations of some of the drugs. However, cells lacking RluC enzyme showed increased susceptibility to all the tested PTC inhibitors. The lack of pseudouridine at position 2504 of 23S rRNA was found to significantly increase the susceptibility of bacteria to peptidyl transferase inhibitors (Toh and Mankin, 2008). This effect was clearly specific to the PTC-targeting antibiotics since *rluC* inactivation had no effect on the minimal inhibitory concentrations of streptomycin, an antibiotic that targets the small ribosomal subunit. The mutations at other positions specific

for RluC, such as U955C and U2580C, had only a small effect on antibiotic sensitivity (Toh and Mankin, 2008). The BipA protein of *E. coli* has similarities to the elongation factor subfamily of GTPases, including EF-Tu, EF-G and LepA. A *bipA* deletion mutant exhibits a cold-sensitive growth phenotype, which is suppressed by deletion of *rluC* gene. The suppressor effect is specific to *rluC*, as deletion of other *rlu* genes did not relieve cold sensitivity. It is also possible that more than a single pseudouridine residue is involved in cold sensitivity phenotype, as alteration of single residues did not produce suppressors. Wild-type ribosomes are dependent on BipA for efficient expression of target mRNAs and the lack of pseudouridylation at these sites seems to make the ribosomes BipA-independent (Krishnan and Flower, 2008).

In the light of the findings described in this literature overview, getting insight into the mechanisms of substrate recognition and catalytic activity of RluC and RluD pseudouridine synthases is important for the basic understanding of ribosome biogenesis and the control of gene expression at the translational level, especially in response to stress. Moreover, detailed knowledge of ribosome biogenesis could contribute to the field of antibiotic resistance and drug design.

2. Experimental part

2.1. Aims of the study

Pseudouridine synthases RluD and RluC are homologous enzymes which belong to the RluA family. RluD and RluC each convert three specific uridine bases in *E. coli* ribosomal 23S RNA to pseudouridines: bases 955, 2504, and 2580 (RluC), bases 1911, 1915, and 1917 (RluD). Both synthases have an N-terminal S4 RNA binding domain. Up to date, nothing is known about the mechanisms of specificity of RluC, how this synthase chooses its target uridines and distinguishes them from others. Also, little is known about the substrate recognition mechanisms of RluD. Furthermore, the role of S4-like domain in the specificity of both synthases is not clear. In order to shed light on the role of S4-like domain in the specificity of both RluD and RluC and asses importance of this domain working together with its native catalytic domain, chimeric proteins RluCD and RluDC were generated by the Prof. J. Remme group. In RluCD and RluDC chimeric proteins the two main domains were exchanged: RluCD protein has the catalytic domain from RluD synthase and the S4-like domain from the RluC synthase. And *vice versa*: RluDC chimeric protein has the catalytic domain from RluC and the S4-like domain from the RluD synthase. The C-terminus of both chimeric proteins carries a hexa-histidine tag (His_6). The aim of the current work was to describe the acting of chimeric proteins *in vitro*, in order to understand the role of S4-like and catalytic domains in specificity of both pseudouridine synthases.

The main objectives of the present study were:

1) to purify active chimeric proteins RluCD and RluDC from *E. coli*

2) to assess the efficiency of pseudouridine synthesis *in vitro* by RluCD and RluDC chimeric proteins on different substrates

3) to map the positions of pseudouridines in 23S rRNA produced by RluCD synthase

4) to measure the effect of magnesium cations concentration on RluCD and RluDC ability to synthesize pseudouridines *in vitro*

2.2. Materials and methods

2.2.1. Bacterial strains and plasmids

In present work, *Escherichia coli* M15 (Qiagen) and MC452 ΔΨ7 strains were used, the latter constructed by Michael O'Connor working group. The M15 strain, that carried pREP4 plasmid encoding *lac* repressor and kanamycin resistance gene, was used for the expression of pQE-60 (Qiagen) vector carrying the RluCD and RluDC chimeric pseudouridine genes (Figure 9) and for the purification of recombinant proteins. The MC452 ΔΨ7 strain was obtained from *E. coli* BW25113 strain (*Δ(araD-araB)567 ΔlacZ4787(::rrnB-3) λ⁻ rph-1 Δ(rhaD-rhaB)568 hsdR514*), where *prfB* gene (coding RF2 protein) specific for *E. coli* K-12 strain with threonine at position 246 was exchanged to *prfB* allele specific for *E. coli* B strain carrying alanine at position 246. MC452 ΔΨ7 strain had all seven pseudouridine synthase genes deleted (knocked out), as result, 23S rRNA in those cells did not contain any pseudouridines. This strain was used for the collection of ribosomal subunits fractions. The dissociated 50S large subunits of MC452 ΔΨ7 strain were used for the purification the 23S rRNA. Both dissociated 50S subunits and 23S rRNA extracted from dissociated 50S subunits were used as substrates for chimeric proteins RluCD and RluDC.

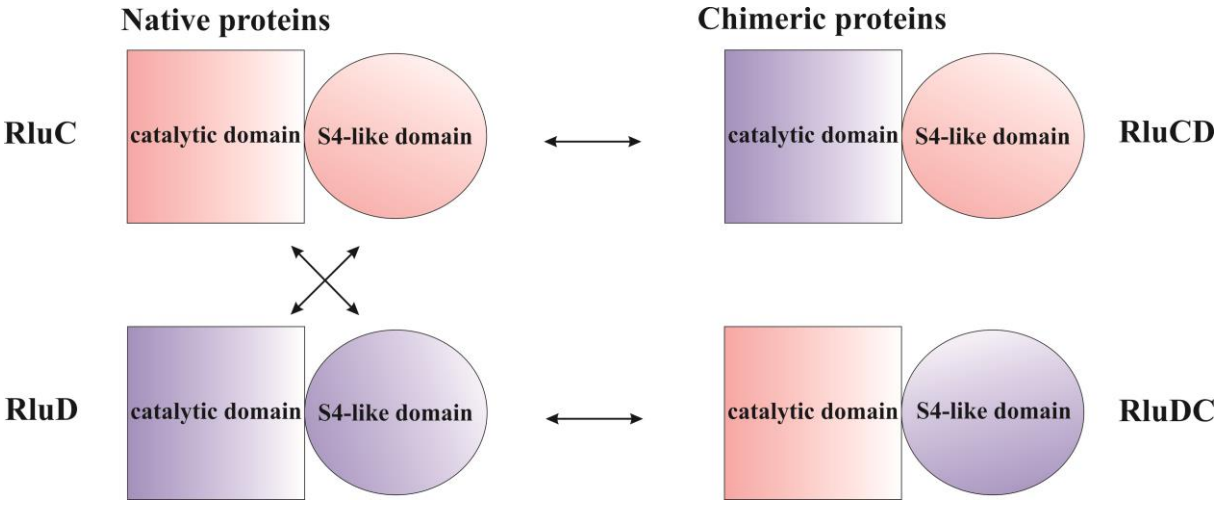

Figure 9 Schematic representation of chimeric proteins RluDC and RluCD. The RluD (blue) and RluC (red) pseudouridine synthases contain catalytic domain (rectangle) and S4-like domain (ellipse). The chimeric proteins were made in Jaanus Remme lab by combining catalytic domain and S4-like domain from two different pseudouridine synthases RluD and RluC, and fusing the chimeric enzymes into expression vector.

2.2.2. Transformation

E. coli M15 competent cells were stored at -80°C. Before transformation, cells were thawed on ice for approximately 20-30 minutes. Plasmid DNA (10-20 ng) carrying chimeric protein gene, was added to 100 µl of competent cells. The competent cell/DNA mixture was incubated on ice for 20-40 minutes. The tube was placed at 42°C for a heat shock for 1 minute (heat shock changes the fluidity of the membrane and enhances entrance of the plasmid DNA into cell), and then placed on ice for about 5 minutes. The cells were plated onto LB agar plate containing antibiotic (ampicillin 200 µg/ml) and incubated at +37°C overnight.

2.2.3. Recombinant protein purification from *E. coli*

In order to express and purify recombinant synthase proteins, *E. coli* M15 cells expressing chimeric pseudouridine synthase from pQE-60 plasmid were grown in 2 ml of 2×YT liquid medium (per 1 L of medium 16 g tryptone, 10 g yeast extract, 5 g sodium chloride) supplemented with ampicillin (100 µg/ml) and kanamycin (25 µg/ml) for the selection of the clones with both plasmids (pREP4 carrying *lac* repressor and pQE-60 carrying chimeric protein gene), and incubated at +37°C, 180 rpm in the InforceHT shacking incubator overnight. After incubation, 500 ml 2×YT liquid medium was inoculated with the overnight culture and cells were grown at 37 C° to the density OD_{600} ~ 1 in presence of antibiotic (Amp 100 µg/ml). Chimeric protein transcription from the plasmid promoter was induced with IPTG (final concentration 1 mM) and the cells were grown for additional 2 hours. Cells were collected by centrifugation at 5 000 rpm (Hettich, Rotina 420R) for 10 minutes at +4°C and re-suspended in 10 ml of buffer A (22.9 mM NaH_2PO_4, 19.8 mM imidazole, 0.5 M NH_4Cl; pH 7.4) to which lysozyme (Applichem) was added to the final concentration 1mg/ml. Cells were incubated on ice for 20 min; after incubation, cells were divided into tubes with glass beads and disrupted in homogenizer (Precellys 24) according to the manufacturer's protocol. In order to remove cellular membranes and glass beads, lysate was centrifuged at 13 000 rpm for 15 minutes at +4°C (Heraeus fresco). Protein mixture was applied to Ni-NTA column (HiLoad 16/60, Superdex 75). Column was washed with buffer A (120 ml) for protein purification and proteins of interest were eluted from the column with buffer B (50 ml) (23.3 mM NaH_2PO_4, 0.5 M imidazole, 0.25 M NH_4Cl; pH 7.5) at 1 ml/min by ÄKTA *prime plus* Liquid Chromatography System. The NaCl in buffer A and buffer B was replaced with NH_4Cl

in order to reduce flocculation of protein. Fractions containing purified protein were pooled, concentrated with Amnicon Ultra 10k filters (Millipore) for 30 minutes and washed 2 times with 2× Enzyme dilution buffer (0.1 M NaCl, 1% SDS, 10 mM $MgCl_2$, 10 mM Tris base pH 7.5). Purified protein was suspended in 1 ml of 2× Enzyme dilution buffer; glycerol was added to the ratio 1:1. Purity of the protein was assessed by SDS-PAGE. Protein concentration was determined by the Bradford method. Protein was flash-frozen in nitrogen and stored at -80°C.

2.2.4. Preparation of ribosomes

In order to collect ribosomal particles from ribosomes carrying 23S rRNA without any pseudouridines, *E. coli* MC452 ΔΨ7 strain culture was grown at +37°C in 2 L of 2×YT liquid medium without antibiotic at 180 rpm in shacking incubator. Ribosomes were isolated from exponentially growing cells at OD_{600} ~ 2. Bacteria were collected by centrifugation at 4 500 rpm (Hettich, Rotina 420R) for 20 minutes at +4°C and suspended in 25 ml ice cold 1×OVERLAY-10 buffer (20 mM Tris base pH 7.5, 100 mM NH_4Cl, 10 mM $Mg(OAc)_2$, 6 mM β-mercaptoethanol). Cells were lysed by Stansted high pressure cell disrupter (pressure 15,000 psi) at +4°C in the presence of lysozyme (Applichem) (final concentration 2 mg/ml) and DNase I (final concentration 40 units/ml) (Amresco). The lysate was cleared by centrifugation at 15 000 rpm for 20 minutes at +4°C in a Beckman Ti50 rotor. The volume of the lysate was increased to 75 ml with ice cold 1×OVERLAY-10 mM buffer. In order to clear ribosomes, a total of 17 150 U (OD_{260}) of lysate was layered onto a 20 % sucrose solution (20 mM Tris base pH 7.5, 500 mM NH_4Cl, 10 mM $Mg(OAc)_2$, 1.1 M sucrose, 6 mM β-mercaptoethanol) in 1×OVERLAY-10 buffer and centrifuged at 36 000 rpm at +4°C overnight in a Beckman Ti45 rotor ($\omega^2 t = 8.5 \times 10^{11}$). The ribosomes were washed twice and suspended in ice cold 1×OVERLAY-10 buffer. The volume was increased to 50 ml with ice cold 1×OVERLAY-10 buffer and a total of ~ 5 227 U of ribosomes were layered onto a 10% - 35% (w/w) sucrose density gradient (10% sucrose, 1× OVERLAY-10 mM $Mg(OAc)_2$, 6 mM β-mercaptoethanol; 35% sucrose, 1× OVERLAY-10 mM $Mg(OAc)_2$, 6 mM β-mercaptoethanol) in 1×OVERLAY-10 buffer followed by centrifugation at 18 000 rpm at +4°C overnight in a Beckman Ti15 rotor ($\omega^2 t = 2.0 \times 10^{11}$). Gradients were analyzed with continuous monitoring of absorbance at 260 nm (Uvis-920). Gradient fractions containing 70S, 50S and 30S were collected (Figure 10), and ribosomal particles were sedimented by centrifugation at +4°C in a Beckman Ti45 rotor overnight ($\omega^2 t = 1.2 \times 10^{12}$). Dissociated 50S

and 30S subunits were obtained by dissociating 70S ribosomes suspended in ice cold 1×OVERLAY-1 buffer (20 mM Tris base pH 7.5, 100 mM NH_4Cl, 1 mM $Mg(OAc)_2$, 6 mM β-mercaptoethanol) and layered onto a 10% - 35% (w/w) sucrose density gradient in 1×OVERLAY-1 buffer followed by centrifugation at 19 500 rpm at +4°C for 20 h in a Beckman Ti15 rotor ($\omega^2 t = 3.0 \times 10^{11}$). Fractions of dissociated 50S (D50S) and 30S (D30S) were collected, supplemented with $Mg(OAc)_2$ (final concentration in solution 10 mM $Mg(OAc)_2$). Gradient fractions containing D50S and D30S were sedimented by centrifugation at +4°C in a Beckman Ti45 rotor overnight ($\omega^2 t = 1.2 \times 10^{12}$) and washed with 10 ml of 1×OVERLAY-10 buffer. Ribosomal particles were dissolved in 1×OVERLAY-10 buffer with final concentration of 0.3U/ml, flash-frozen and stored at -80°C.

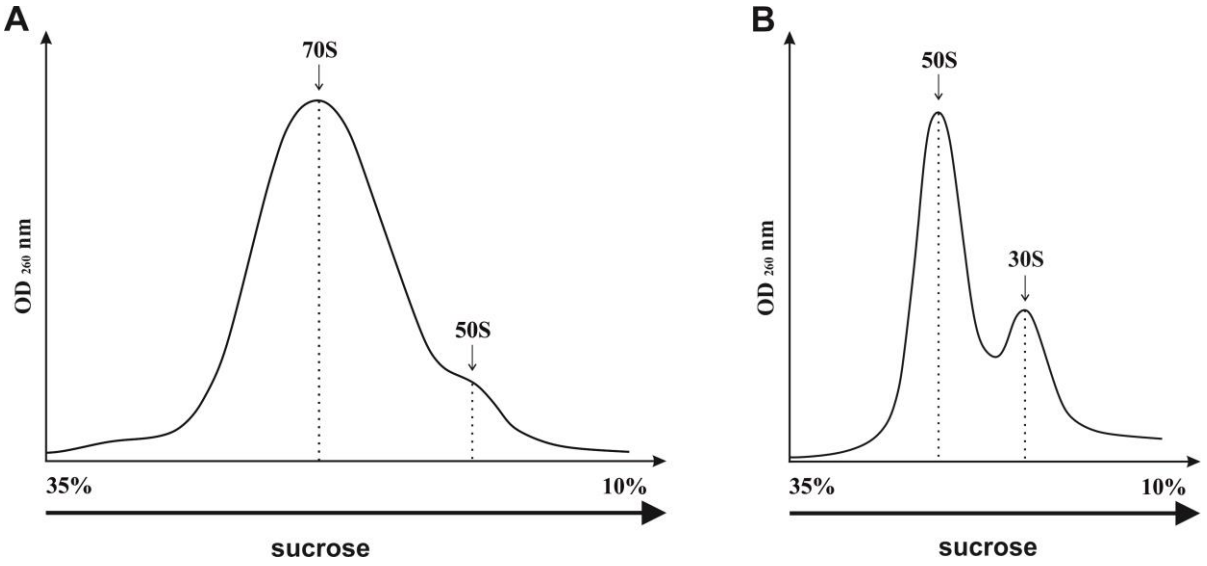

Figure 10 Separation of ribosomal particles in sucrose density gradient from *E. coli* MC452 ΔΨ7 strain. A. Separation of 70S ribosomal particles from lysate of cells. **B.** Separation of 50S and 30S ribosomal particles from 70S particles. 70S, 50S and 30S ribosomal fractions are indicated. 70S, 50S and 30S ribosomal fractions were collected and stored separately.

2.2.5. Purification of 23S rRNA from 50S ribosomal subunits

The 50S subunits dissociated from the 70S ribosomes (D50S) from *E. coli* MC452 ΔΨ7 strain were used for rRNA purification. In this work, two methods of rRNA purification were used: silica and phenol:chloroform purification. rRNA purification by silica was used to purify 23S rRNA from 50S large ribosomal subunits, 2-3 units of 50S ribosomal subunits were used, the volume was increased to 200 µl by 1× Dissociation buffer (DB) with 1 mM $Mg(OAc)_2$ (60 mM KCl, 60 mM NH_4Cl, 10 mM Tris HCl pH 8, 1 mM $Mg(OAc)_2$, 6 mM β-ME) and 800 µl

of PN buffer (Qiagen) was added for protein denaturation. The mixture was incubated at room temperature in Eppendorf shaker for 20 minutes. In order to bind rRNA from the solution, 20 µl of 50% silica was added and the mixture was incubated at room temperature in Eppendorf shaker for 20 minutes. The rRNA/silica complexes were spun down by centrifugation at 13 000 rpm for 1 minute at +4°C and the supernatant was removed. These RNA/silica complexes were then washed twice with 1 ml of 70% ethanol and dried at +37°C for ~ 10 minutes. Then, rRNA was eluted from silica in 50 µl of Milli-Q water at +37°C for 3 minutes. Silica was sedimented by centrifugation at 13 000 rpm for 3 minutes at +4°C. Supernatant with rRNA was transferred into a new tube. Finally, rRNA concentration was determined by measuring optical density at 260 nm with the spectrophotometer (NanoDrop) and rRNA was stored at -20°C.

In order to purify rRNA for the HPLC analysis after the reaction with chimeric proteins RluCD and RluDC phenol/chloroform method was used. Equal volume of phenol (pH 5.0) (Amresco) was added to the RNA mixture. The mixture was incubated at room temperature in the Eppendorf shaker for 3 minutes to mix the phases. To achieve organic and aqueous phase separation, the mixture was centrifuged at 13 000 rpm for 3 minutes at room temperature and the aqueous phase containing rRNA was transferred into a fresh reaction tube. Then, 100 µl of Milli-Q water was added to the organic phase; the mixture was placed in the Eppendorf shaker for 1 minute and then centrifuged at 13 000 rpm for 3 minutes at room temperature. The second aqueous phase was combined with the first. All the procedure was performed three times. The equal volume of chloroform (Applichem) was added to the aqueous phase and shaken in the Eppendorf shaker for 1 minute at room temperature. The mixture was then centrifuged at 13 000 rpm for 1 minute at room temperature to separate aqueous phase containing rRNA from chloroform. The aqueous phase was placed into new reaction tube and rRNA was precipitated with 2.5 volumes of 96% ethanol at +4°C for 15 minutes. The rRNA was pelleted by centrifugation at 13 000 rpm for 15 minutes at +4°C. Supernatant was removed and rRNA pellet was dried at +37°C for 2-10 minutes before being resuspended in 80 µl of Milli-Q water. rRNA concentration was measured by spectrophotometer at 260 nm and rRNA was stored at -20°C.

2.2.6. *In vitro* treatment of 23S rRNA, 50S subunit and poly-uridine oligonucleotide by purified chimeric pseudouridine synthases

The D50S large subunit, 23S rRNA (purified as described above) and poly-uridine

oligonucleotide were treated with chimeric pseudouridine synthases RluCD and RluDC at a molar ratio of enzyme to substrate 1:1 or 2:1 in 1×DB buffer with 1mM $Mg(OAc)_2$ at +37°C for 3 hours. Also, 23S rRNA was treated with wild type RluD, chimeric RluCD and RluDC at a molar ratio of rRNA to enzyme 2:1 in 1×DB buffer with three different $Mg(OAc)_2$ concentrations at +37°C for 3 hours. Then, rRNA was purified from the reaction mixture by phenol/chloroform method as described above.

2.2.7. High-performance liquid chromatography (HPLC)

HPLC analysis was used for determination of nucleoside composition of 23S rRNA after treatment by pseudouridine synthases. For HPLC analysis, rRNA fragments were prepared according to the method of Gehrke and Kuo (Figure 11) (Gehrke and Kuo, 1989); 2 units (72 pmol) (5 units for polyuridine) of phenol:chloroform purified rRNA were digested with 10 µl of P1 nuclease (200 U/ml) (Sigma Aldrich) supplemented with 10 µl of 10 mM $ZnSO_4$. The total volume was increased to 100 µl with Milli-Q water. The mixture was incubated at +37°C overnight. In order to remove phosphate group from nucleotides, the 1 µl (1 unit) of thermosensitive alkaline phosphatase (Thermo Scientific), diluted in 12 µl of 10× alkaline phosphatase buffer (Fermentas) and 7 µl of Milli-Q water, was added to the mixture with total volume 120 µl. The mixture was incubated at +37°C for 3.5 hours. rRNA nucleoside composition was determined by Reversed phase HPLC (RP-HPLC) on a Supelcosil LC-18-S HPLC column (25 cm × 4.6 mm, 5 µm) equipped with a precolumn (20 × 4.6 mm) at 30°C on a SHIMADZU Prominence HPLC system. RP-HPLC analysis was performed using the gradient conditions of Gehrke and Kuo (Gehrke and Kuo, 1989) (Figure 11). The composition of the HPLC elution buffers were as follows: (A) 2.5 % methanol, 0.01 M $NH_4H_2PO_4$; pH 5.3, (B) 20 % methanol, 0.01 M $NH_4H_2PO_4$; pH 5.1, C 35 % acetonitrile, 0.01 M $NH_4H_2PO_4$; pH 4.9.

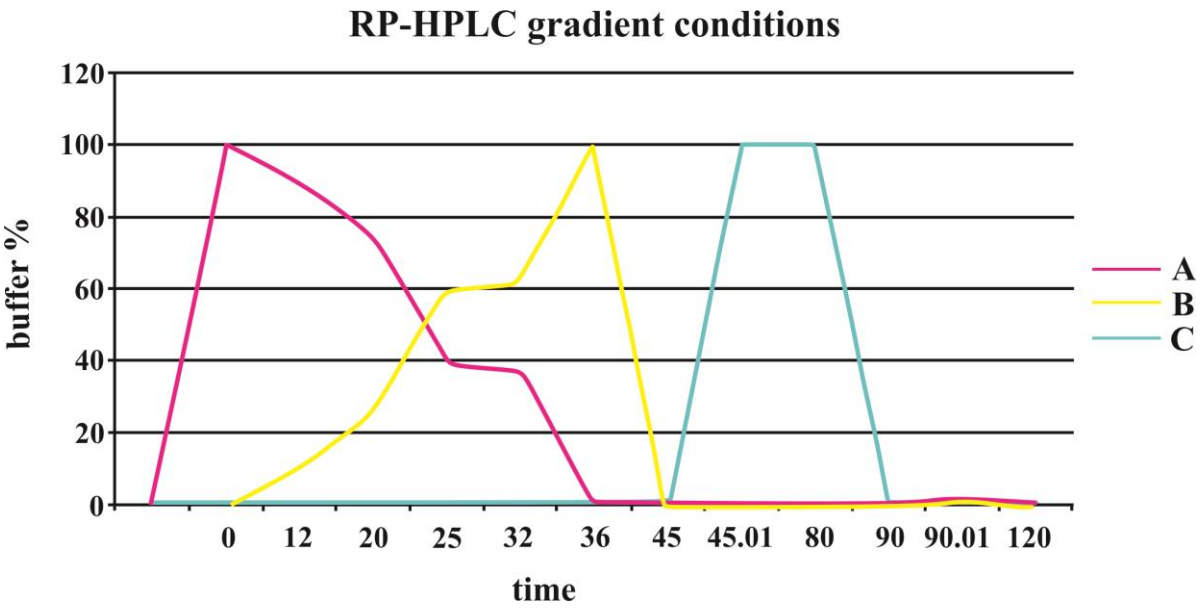

Figure 11 RP-HPLC gradient conditions. The program length is 120 min, flow-rate 1 ml/min and temperature +30°C based on Gehrke and Kuo article (Gehrke and Kuo, 1989). Different colors were used to indicate different buffers; A buffer (red), B buffer (yellow), C buffer (blue). The illustration was taken from Anneli Rander bachelor's thesis (Rander, 2009).

Nucleoside absorbance profiles were recorded at 260 nm, and peak areas were integrated. The calculation of the relative amounts of pseudouridines for 23S rRNA was based on the following formulas:

$$\Psi\text{mol} = \Psi\text{coef} \times \Psi\text{area}$$
$$\text{Umol} = \text{Ucoef} \times \text{Uarea}$$
$$\Psi = \frac{\text{U} \times \Psi\text{mol}}{(\Psi\text{mol} + \text{Umol})}$$

where Ψmol and Umol are the number of moles of pseudouridine and uridine in 1 mole of analyzed RNA; Ψcoef and Ucoef are molar extinction coefficients at 260 nm (pH 5.0) of pseudouridine and uridine (kindly provided by Christy Chow); Ψarea and Uarea are peak areas of pseudouridines and uridines respectively from chromatograms; Ψ is relative amount of pseudouridines and U is relative amount of uridines in *E. coli* 23S rRNA and 5S rRNA.
In order to calculate the relative amounts of pseudouridines for poly-uridine oligonucleotide, the relative amount of uridines in oligonucleotide was count as 100 nucleotides.
HPLC results were compared using *Student* t-test.

2.2.8. Detection of pseudouridines by CMCT/alkali treatment

The CMCT/alkali treatment of the 23S rRNA followed by the primer extension analysis was used to determine the location of the pseudouridines in 23S rRNA. The CMCT/alkali treatment was performed as described by Ofengand and coworkers (Figure 12) (Ofengand *et al*., 2001b). 20 µg of rRNA was dissolved in 20 µl of Milli-Q water. 80 µl of BEU buffer (7 M UREA, 4 mM EDTA, 50 mM Bicine/NaOH pH 8.5) and 20 µl of CMCT/BEU (1 M CMCT (SigmaAldrich) in BEU buffer) were added to the rRNA solution. In parallel, 100 µl of BEU buffer was added to the rRNA solution, serving as the negative control. Both samples were incubated at +37°C for 10 minutes for the CMCT modification of the U, G and Ψ residues. rRNA was precipitated with the addition of 2 µl of dextran, 38 µl of 4 M Na-acetate (pH 5.5) and 600 µl of ice cold 96% ethanol, and by incubation at -20°C for 10 minutes. rRNA precipitate was collected by centrifugation at 13 000 rpm for 10 minutes at +4°C. The supernatant was carefully removed, rRNA was washed twice with 1 ml of ice-cold 70% ethanol and centrifuged at 13 000 rpm for 10 minutes at +4°C. rRNA precipitate was dried at +37°C for ~ 10 minutes, then dissolved in 50 µl of NPK buffer (20 mM $NaHCO_3$, 30 mM Na_2CO_3, 2 mM EDTA, pH 10.4) and incubated at +37°C for 4 hours to allow for the removal of the CMCT group from the U and G residues. rRNA was precipitated with addition of 2 µl of dextran, 6 µl of 4 M Na-acetate (pH 5.5) and 110 µl of ice cold 96% ethanol and incubated at -20°C for 10 minutes. Then, rRNA was washed and dried as described above. Finally, rRNA precipitate was dissolved in 20 µl of Milli-Q water, rRNA concentration was determined by measuring optical density at 260 nm with spectrophotometer (NanoDrop). rRNA was stored at -20°C.

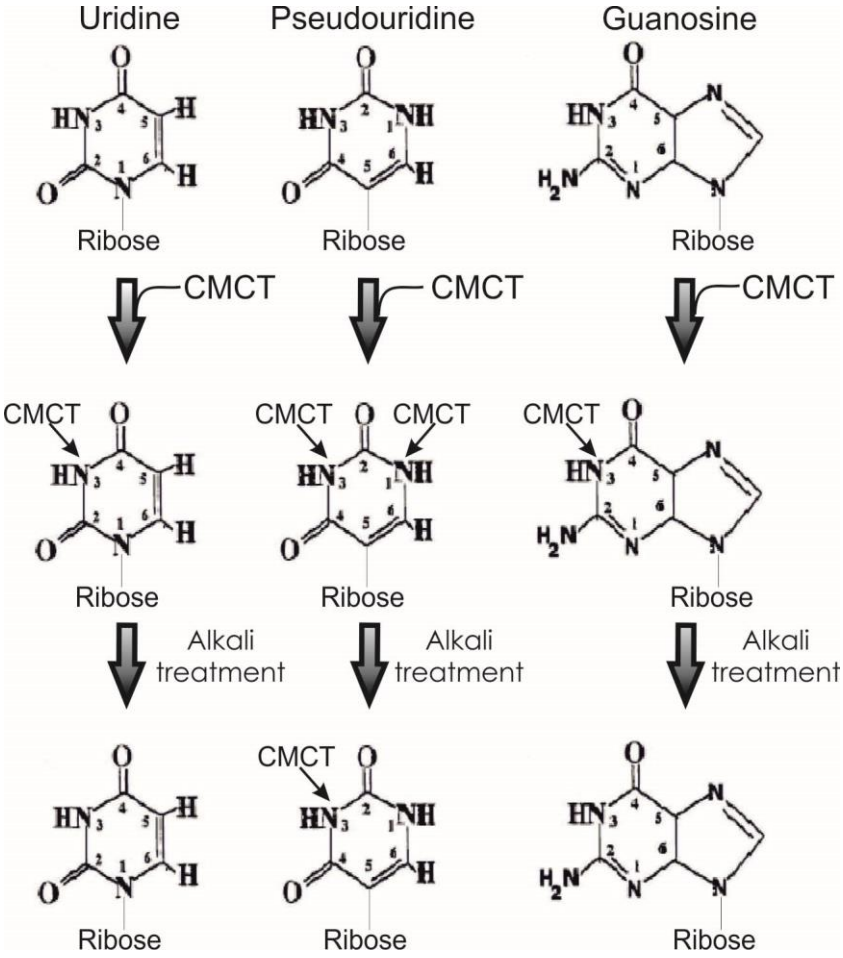

Figure 12 CMCT/alkali treatment. CMC group is added to uridine, pseudouridine and guanosine during the CMCT treatment of rRNA; the CMCT binding sites are shown with arrows, uridine and guanosine at position N3, pseudouridine at position N1 and N3; modified nucleotides are treated with alkali to allow the removal of the CMC group; CMC group remains only at position N3 of pseudouridine (Leppik, 2007b).

2.2.9. Primer extension analysis

In order to map the positions of pseudouridines in 23S rRNA produced by chimeric protein RluCD the primer extension analysis was performed after the CMCT/alkali tratment. Pseudouridine sequencing of 23S rRNA was carried out in four steps. Hybridization was carried out by using 2 pmol of primer, 1 pmol of CMCT-modified rRNA and 2 µl of 10× hybridization buffer (RT-HB) (450 mM K-HEPES pH 7, 900 mM KCl). The total volume was increased to 9 µl with Milli-Q water. The solution was incubated at +90°C for 5 minutes, and cooled down slowly to +47°C. For labeling reaction, 1.2 µl of 10× reverse transcriptase reaction buffer (10×RB) (1.3 M Tris-HCl pH 8.5, 100 mM $MgCl_2$, 100 mM DTT (DTE)), 0.8 µl of dNTP(-C) (110 mM dATP, 110 mM dGTP, 110 mM dTTP, 6 mM dCTP), 0.3 µl of [α-

^{32}P]dCTP (Hartmann Analytic) and 0.16 µl (4 U) of AMV reverse transcriptase (Seikagaku Corp., Tokyo, Japan) was added to the solution. The total volume was increased to 12 µl with Milli-Q water and incubated at 42°C for 30 minutes. 2 µl of 1 mM dNTP was added to the mixture and incubated at 42°C for 15 minutes to finish all the reverse transcriptase reactions. Reaction was stopped with 120 µl of RT-STOP solution (1 part of 300 mM NaOAc and 3 parts 96% ethanol) and incubated at -20°C overnight. The resulting DNA fragments were sedimented by centrifugation at 13 000 rpm for 15 minutes at +4°C. DNA fragments were dried at +37°C for ~ 10 minutes, resolved in 10 µl of formamide STOP (FS) (80% deionized formamide, 10 mM EDTA, 0.25% bromophenol blue), denaturated at 95°C for 5 minutes and 1 µl of material was separated on a 7% poly-acrylamide/8 M urea denaturing gel (1× TBE (90 mM Tris base, 90 M boric acid, 1 mM EDTA, pH 8.3, 8 M urea, 7 % acrylamide/bis-acrylamide, 19:1) at 2000 Volts for ~ 2 hours. The gel was transferred to Whatman 3 MM paper and vacuum-dried. Radioactivity was visualized by a Typhoon Trio Phosphorimager (Amersham Biosciences) and analyzed by ImageQuant program (Amersham Biosciences).

In order to determine the positions of pseudouridines in 23S rRNA molecule synthesized by the chimeric pseudouridine synthase RluCD, the following sequencing primers were used (see Table 3). Primers were ordered from DNA Technology A/S, GENSET, TIB MOLBIOL and Integrated DNA Tech.

Table 3. Primers for Primer extension analysis.

Primer	Oligonucleotide sequence 5' - 3'	Domains of 23S rRNA
C8	5′ TCGCCTCATTAACCTATGG 3′	**I**
C20	5′ CAGCATGTGCATTTTTGTGTACGG 3′	**I**
MINI TAQ	5′ CAAAAGGTACGCAGT 3′	**I**
C2	5′ GTCGGTTCGGTCCTCCAG 3′	**II**
C12	5′ CGCAGTTTGCATCGGGTTGG 3′	**II**
T3	5′ GCTTTCTTTAAATGATGGCTGCTT 3′	**II**
C13	5′ GCACTTCTGATACCTCCAGC 3′	**II**
C3	5′ AGTAACACCAAGTACAGG 3′	**II/III**
C14	5′ GTTTGGGGTACGATTTGATG 3′	**III**
C4	5′ ACAGTTGCAGCCAGCTGG 3′	**III/IV**

C15	5′ GGACCGTTATAGTTACGGCC 3′	**IV**
U1	5′ CAGCCTGGCCATCATTACGCC 3′	**IV**
C5	5′ TCAAGGTCGGCTCCATGCAG 3′	**IV/V**
C16	5′ GCACTAACCTCCTGATGTCC 3′	**V**
C6	5′ CTTGGGCGGTATCAGCCTG 3′	**V**
C17	5′ ACCACTTTAAATGGC 3′	**V**
C7	5′ ACACCAGTGATGCGTCCAC 3′	**V/VI**
C18	5′ GGGAGAACTCATCTCGGGGC 3′	**VI**
C19	5′ GGTTAAGCCTCACGGTTC 3′	**VI**
C11	5′ CACACACTGATTCAGGCTCTG 3′	**II**
C33	5′ GTTTGATTGGCCTTT 3′	**II**
C30	5′ CTTCGATCAAGAGCT 3′	**IV**
C31	5′ GACAAGGAATTTCGC 3′	**IV**
C32	5′ CATCATTACGCCATT 3′	**IV**

2.3. Results

2.3.1. Purification of active RluCD and RluDC chimeric proteins

In current work, recombinant chimeric pseudouridine synthases RluCD and RluDC were purified from *E. coli*, and their specificity and activity was assessed *in vitro*. His_6-tagged chimeric RluCD and RluDC enzymes were expressed from the pQE-60 vector in the *E. coli* M15 strain and purified by Ni-NTA chromatography. To achieve optimal purification, the NaCl in the wash buffer and the elution buffer was replaced with NH_4Cl in order to reduce flocculation of protein, observed in the earlier experiments. During the elution of RluCD protein two peaks were observed on chromatogram and elute was fractionated to 5 ml fractions by ÄKTA Prime. All fractions were collected separately and analyzed with the SDS-PAGE for the presence of the recombinant protein with the predicted molecular mass. The RluCD protein was present in fractions 4 and 5 (Figure 13 A and B).

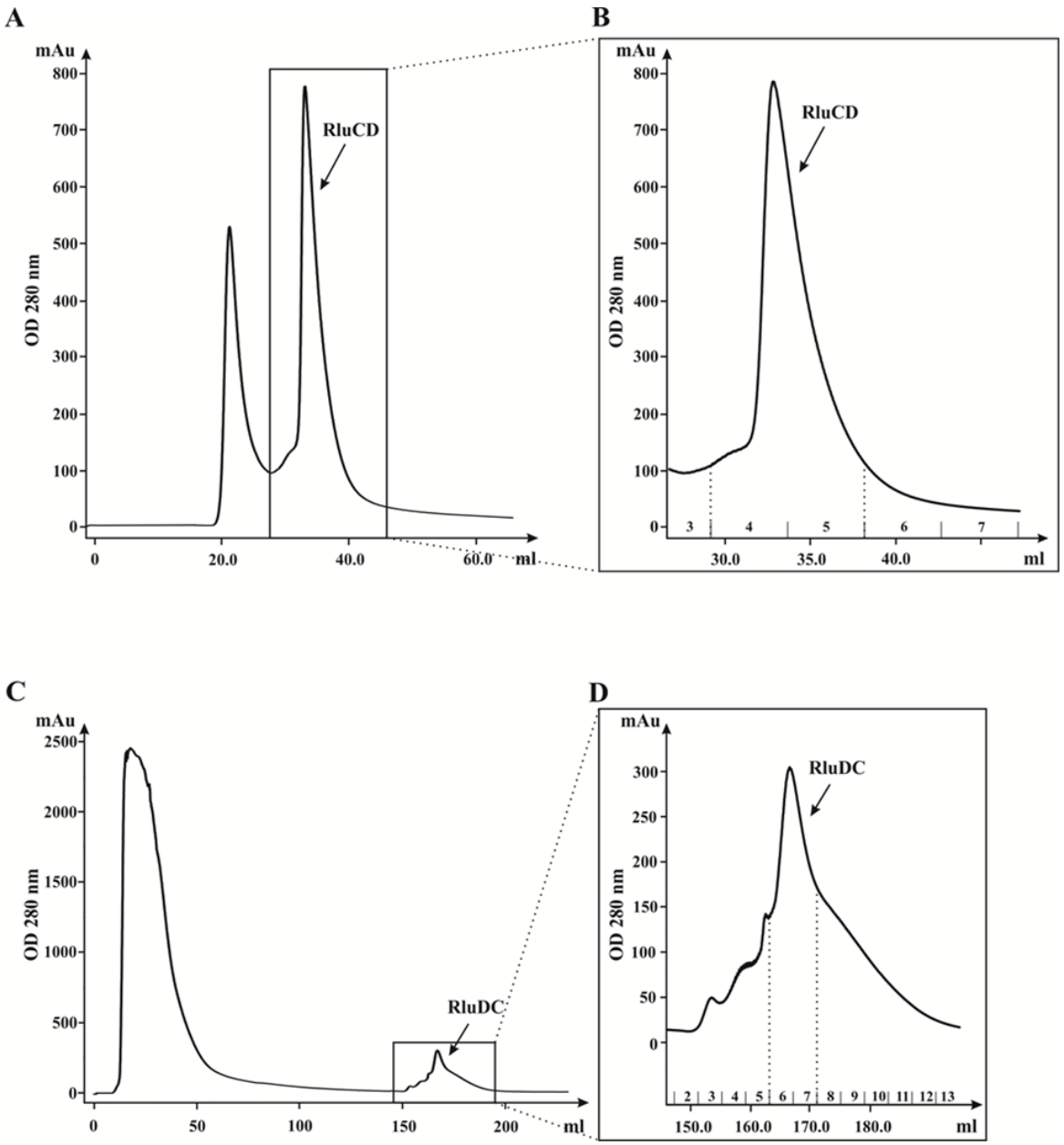

Figure 13. Purification of the RluCD and RluDC chimeric proteins by Ni-NTA chromatography. A. RluCD protein purification chromatogram. **B.** Fractions 4 and 5 contain RluCD chimeric protein. **C.** RluDC protein purification chromatogram. **D.** Fractions 6 and 7 contain RluDC chimeric protein.

The protein migrated as a single band on the gel and its mass was in agreement with the predicted molecular mass of 38.74 kDa (Figure 14). The RluDC protein was eluted in fractions 6 and 7 (Figure 13 C and D) and the protein migrated as single band on SDS-PAGE (Figure 14) in agreement with the predicted molecular mass of 36.24 kDa. The concentration of both chimeric proteins was measured with the Bradford method. Totally, 2 mg of RluCD and 4.5 mg of RluDC were purified.

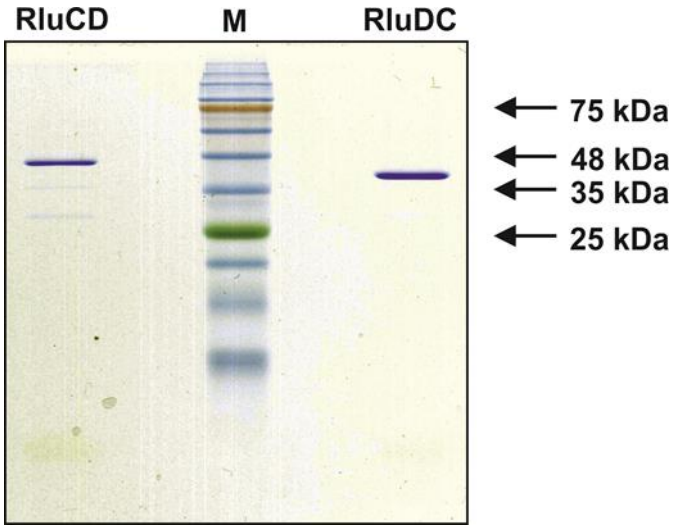

Figure 14. Detection of RluCD and RluDC chimeric proteins on Tris-Glycine SDS-PAGE. RluCD and RluDC proteins were detected in the range of 35-48 kDa, in agreement with the predicted molecular mass. Single band in eluted protein fractions is showing the purity of proteins.

In order to assess the enzymatic activity of the purified chimeric proteins RluCD and RluDC, 23S rRNA of 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was purified and used as a substrate for the enzymes. Reactions were performed with various enzyme/substrate ratios (1:1 and 2:1) and with 1 mM Mg^{2+} concentration in the reaction buffer. The nucleoside composition of the 23S rRNA after the treatment with chimeric proteins RluCD and RluDC was analyzed by HPLC (Figure 15B) as described in Materials and Methods (section 2.2.7). Since the 23S rRNA purified from 50S subunits of MC452 ΔΨ7 strains does not contain any pseudouridines, the amount of pseudouridines in 23S rRNA determined by HPLC analysis reflect the activity of chimeric proteins RluCD and RluDC on free 23S rRNA molecule. The activity was calculated using the formulas described in Materials and Methods (section 2.2.7.).

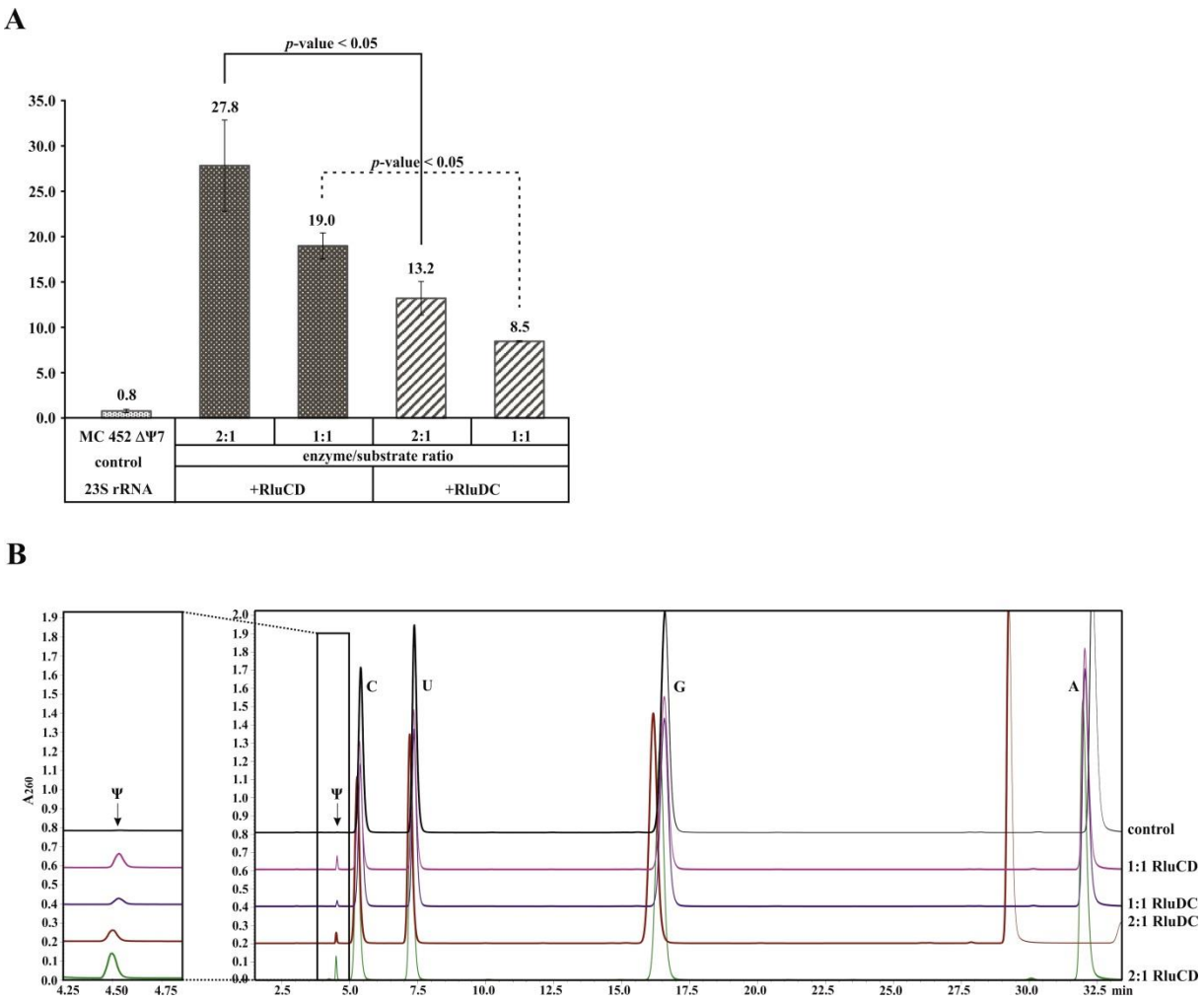

Figure 15. RluCD and RluDC enzymatic activity assessment. 23S rRNA of the 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was used as a substrate and as a control. Mg^{2+} concentration in buffer was 1 mM. For each synthase two different enzyme/substrate ratio (2:1 and 1:1) were used. **A.** Y axis indicates the relative amount of pseudouridines in *E. coli* 23S rRNA and 5S rRNA calculated from HPLC chromatograms. The RluCD and RluDC activity comparison showed significant difference between two proteins and each protein with different enzyme/substrate ratio (*p*-value < 0.05). **B.** HPLC chromatogram of nucleotides in 23S rRNA treated with RluCD and RluDC at different enzyme/substrate ratio; peaks corresponding to four standard nucleotides (C, U, G, and A) and pseudouridines (Ψ) are indicated.

As it can be seen in Figure 15A, RluCD synthase activity *in vitro* is significantly higher than RluDC (fold change 2 times, *p*-value < 0.05). Also, enzyme/substrate ratio significantly influence the activity of chimeric proteins (fold change 1.5 times, *p*-value < 0.05), since increase in the amount of pseudouridines was detected in case of enzyme/substrate ratio 2:1 in comparison with enzyme/substrate ratio 1:1 for both chimeric proteins. In summary, RluCD protein produced about 18-27 pseudouridines in 23S rRNA, in contrast the RluDC protein

which displayed two-fold less activity on 23S rRNA, modifying approximately 7-12 uridines to pseudouridines.

2.3.2. RluCD and RluDC activity *in vitro* on different substrates

In order to assess the activity of chimeric proteins RluCD and RluDC on different substrates, free 23S rRNA was purified from dissociated 50S large subunits of the MC452 ΔΨ7 strains. Also, the dissociated 50S large subunits of MC452 ΔΨ7 strains and synthetic poly-uridine oligonucleotide were used as substrates. The enzyme/substrate ratio was kept 2:1. Each type of substrate was treated by chimeric proteins RluCD and RluDC in 1×DB buffer containing 1 mM $Mg(OAc)_2$ at +37°C for 3 hours. After incubation with chimeric proteins, the 23S rRNA was extracted from 50S large subunits as well as free 23S rRNA and poly-uridylic acid were purified by phenol/chloroform method. In order to get nucleosides, 23S rRNA and poly-uridylic acid were treated with P1 nuclease and thermosensitive alkaline phosphatase. The nucleoside composition of the both type of 23S rRNAs and poly-uridine oligonucleotide after the treatment with RluCD and RluDC was analyzed by HPLC (Figure 16, panel C and E).

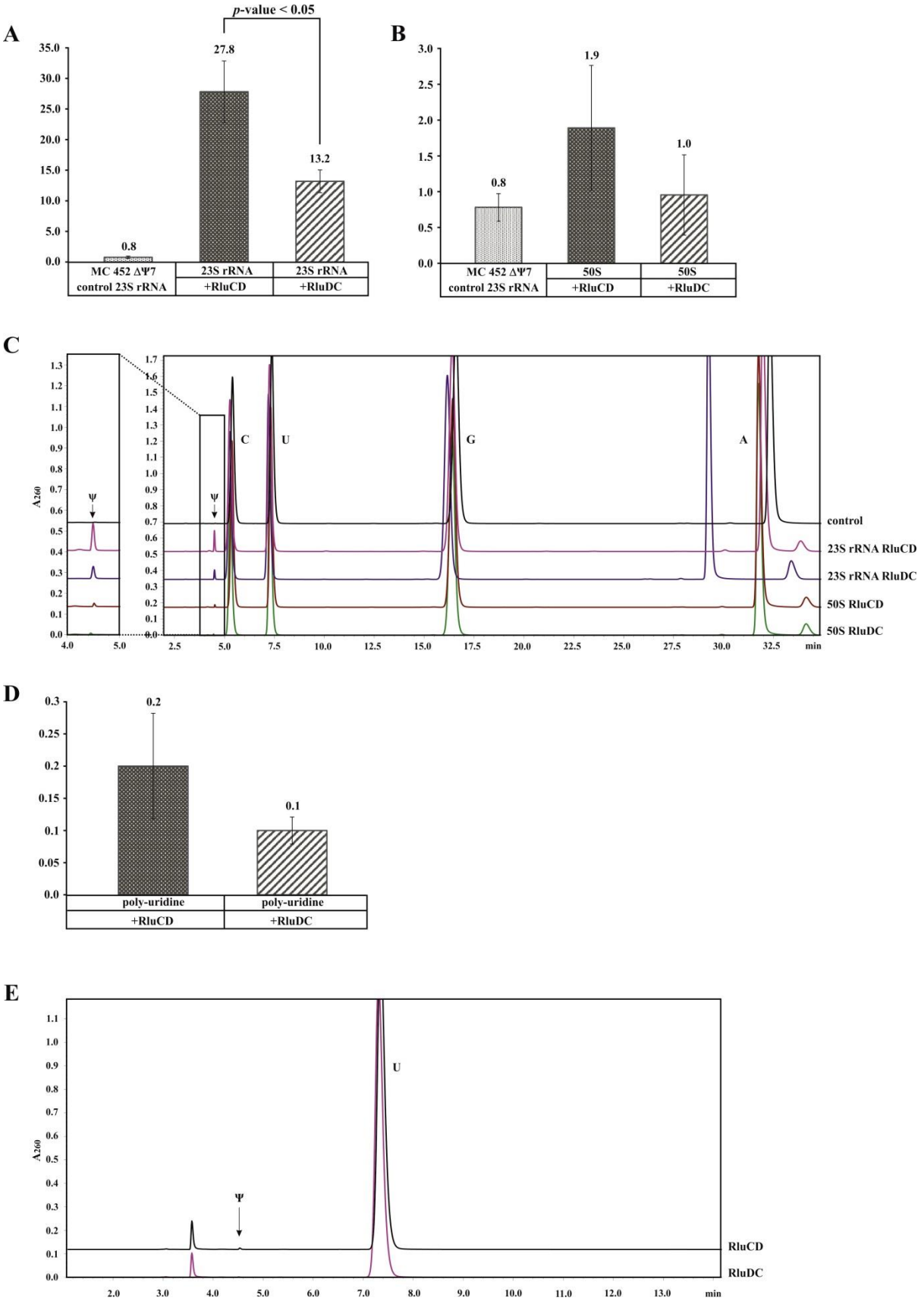

Figure 16. RluCD and RluDC activity *in vitro* on different substrates. A. 23S rRNA of 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was treated with RluCD and RluDC proteins. **B.** 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains

was treated with RluCD and RluDC proteins. The y axis shows the relative amount of pseudouridines in *E. coli* 23S rRNA and 5S rRNA calculated from HPLC chromatogram. 23S rRNA of 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was used as a control (panel A and B). **C.** HPLC chromatogram of nucleotides in free 23S rRNA and 23S rRNA purified from 50S subunit, which was treated with RluCD and RluDC; peaks corresponding to four standard nucleotides (C, U, G, and A) and pseudouridines (Ψ) are indicated. **D.** Poly-uridine oligonucleotide was treated with RluCD and RluDC proteins. The y axis shows the relative amount of pseudouridines in poly-uridine oligonucleotide conventionally consisting of 100 nucleotides calculated from HPLC chromatogram (panel C). Substrate/enzyme ratio was 2:1. Mg^{2+} concentration in buffer was 1 mM. For each chimeric protein and control standard deviations are denoted. The RluCD and RluDC activity comparison on 23S rRNA showed significant difference between two proteins (p-value < 0.05). The RluCD and RluDC activity comparison on 50S showed no significant differences (p-value > 0.05). **E.** HPLC chromatogram of nucleotides in poly-uridine oligonucleotides treated with RluCD and RluDC; peaks corresponding to four standard nucleotides (C, U, G, and A) and pseudouridines (Ψ) are indicated.

As shown in Figure 16 (A and B), RluCD pseudouridine synthase is 13 times more active on the free 23S rRNA than on the 50S subunit, producing about 27 pseudouridines on free 23S rRNA as compared to approximately 1 pseudouridines on 50S large subunit. The activity on the free 23S rRNA of both pseudouridine synthases differs, with RluDC producing about twice less pseudouridines than RluCD synthase (Figure 16 A). Also, RluDC was not active on the 50S large subunit and did not produce pseudouridines in 23S rRNA in comparison with 1 pseudouridine in case of RluCD (Figure 16 B). In order to measure the competence of RluCD and RluDC to convert uridines to pseudouridines outside of sequence context and structure specific for ribosomal RNA, the poly-uridine oligonucleotide was used as a substrate (Figure 16 D). In Figure 16 (panel D) can be seen that there are 0.2 and 0.1 pseudouridines from 100 uridines isomerized by chimeric pseudouridine synthases RluCD and RluDC respectively, which comprise 0.2% and 0.1 % of uridines isomerized to pseudouridines. The experiment showed that both RluCD and RluDC do not produce any pseudouridines on the poly-uridine oligonucleotide substrate.

2.3.3. Activity of RluD, RluCD and RluDC at various Mg^{2+} concentration *in vitro*

In order to study if the enzymatic activity of chimeric proteins RluCD and RluDC as well as enzymatic activity of native pseudouridine synthase RluD depends on Mg^{2+} concentration

during the reaction, the activity of above-mentioned pseudouridine synthases was tested *in vitro* on 23S rRNA with different Mg^{2+} concentrations. Three different $Mg(OAc)_2$ concentrations in 1×DB buffer were used: 1 mM, 10 mM, and 20 mM. The 23S rRNA extracted from dissociated 50S subunits of MC452 ΔΨ7 strain was used as a substrate. The native pseudouridine synthase carrying a His_6-tag was purified in our laboratory by the colleagues.

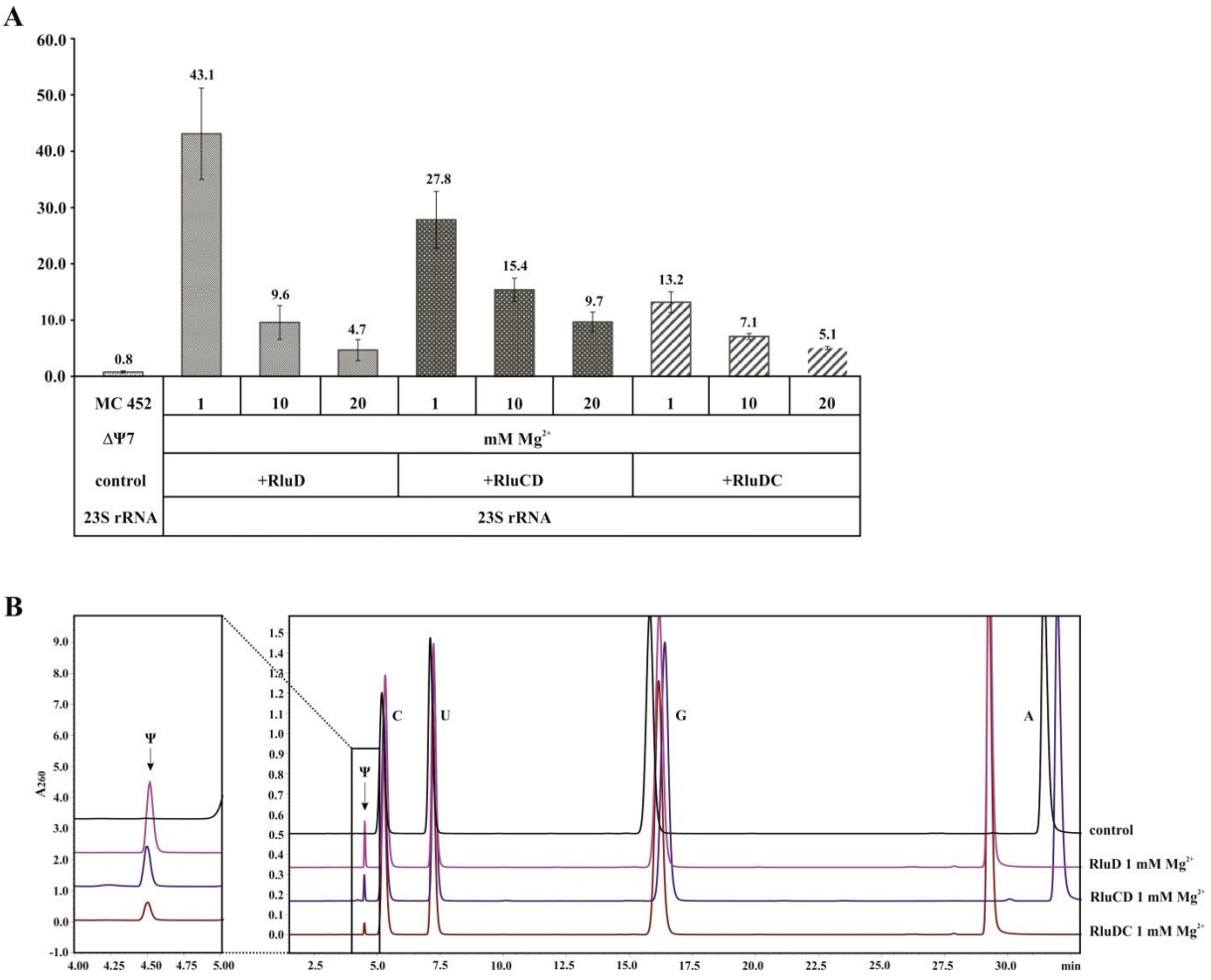

Figure 17. Effect of Mg^{2+} concentrations on RluD, RluCD and RluDC synthases activity. A. 23S rRNA of 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was used in all reactions and as a control. Substrate/enzyme ratio was 2:1. Reactions were performed at 1 mM, 10 mM and 20 mM Mg^{2+} concentrations for each pseudouridine synthase. The RluD, RluCD and RluDC activity comparison at all mM Mg^{2+} concentration showed significant difference between all proteins (p-value < 0.05). Only activity of RluD and RluDC at 10 mM Mg^{2+} and 20 mM Mg^{2+} concentrations showed no significant difference (p-value > 0.05). **B.** HPLC chromatogram of nucleotides in 23S rRNA treated with RluD, RluCD and RluDC at 1 mM Mg^{2+}; peaks corresponding to four standard nucleotides (C, U, G, and A) and pseudouridines (Ψ) are indicated.

As can be seen from Figure 17, RluD had the highest activity at 1 mM $Mg(OAc)_2$ concentration as compared to other Mg^{2+} concentrations for the same protein or pseudouridine synthases RluCD and RluDC. The native RluD isomerized about 42 pseudouridines at 1 mM $Mg(OAc)_2$ concentration, 9 pseudouridines at 10 mM $Mg(OAc)_2$ concentration, and 4 pseudouridines at 20 mM $Mg(OAc)_2$ concentration. As can be seen from this data, activity of RluD synthase decreased with increasing of $Mg(OAc)_2$ concentration. This pattern was observed also for RluCD and RluDC. RluCD pseudouridine synthase modified on average 27 pseudouridines at 1 mM $Mg(OAc)_2$ concentration, 14 pseudouridines at 10 mM $Mg(OAc)_2$ concentration, and 9 pseudouridines at 20 mM $Mg(OAc)_2$ concentration. In the same manner RluDC produced at the average 12 pseudouridines at 1 mM $Mg(OAc)_2$ concentration, 6 pseudouridines at 10 mM $Mg(OAc)_2$ concentration and 4 pseudouridines at 20 mM $Mg(OAc)_2$ concentration. The activity of RluCD was approximately twice higher than activity of RluDC on 23S rRNA at all $Mg(OAc)_2$ concentrations of 1×DB buffer. For both RluCD and RluDC, 10-fold increasing of $Mg(OAc)_2$ concentration was correlated with about 2-fold decrease in chimeric pseudouridine synthases activity. The $Mg(OAc)_2$ concentration increasing from 10 mM to 20 mM induced approximately 1.5-fold decrease of enzymatic activity. In case of native pseudouridine RluD, increasing of $Mg(OAc)_2$ concentration from 1 mM to 10 mM induced approximately 4.5-fold decreasing of pseudouridine synthase activity and 2-fold decreasing of enzyme activity when $Mg(OAc)_2$ concentration was increased from 10 mM to 20 mM.

2.3.4. Mapping of uridines isomerized by RluCD chimeric synthase in 23S rRNA *in vitro*

The aim of this work was not only to purify and test the activity of the chimeric proteins, but also to find the positions of pseudouridines, which were isomerized by chimeric synthases. The RluCD pseudouridine synthase was chosen for this experiment, because of its high activity *in vitro*. Moreover, similar experiments with RluCD were initiated *in vivo* in our lab. Mapping of the positions that were modified by RluCD provided insight into the role of the S4-like domain and catalytic domain in the specificity of pseudouridine synthase RluD. For this analysis, the 23S rRNA extracted from dissociated 50S large subunits (D50S) of MC452 ΔΨ7 strain was used as substrate. After incubation with RluCD *in vitro*, 23S rRNA was purified by phenol/chloroform method and treated with CMCT, followed by alkaline treatment to remove CMC from uridines and guanosines, leaving the CMC modification only

on pseudouridines. The sequencing was performed by primer extension analysis using reverse transcriptase. Different primers (see Table 3) were used in this analysis, covering the whole 23S rRNA molecule for sequencing. From this analysis, 27 positions of pseudouridines were identified (Figure 18 and Figure 19).

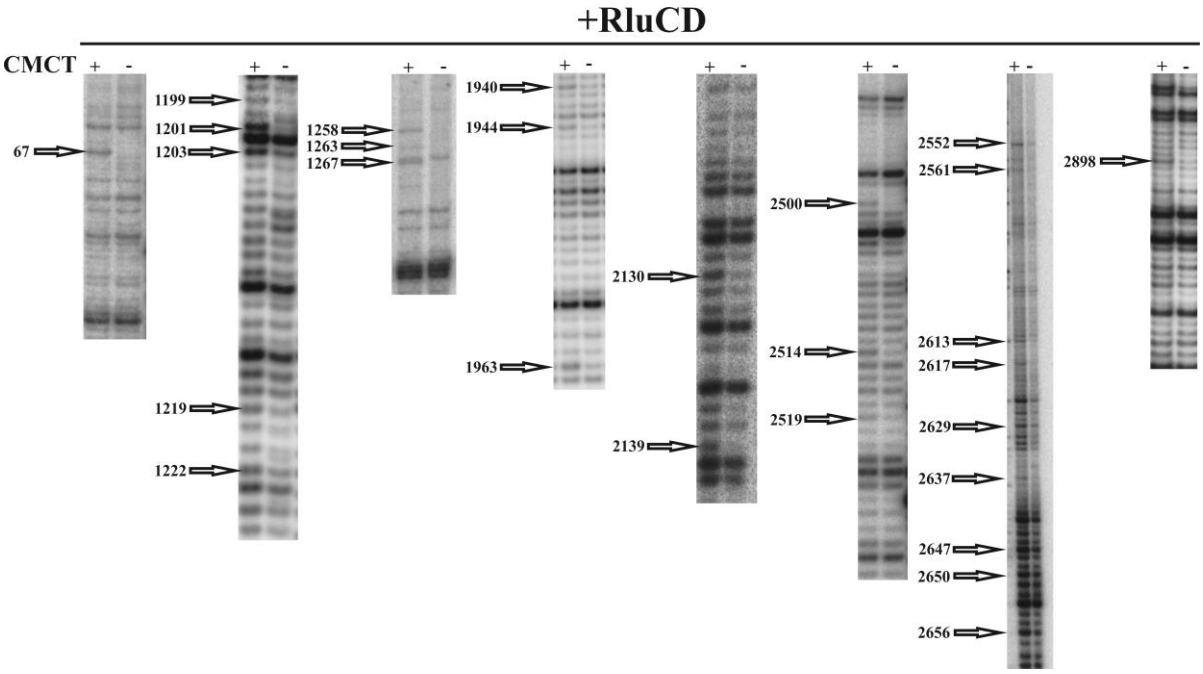

Figure 18. Pseudouridine positions on 23S rRNA after *in vitro* treatment with RluCD. 23S rRNA of 50S large subunit dissociated from 70S ribosomal subunits of MC452 ΔΨ7 strains was incubated with RluCD for 3 hours and treated with CMCT, followed by alkaline treatment to remove CMC from uridines and guanosines. The 27 positions of pseudouridines were identified. The "+" indicates 23S rRNA treated with CMCT/alkali; "-" indicates untreated 23S rRNA. Pseudouridines are at positions containing band on "+" CMCT line which is absent in "-" CMCT line. Bands corresponding to the modified positions on 23S rRNA are indicated with arrows.

In Figure 18 there are only positive results on gel are shown, the part of gel where pseudouridines are not present is not shown. As it can be seen in Figure 18, pseudouridines are at positions which contain band in the CMCT present line, and do not contain band in the line where CMCT absent. For example, at position 67 the band is present only in the "+" CMCT line, and absent in the "-" CMCT line. If bands appear in both CMCT "+" and CMCT "-" lines on gel, it is result of unspecific stops of reverse transcriptase during primer extension reaction caused by rRNA secondary structure. The RluCD did not isomerize either of RluD-specific uridines at positions 1911, 1915 and 1917 or RluC specific uridines at positions 955, 2504 and 2580.

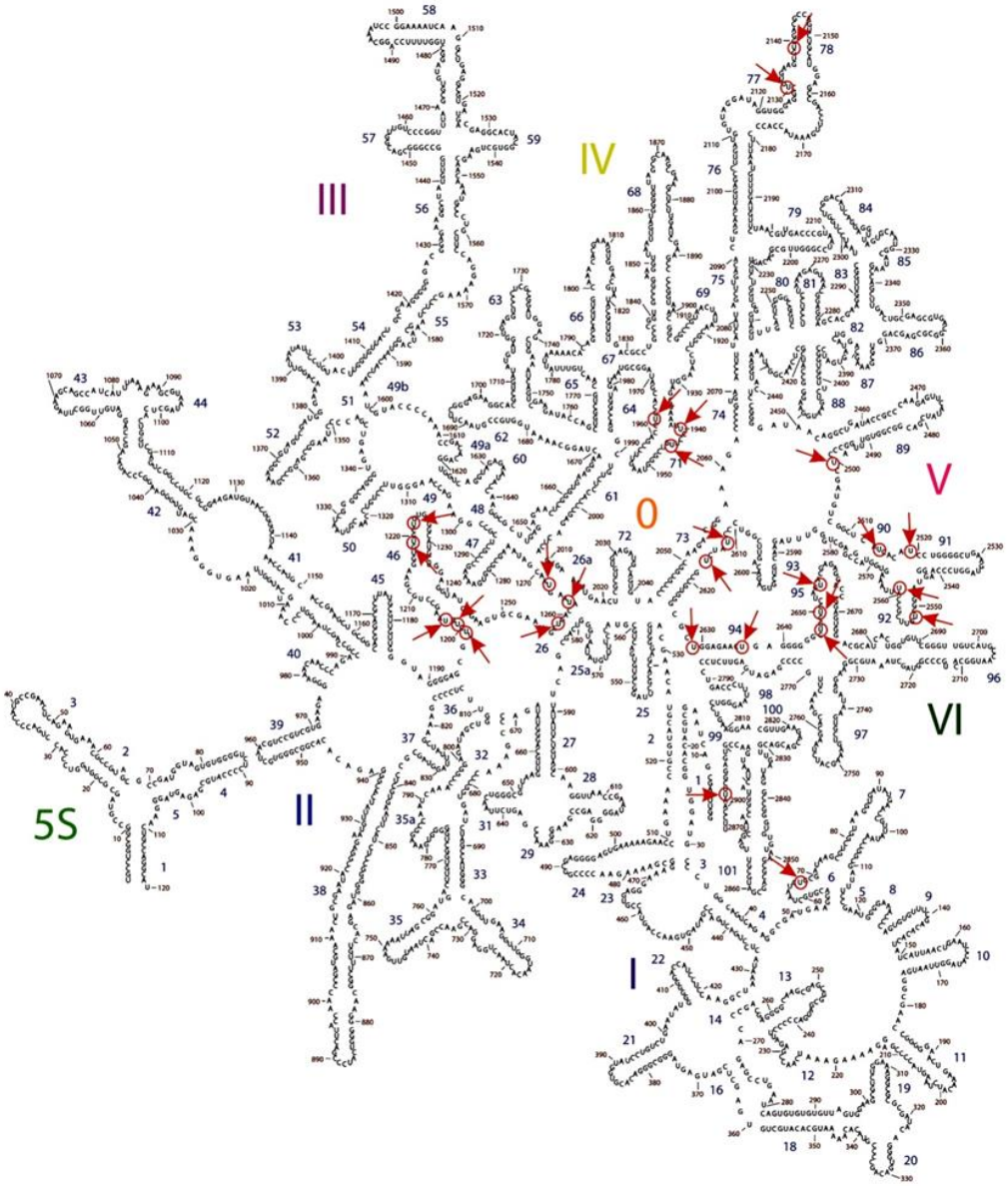

Figure 19. Mapping of identified pseudouridines on the secondary structure of the 23S rRNA after RluCD pseudouridine synthase treatment. 27 pseudouridines were mapped to the secondary structure of the 23S rRNA. RluCD specific pseudouridines were detected in all domains, with III domain being the only exception.

Figure 19 shows, that there are 2 positions modified in the domain I, 5 positions in the domain II, 3 positions in the domain IV, 7 positions in the domain V, 5 positions in the domain VI, and 5 positions in the domain 0. However, no pseudouridines were isomerized in the domain III. The modification sites were on helices and loops, two to three adjacent pseudouridines were

synthesized more frequently than single (see Supplement 1).

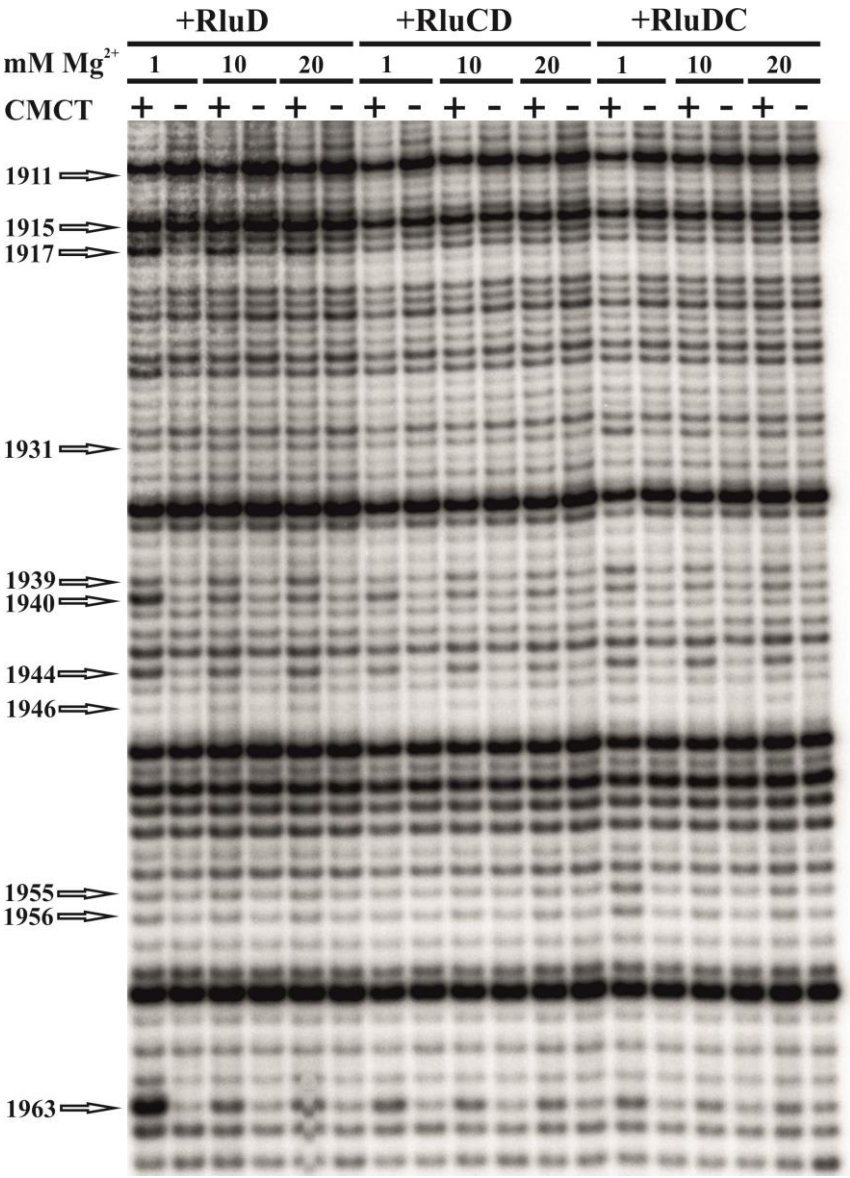

Figure 20. Pseudouridine positions on 23S rRNA after treatment with RluD, RluCD and RluDC synthases at different Mg^{2+} concentrations in reaction buffer. Bands corresponding to the modified positions on 23S rRNA indicated with arrows. 23S rRNA treated with CMCT/alkali was indicated with "+"; untreated 23S rRNA was indicated with "-". Gel is shown only with positive results.

In addition to total sequencing of 23S rRNA after treatment with chimeric protein RluCD, the incomplete sequencing of domain IV in 23S rRNA using U1 primer (see Table 3) were performed for RluD, RluCD and RluDC proteins at different Mg^{2+} concentrations (Figure 20). Sequencing using U1 primer covers region with helix 69 and the results provide information about isomerization of 1911, 1915 and 1917 positions in 23S rRNA. In Figure 20 it can be seen that RluD modified its native position 1917, but not position 1911 at different Mg^{2+}

concentrations. For the position 1915 the results of the primer extension analysis were difficult to interpret, because a strong stop signal on the gel was detected in both lines, "+" and "-" CMCT. Furthermore, RluD modified other uridines in helix 71 and near this region, which are not its natural substrates. In the case of RluCD and RluDC, another picture can be observed; both of them could not modify uridines neither at position 1917 nor 1911. Positions 1939, 1940, 1944, 1946 and 1963 were modified by RluD, RluCD and RluDC. Position 1955 and 1956 were modified by RluD and RluDC. Position 1931 was modified only by RluDC. For some pseudouridine positions band on gel was stronger at 1 mM Mg^{2+} concentration than at higher concentrations. For example, positions 1940 and 1963 modified by RluD have stronger band at 1 mM Mg^{2+} concentration than at 10 mM and 20 mM Mg^{2+} concentration. Also, at the same Mg^{2+} concentration pseudouridine bands on gel have different brightness, as it can be observed for positions 1201, 1203, 2130 and 2139 modified by RluCD. The bands at these positions are more evident than at other positions.

2.4. Discussion

Chimeric pseudouridine synthases RluCD and RluDC exhibit high activity and 23S rRNA-substrate specificity *in vitro*

In the presented study, chimeric synthases with exchanged RNA binding S4-like domains and catalytic domains were designed to dissect the role of each domain in activity and specificity of the bacterial RluC and RluD synthases. The presented work is the first report on the activity and specificity of recombinant chimeric pseudouridine synthases RluCD and RluDC *in vitro*. The first step of the analysis was to express in *E. coli* M15 strain and to purify active chimeric RluCD and RluDC recombinant proteins and test their activity on different substrates. Recombinant chimeric proteins were successfully expressed in *E. coli* and purified with the predicted molecular mass and did not co-purify with any other protein. As results showed, exchanging of the S4 domains as well as purification process did not affect the activity of these synthases, since they were capable of binding to 23S rRNA molecule and modifying uridines to pseudouridines. RluD-directed isomerization of uridines in 50S subunits *in vivo* by endogenous enzyme occurs at significantly lower RluD concentrations within 1 min (Ero *et al*., 2010). However, the experimental design of this study aimed at detecting the maximal activity of the synthases *in vitro* using the concentration of chimeric proteins two-fold higher than the concentration of substrate and incubation time 3 hours. As it was determined by HPLC analysis, RluCD synthesized about 27 pseudouridines in 23S rRNA and only 1 pseudouridine on the 50S subunit which comprise 4.4% and 0.16% of all uridines in LSU subunit respectively (Figure 16A and B). RluDC was less active in comparison with RluCD, modifying approximately 12 pseudouridines in 23S rRNA which is 1.9% of all uridines in LSU, and did not modify pseudouridines in the 50S subunit (Figure 16A and B). Noteworthy, previously it has been shown that RluD and RluC synthases "expand" their specificity *in vitro* and modify both free 16S rRNA and 23S rRNA at positions that are not their endogenous uridine substrates as well as their specific positions in 23S rRNA (Huang *et al*., 1998a; Wrzesinski *et al*., 1995; Vaidyanathan *et al*., 2007). For RluD the situation changes when 50S subunit is used as substrate. RluD exhibits fast and specific activity *in vitro* towards the helix 69 of 23S rRNA (Vaidyanathan *et al*., 2007; Ero *et al*., 2010). It can be explained by fact that RluD is a “late” enzyme, which is active on 50S subunit *in vivo*, but not on 23S rRNA (Ofengand and Del Campo, 2004; Siibak and Remme, 2010). Thus, the activity of RluD can be regulated to some extent by 23S rRNA structure and by presence of r-proteins. RluC, on the contrary, is active on 23S rRNA *in vivo* (Siibak and Remme, 2010). In spite of RluC being less active than RluD on 23S rRNA *in vitro*, it modifies extra uridines besides its

native positions, thus, its activity and specificity *in vivo* might be regulated by other factors inside the cell. In this work, activity of chimeric proteins RluCD and RluDC was dropped when 50S was used as substrate, which is pointing to the fact that r-proteins and particular 23S rRNA structure are restricting the activity of chimeric proteins.

Neither RluCD, nor RluDC were active on the poly-uridine oligonucleotide (Figure 16D), which is in agreement with the findings of Johnson and Söll about the activity of pseudouridine synthases on polyuridylic acid (Johnson and Söll, 1970). Therefore, it had been suggested that pseudouridine synthases require the 3D structure of the RNA molecule for their enzymatic activity. If we take into account the fact that RluD and RluC remained specific to the ribosomal RNA and did not modify uridines in tRNA (Huang *et al*., 1998a), chimeric proteins RluCD and RluDC also may be specific only to the ribosomal RNA and require a proper secondary structures of this molecule. Also, S4-like domain may require helical junctions for binding (Powers and Noller, 1995), which do not obviously exist in poly-uridine oligonucleotide.

RluCD and RluDC do not isomerize positions 1911 and 1917 in 23S rRNA *in vitro*.

In this work, the total sequencing was performed for 23S rRNA treated with RluCD protein. Uridines modified by RluCD were mapped after the treatment of 23S rRNA with CMCT/alkali and primer extension. As a result, 27 uridines isomerized to pseudouridines on 23S rRNA by RluCD were found on the sequencing gel. However, uridines at positions 1911 and 1917 which are specific for RluD and uridines at positions 955, 2504, and 2580 specific for RluC were not isomerized. For the position 1915 the results of the primer extension analysis were difficult to interpret, because a stop signal on the gel was detected both in the presence and in the absence of CMCT. Supposedly, this position is difficult to be read by the reverse transcriptase because of the methylation at this position by RlmH (Ero *et al*., 2008). For RluDC and RluD sequencing was performed only for the region comprising helix 69 and the adjacent area in domain IV. RluDC also did not produce any pseudouridines at positions 1911 and 1917. In contrast to these results, C -terminal His_6-tagged purified recombinant RluD protein could modify position 1917, but not 1911 *in vitro*. This is in agreement with the results obtained by Raychaudhuri *et al*. where N-terminal tag-containing RluD also could not isomerize uridine at position 1911 *in vitro* (Raychaudhuri *et al*,. 1998). Thus, according to the results of the present work using RluD, RluCD and RluDC, none of these enzymes could isomerize uridine at position 1911. For RluDC, not all positions of pseudouridines on 23S rRNA were determined and this should be done in the future. As was shown by Vaidyanathan *et al*., RluD with truncated S4-like domain exhibited low activity on 50S subunit and

synthetic 23S rRNA and did not modify uridines at positions 1911, 1915 and 1917, but unexpectedly modified uridine at position 2457 which is a natural substrate for RluE pseudouridine synthase (Vaidyanathan *et al*., 2007). Interestingly, RluE pseudouridine synthase has a S4-like domain at the N-terminus. In this study, 27 mapped positions in 23S rRNA revealed that RluCD did not isomerized any uridines to pseudouridines which are natural substrates to other pseudouridine synthases in *E. coli*. RluCD modified 1 uridine to pseudouridine on 50S subunit, but the exact modified position is not known and it needs to be studied if this position is a natural substrate for RluD or RluC proteins. Interestingly, *in vivo* experiments with RluCD which were carried out by Natalja Garber in the Prof. Remme group showed that RluCD is capable of synthesizing pseudouridines at positions 1917 and 2504, which are the substrates for RluD and RluC respectively (Garber, 2014). It is remarkable that pseudouridine at position 1917 is highly conserved and pseudouridine at position 2504 plays an important role in the susceptibility of bacteria to peptidyl transferase inhibitors. Finally, it is noteworthy that mapping the modified positions produced by RluCD on the secondary structure of 23S rRNA revealed that RluCD isomerized groups of 2-3 uridines more often than single uridines located far away from others (Figure 19). Modified positions were located in almost all domains of 23S rRNA, the only exception being domain III. Modified pseudouridines were more concentrated in the center of the secondary structure of 23S rRNA. The mapping positions modified by RluD, RluCD and RluDC on 23S rRNA in part of domain IV at different Mg^{2+} concentrations revealed that five positions (1939, 1940, 1944, 1946 and 1963) were isomerized by all three pseudouridine synthases at all three Mg^{2+} concentrations, and two positions (1955 and 1956) were made only by RluD and RluDC which share the S4-like domain (Figure 20). Further investigation of exact positions modified by RluD, RluCD and RluDC is necessary for activity and specificity comparison of three pseudouridine synthases.

Catalytic activity suppression of chimeric proteins RluCD and RluDC by magnesium cations during pseudouridine isomerization reaction *in vitro*

Activity and specificity of RluD *in vitro* has been earlier found to depend on the concentration of magnesium ions in solution: a significant increase in the activity of RluD at lower Mg^{2+} concentrations was observed together with the reduced specificity (Wrzesinski *et al*., 2000). In this study, the treatment of 23S rRNA with wild type (wt) RluD, RluCD and RluDC showed that at various Mg^{2+} concentrations these enzymes exhibit different activity. All three proteins were highly active on 23S rRNA at 1 mM Mg^{2+} concentration. The Mg^{2+} concentration increase from 1 mM to 10 mM caused 2-fold activity decrease for chimeric

proteins and 4.5-fold for wt RluD; Mg^{2+} concentration increase from 10 mM to 20 mM induced activity decrease 1.5-fold for chimeric synthases and 2-fold for wt RluD synthase. Thus, there seems to be a strong correlation of Mg^{2+} concentration and the activity of pseudouridine synthases tested. Interestingly, the concentration of free Mg^{2+} in the cytoplasm of *E. coli* is in the range of 1 mM to 5 mM and the total Mg^{2+} concentration in the cell is 15-25 mM (Alatossava *et al*., 1985). It can be suggested that the reason of such correlation is the fact that Mg^{2+} cations bind to ribosomal RNA and alter it to more structured form, so that in higher Mg^{2+} concentration conditions pseudouridine synthases cannot reach some uridines and modify them *in vitro*. But it is not yet known whether Mg^{2+} cations play the same role inside the cell. It is notable that if RluD S4-like domain is changed as it was done for RluCD chimeric protein, the activity of chimeric protein decreases at 1 mM Mg^{2+} concentration in comparison with wt RluD, but in the same time RluCD activity is higher than the activity of wt RluD at 10 mM and 20 mM Mg^{2+} concentrations. Thus, exchanging S4 domains with RluC protein helps RluCD chimeric synthase to bind and modify more uridines than with endogenous S4 domain at two abovementioned Mg^{2+} concentrations. As for RluDC, its activity at 10 mM and 20 mM Mg^{2+} concentrations are approximately the same as RluD activity at these concentrations. So, it can be suggested that at higher Mg^{2+} concentrations chimeric proteins RluCD and RluDC are more unspecific than RluD. Also, bands on gel had different intensities, suggesting that some uridines were modified by all synthases more frequently than others at different Mg^{2+} concentrations or at the same Mg^{2+} concentration. For example, positions 1940 and 1963 modified by RluD have more visible band at 1 mM Mg^{2+} concentration than at 10 mM and 20 mM or at positions 1201, 1203, 2130 and 2139 for RluCD at 1 mM Mg^{2+} concentration bands are more evident than at other positions. It should be mentioned that pseudouridine specific stop signals on gel can be detected when modified 23S rRNA constituted approximately 20% of the total rRNA population (Leppik *et al*., 2012). The detection limit of the CMCT/alkali-reverse transcriptase method is about 20% level of the pseudouridine formation and this method can be considered to be roughly quantitative (Leppik *et al*., 2012).

In summary, this study revealed that exchanging of S4 domain does not deprive chimeric proteins of their ability to modify uridines to pseudouridines, but these proteins do not recognize all the endogenous positions of RluD or RluC synthases *in vitro*. These facts highlight the importance of both native domains (S4-like domain and catalytic domain) working together to assure specificity of the RluD and RluC synthases. It is evident that S4-like domain plays some role in pseudouridine synthase specificity together with catalytic domain; hence, S4-like domain function is broader than simply binding to RNA. Apparently,

the substrate recognition mechanism in the case of RluC and RluD pseudouridine synthases is more complex than just being sequence dependent like it is for RluA, and some factors inside the cell help to facilitate the specificity of these synthases. Also, high Mg^{2+} concentration suppresses the activity of RluD, RluCD and RluDC *in vitro*, probably changing substrate accessibility for these synthases or maybe blocking the active center in the synthases. Chimeric proteins RluCD and RluDC lose their activity at low and at high Mg^{2+} concentration. In contrast, RluD loses its activity more rapidly but becomes more specific to position 1917 with increasing of Mg^{2+} concentration. Current work represents the first effort to explore the chimeric pseudouridine synthases RluCD and RluDC *in vitro* and the results of this work make the contribution to understanding the role of S4-like and catalytic domains in specificity of both RluD and RluC pseudouridine synthases as well as pose a question about extra factors inside the cell which may contribute to the specificity of these pseudouridine synthases.

In summary, getting insight into the substrate recognition and catalytic mechanism of RluC and RluD pseudouridine synthases as well as the role of their RNA binding domain is important for the basic understanding of the ribosome biogenesis and the control of gene expression at the translational level. In the light of recent findings about the presence of pseudouridines in mRNA and the association of pseudouridines with different disease in eukaryotes, it is very important to understand the basis of this reaction and variability in substrate recognition mechanisms.

Summary

This work focuses on bacterial pseudouridine synthases RluD and RluC – homologous enzymes which belong to the RluA family. RluD and RluC each convert three specific uridine bases in *E. coli* ribosomal 23S RNA to pseudouridines: bases 955, 2504, and 2580 (RluC), bases 1911, 1915, and 1917 (RluD). Both synthases have an N-terminal S4-like RNA binding domain and catalytic domain (Mizutani *et al*., 2004). Up to date, nothing is known about the mechanisms of specificity of RluC and RluD, that is: how these synthases choose their target uridines and distinguishes them from other uridines in rRNA. In order to shed light on the role of the S4-like and catalytic domains in the specificity of both RluD and RluC, chimeric proteins RluCD and RluDC were used in this study.

The main objective of the presented study was to describe the activity of the chimeric proteins RluCD and RluDC *in vitro* for various substrates and at different magnesium concentrations, and to map the uridines in 23S rRNA which were modified by the chimeric proteins.

The following summarizes briefly the essence of the presented work:

- Results demonstrate that exchanging of the S4-like domains between RluD and RluC producing chimeric proteins RluCD and RluDC do not inactivate these synthases, since they are capable of binding to 23S rRNA molecule and modify uridines to pseudouridines *in vitro*.

- It was shown that chimeric pseudouridine synthases RluCD and RluDC are highly active on protein-free 23S rRNA and not active on poly-uridine oligonucleotide *in vitro*. Whereas, RluCD is much less active and RluDC is not able to synthesize any pseudurudines on dissociated 50S subunit.

- With primer extension sequencing was established that either chimeric proteins RluCD and RluDC do not isomerize positions native for RluD (1911 and 1917) in 23S rRNA *in vitro*. Also, RluCD is not able to modify any of position specific for RluC (955, 2504 and 2580) *in vitro*.

- In this study, from 27 mapped positions in 23S rRNA modified by RluCD were not any uridines isomerized to pseudouridines which are natural substrates to other

pseudouridine synthases in *E. coli*.

- Mapping the modified positions produced by chimeric protein RluCD on the secondary structure of 23S rRNA revealed that modifications place by groups of 2-3 pseudouridines more often than single pseudouridines located far away from others and more concentrated in the center of the secondary structured 23S rRNA.

- RluCD and RluDC catalytic activity decreasing with increasement of magnesium cations concentration in the solution during pseudouridine isomerization reaction *in vitro*.

- Chimeric proteins RluCD and RluDC are more active than RluD at higher Mg^{2+} concentrations.

- The results show that S4-like domain plays important role in pseudouridine synthase specificity together with catalytic domain and highlight the importance of both native S4-like and catalytic domains work together to assure specificity of the RluD and RluC synthases.

E. coli pseudouridiini süntaaside RluC ja RluD S4-sarnase ning katalüütilise domeeni roll nende ensüümide aktiivsuses ning spetsiifilisuses _in vitro_

Jekaterina Aid

Resümee

Valgusüntees on üks olulisemaid ja ürgsemaid protsesse rakus. Selle käigus sünteesitakse mRNAs kodeeritud nukleotiidse järjestuse järgi vastavalt geneetilisele koodonile valkude aminohappeline järjestus. Kõik organismid, bakteritest eukarüootideni, kasutavad valgusünteesiks ribosoome, mis koosnevad valkudest ja ribosomaalsest RNA-st (rRNA). Bakterialne ribosoom koosneb väiksemast 30S ja suuremast 50S subühikust. rRNA moodustub suurema osa riboomist ja sisaldab mitmeid posttranskriptsiooniliselt modifitseeritud nukleotiide. On märgatud, et rRNA modifikatsioonid asuvad ribosoomis funktsionaalselt olulistes kohtades ja enamus neist positsioonidest on kõrgelt konserveerunud, viidates nende tähtsale rollile rakus.

Pseudouridiin (Ψ) on kõige rohkem levinud RNA posttranskriptsiooniline modifikatsioon mida on leitud enamikust struktuursetest RNA-dest: rRNA-st, tRNA-st ja snRNA-st. Pseudouridiinid sünteesitakse uridiinist isomerisatsiooni reaktsiooni käigus, kasutamata selleks lisaenergiat või lisafaktoreid. Isomerisatsiooni reaktsiooni viivad läbi pseudouridiini süntaasid. Bakteri pseudouridiini süntaasid on valgulised ensüümid, mis omavad nii katalüütilist, kui ka substraadi äratundmise ja seondumise aktiivsust. Algselt leiti ja identifitseeriti pseudouridiini süntaasid bakteris *Escherichia coli*. Pseudouridiini süntaasid klassifitseeritakse kuuesse perekonda: viis perekonda nimetatakse *E. coli* ensüümide järgi RluA, RsuA, TruA, TruB, ja TruD ning kuues perekond Pus10 on leitud ainult arheades ja eukarüootides.

E. coli ribosoomi suurema subühiku RNA (23S rRNA) sisaldab 10 pseudouridiini, mis sünteesitakse kuue pseudouridiini süntaasi abil. Pseudouridiini süntaasid RluC ja RluD sünteesivad kumbki kolme pseudouridiini. Mõlemad pseudouridiini süntaasid isomeriseerivad uridiini ribosoomi funktsionaalselt olulistes kohtades. RluD katalüüsib Ψ tekkimise 23S rRNA positsioonidesse 1911, 1915 ja 1917, mis asuvad kõrgelt konserveerunud rRNA heeliksis 69 (H69). Pseudouridiinid 1915 ja 1917 positsioonides on äärmiselt konserveerunud. RluC isomeriseerib positsioonides 955, 2504 ja 2580 asuvaid uridiine. Positsioonis 2504

asuva pseudouridiini eemaldamine põhjustab bakteri tundlikkust peptidüültransferaasi keskuse inhibiitorite vastu. Mõlemad pseudouridiini süntaasid koosnevad N-terminaalses osas S4-sarnasest domeenist ja C-terminaalses osas katalüütilisest domeenist. Vaatamata sellele, et RluC ja RluD pseudouridiini süntaase on juba pikalt uuritud, ei ole siiani suudetud välja selgitada kuidas RluC ja RluD oma substraadid ära tunnevad ja mis on nende pseudouridiinide täpne funktsioon. Samuti ei ole veel teada missugust rolli mängivad ülalmainitud domeenid pseudouridiini süntaasi spetsiifilisuses ja aktiivsuses.

Käesolevas töös uuriti kahe varasemalt konstrueeritud kimäärset pseudouridiini süntaasi RluCD ja RluDC, mille S4-sarnased domeenid olid vahetatud *in vitro* spetsiifikat. RluCD sisaldab RluD katalüütilist domeeni ja RluC S4-sarnast domeeni ning RluDC – vastupidi. RluCD ja RluDC kimäärsete valkude uurimine võimaldab selgitada S4-sarnase ja katalüütilise domeeni tähtsust pseudouridiini süntaaside spetsiifilisuse ja aktiivsuse jaoks. Selle töö peamiseks eesmärgiks oli RluCD ja RluDC kimäärsete valkude aktiivsuse kirjeldamine erinevatel substraatidel ja erinevatel magneesiumi kontsentratsioonidel *in vitro* ning uridiinide kaardistamine 23S rRNA molekulis, mis on *in vitro* modifitseeritud kimäärsete valkude poolt.

Tulemused võib kokku võtta järgmiselt:

- S4-sarnase domeeni vahetamisel RluD ja RluC valkude vahel ning rekombinantsete kimäärsete valkude RluCD ja RluDC rakkudest puhastamise käigus säilis ensüümi aktiivsus substraadi suhtes, kuna kimäärsed süntaasid olid võimelised 23S rRNA-le seonduma ja uridiine isomeriseerima.

- Kimäärsed pseudouridiini süntaasid RluCD ja RluDC näitavad vaba 23S rRNA suhtes üles kõrge aktiivsust, kuid polü-uridiini oligonukleotiidi suhtes *in vitro* aktiivsust ei tuvastatud. Võrrelduna vaba 23S rRNA-ga on RluCD dissotsieeritud 50S subühiku suhtes oluliselt vähem aktiivne ning RluDC-l puudub dissotsieeritud 50S subühiku suhtes aktiivsus üldse.

- Sekveneerimise analüüsi abil leiti, et RluCD ja RluDC süntaasid ei ole võimelised isomeriseerima uridiine *in vitro* positsioonides 1911 ja 1917, mis on RluD looduslikud substraadid. Samuti ei tunne RluCD *in vitro* substraatidena ära uridiine, mis asuvad positsioonides 955, 2504 ja 2580 ja on RluC looduslikud substraadid.

- Käesolevas töös avastati 27 positisooni, milles asuvad uridiinid modifitseeriti RluCD süntaasi poolt pseudouridiinideks. Seejuures ei modifitseeritud mitte ühtegi uridiini positsioonides, mis on RluD looduslikud substraadid. Samuti, on märkamist väärt et mitte ükski nendest positsioonidest ei ole *E. coli* pseudouridiini suntaaside looduslikud substraadid.

- RluCD poolt modifitseeritud positisioonide kaardistamine 23S rRNA sekundaarstruktuuris näitas, et RluCD modifitseerib uridiine eelistatult 2-3 nukleotiidiliste gruppidena, mis asuvad üksteise läheduses. Samuti selgus, et modifikatsioonid klasterdusid 23S rRNA sekundaarstruktuuri keskele.

- Magneesiumi kontsentratsiooni tõstmine põhjustab *in vitro* RluCD ja RluDC katalüütilise aktiivsuse langemise.

- Kõrgematel magneesiumi kontsentratsioonidel on kimäärsed valgud RluCD ja RluDC aktiivsemad kui pseudouridiini süntaas RluD.

Kokkuvõtteks võib väita, et S4-sarnane domeen mängib olulist rolli RluC ja RluD pseudouridiini süntaaside spetsiifilisuse määramises. S4-sarnase domeeni koordineeritud koostöö katalüütilise domeeniga on vajalik pseudouridiini süntaaside spetsiifilisuse tagamiseks.

Acknowledgements

I would like to express sincerest gratitude to:

Professor Jaanus Remme for his wise supervision, patience and for the great team of Molecular Biology lab that he has created.

Dr. Margus Leppik to whom I own the success in this work and the knowledge I have acquired in the field of ribosomology. His supervision, psychological support and readiness to answer all my questions at any time are immeasurable value to me.

Molecular Biology lab members, who became great companions during these two years.

I also want to thank my sister, Dr. Tamara Aid-Pavlidis (Institute of Molecular Biology and Biotechnology, FORTH, Irakleon), who always supported, believed in me and helped me during my whole life. Finally, I thank my daughter, Sofia Aid, for her understanding, unlimited patience and love.

References

Agrawal, R.K., Sharma, M.R., Kiel, M.C., Hirokawa, G., Booth, T.M., Spahn, C.M., Grassucci, R.A., Kaji, A. and Frank, J. (2004). Visualization of ribosome-recycling factor on the Escherichia coli 70S ribosome: functional implications. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 8900–8905.

Alatossava, T., Jütte, H., Kuhn, A. and Kellenberger, E. (1985). Manipulation of Intracellular Magnesium Content in Polymyxin-B Nonapeptide-Sensitized Escherichia-Coli by Ionophore A23187. *Journal of Bacteriology*, 162, 413–419.

Ali, I.K., Lancaster, L., Feinberg, J., Joseph, S., and Noller, H.F. (2006). Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Molecular Cell*, 23, 865–874.

Alian, A., DeGiovanni, A., Griner, S.L., Finer-Moore, J.S. and Stroud, R.M. (2009).Crystal structure of an RluF-RNA complex: a base-pair rearrangement is the key to selectivity of RluF for U2604 of the ribosome. *Journal of molecular biology*, 388, 785–800.

Andersen, N.M. and Douthwaite, S. (2006). YebU is a m5C Methyltransferase Specific for 16 S rRNA Nucleotide 1407. *Journal of Molecular Biology*, 359, 777–786.

Aravind, L. and Koonin, E. V. (1999). Novel predicted RNA-binding domains associated with the translation machinery. *Journal of molecular evolution*, 48, 291–302.

Arluison, V., Hountondji, C., Robert, B. and Grosjean, H. (1998). Transfer RNA-pseudouridine synthetase Pus1 of Saccharomyces cerevisiae contains one atom of zinc essential for its native conformation and tRNA recognition. *Biochemistry*, 37, 7268–7276.

Arps, P.J., Marvel, C.C., Rubin, B.C., Tolan, D.A., Penhoet, E.E. and Winkler, M.E. (1985). Structural features of the hisT operon of Escherichia coli K-12. *Nucleic Acids Research. DNA Sequence*, 5297–5315.

Bakin, A., Lane, B.G. and Ofengand, J. (1994). Clustering of pseudouridine residues around the peptidyltransferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry*, 33, 13475–13483.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* , 289, 905–920.

Bousquet-Antonelli, C., Henry, Y., Gelugne, J.-P., Caizergues-Ferrer, M. and Kiss, T. (1997). A small nucleolar RNP protein is required for pseudouridylation of eukaryotic ribosomal RNAs. *EMBO Journal*, 16, 4770–4776.

Brimacombe, R., Mitchell, P., Osswald, M., Stade, K. and Bochkariov, D. (1993). Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 7, 161–167.

Brosius, J., Palmer, M.L., Kennedy, P.J. and Noller, H.F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 75, 4801–4805.

Brosius, J., Dull, T.J. and Noller, H.F. (1980). Complete nucleotide sequence of a 23S ribosomal RNA gene from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 77, 201–204.

Bykhovskaya, Y., Casas, K., Mengesha, E., Inbal, A. and Fischel-Ghodsian, N. (2004). Missense mutation in pseudouridine synthase 1 (PUS1) causes mitochondrial myopathy and sideroblastic anemia (MLASA). *American journal of human genetics*, 74, 1303–1308.

Del Campo, M., Ofengand, J. and Malhotra, A. (2004). Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of Escherichia coli. *Rna*, 10, 231–239.

Del Campo, M., Kaya, Y. and Ofengand, J. (2001). Identification and site of action of the remaining four putative pseudouridine synthases in Escherichia coli. *RNA*, 7, 1603–1615.

Cantara, W. A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A.P., Fabris, D. and Agris, P. F. (2011). The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Research*, 39, 195–201.

Carter, A P., Clemons, W.M., Brodersen, D.E., Morgan-Warren, R.J., Wimberly, B.T. and Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, 407, 340–348.

Charette, M. and Gray, M.W. (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB life*, 49, 341–351.

Charpentier, B., Muller, S. and Branlant, C. (2005). Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Research*, 33, 3133–3144.

Cohn, W.E. (1960). Pseudouridine, a carbon-carbon linked ribonucleoside in ribonucleic acids: isolation, structure, and chemical characteristics. *The Journal of biological chemistry*, 235, 1488–1498.

Conrad, J., Niu, L., Rudd, K. Lane, B.G. and Ofengand, J. (1999). 16S ribosomal RNA pseudouridine synthase RsuA of Escherichia coli: deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. *RNA*, 5, 751–763.

Conrad, J., Sun, D., Englung, N. and Ofengand, J. (1998). The rluC gene of Escherichia coli codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. *Journal of Biological Chemistry*, 273, 18562–18566.

Corollo, D., Blair-Johnson, M., Conrad, J., Fiedler, T., Sun, D., Wang, L., Ofengand, J. and Fenna, R. (1999). Crystallization and characterization of a fragment of pseudouridine synthase RluC from Escherichia coli. *Acta crystallographica. Section D, Biological crystallography*, 55, 302–304.

Czerwoniec, A., Dunin-Horkawicz, S., Purta, E., Kaminska, K.H., Kasprzak, J.M., Bujnicki, J.M., Grosjean, H. and Rother, K. (2009). MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Research*, 37, 118–121.

Davis, D.R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic acids research*, 23, 5020–5026.

Davis, F.F. and Allen, F.W. (1957). Ribonucleic acids from yeast which contain a fifth nucleotide. *Journal of Biological Chemistry*, 227, .907–915.

Decatur, W. A. and Fournier, M.J. (2002). rRNA modifications and ribosome function. *Trends in Biochemical Sciences*, 27, 344–351.

Desaulniers, J.P., Chang, Y.C., Aduri, R. Abeysirigunawardena, S.C., SantaLucia, J., Jr. and Chow, C.S. (2008). Pseudouridines in rRNA helix 69 play a role in loop stacking interactions. *Organic & biomolecular chemistry*, 6, 3892–3895.

Diaconu, M.,Kothe, U., Schlunzen, F., Fischer, N., Harms, J.M., Tonevitsky, A.G., Stark, H., Rodnina, M.V. and Wahl, M.C. (2005). Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTpase activation. *Cell*, 121, 991–1004.

Dabbs, E.R., Ehrlich, R.,Hasenbank, R.H., Schroeter, B.-H., Stoffler-Melnicke, M. and Stoffler, G. (1981). Mutants of Escherichia coli Lacking Protein Ll. *Journal of Molecular Biology*, 149, 553–578.

Ejby, M., Sørensen, M. A. and Pedersen, S. (2007). Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19410–19415.

Ero, R., Peil, L., Liiv, A. and Remme, J. (2008). Identification of pseudouridine methyltransferase in Escherichia coli. *Rna*, 14, 2223–2233.

Ero, R., Leppik, M., Liiv, A. and Remme, J. (2010). Specificity and kinetics of 23S rRNA modification enzymes RlmH and RluD. *Rna*, 16, 2075–2084.

Friedt, J., Leavens, F.M.V., Mercier, E., Wieden, H.J. and Kothe, U. (2014). An arginine-aspartate network in the active site of bacterial TruB is critical for catalyzing pseudouridine formation. *Nucleic Acids Research*, 42, 3857–3870.

Garber, N. (2014). *Kimäärse pseudouridiini süntaasi spetsiifika iseloomustamine*. Tartu Ülikool.

Ge, J. and Yu, Y.T. (2013). RNA pseudouridylation: new insights into an old modification. *Trends in Biochemical Sciences*, 38, 210–218.

Gehrke, C.W. and Kuo, K.C. (1989). Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. *Journal of chromatography*, 471, 3–36.

Glotz, C., Zwieb, C. and Brimacombe, R. (1981). Secondary structure of the large subunit ribosomal RNA from Escherichia coli, Zea mays chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic acids research*, 9, 3287–3306.

Glotz, C. and Brimacombe, R. (1980). An experimentally-derived model for the secondary structure of the 16S ribosomal RNA from Escherichia coli. *Nucleic acids research*, 8, 2377–2395.

Green, R. and Noller, H.F. (1997). Ribosomes and translation. *Annual review of biochemistry*, 66, 679–716.

Gustafsson, C., Reid, R., Greene, P.J. and Santi, D.V. (1996). Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic Acids Research*, 24, 3756–3762.

Gutgsell, N., Englund, N., Niu, L., Kaya, Y., Lane, B.G. and Ofengand, J. (2000). Deletion of the Escherichia coli pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells. *Rna*, 6, 1870–1881.

Gutgsell, N.S., Del Campo, M., Raychaudhuri, S. and Ofengand, J. (2001). A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911, 1915, and 1917 in Escherichia coli 23S ribosomal RNA restores normal growth to an RluD-minus strain. *Rna*, 7, 990–998.

Gutgsell, N.S., Deutscher, M.P. and Ofengand, J., (2005). The pseudouridine synthase RluD is required for normal ribosome assembly and function in Escherichia coli. *Rna*, 11, 1141–1152.

Hamma, T., Reichow, S.L., Varani, G. and Ferré-D'Amaré, A.R. (2005). The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nature structural & molecular biology*, 12, 1101–1107.

Hamma, T. and Ferré-D'Amaré, A.R. (2006). Pseudouridine Synthases. *Chemistry and Biology*, 13, 1125–1135.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell*, 107, 679–688.

Hirabayashi, N., Sato, N.S. and Suzuki, T. (2006). Conserved loop sequence of helix 69 in Escherichia coli 23 S rRNA is involved in A-site tRNA binding and translational fidelity. *Journal of Biological Chemistry*, 281, 17203–17211.

Hoang, C. and Ferre-D'Amare, A.R. (2004). Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *Rna*, 10, 1026–1033.

Hoang, C. and Ferré-D'Amaré, A.R. (2001). Cocrystal structure of a tRNA Ψ55 pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell*, 107, 929–939.

Huang, L., Pookanjanatavip, M., Gu, X. and Santi, D.V. (1998b). A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry*, 37, 344–351.

Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P.J. and Santi, D.V. (1998a). Identification of two Escherichia coli pseudouridine synthases that show multisite specificity for 23S RNA. *Biochemistry*, 37, 15951–15957.

Johnson, L. and Söll, D. (1970). In vitro biosynthesis of pseudouridine at the polynucleotide level by an enzyme extract from Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 67, 943–950.

Kaczanowska, M. and Rydén-Aulin, M. (2007). Ribosome biogenesis and the translation process in Escherichia coli. *Microbiology and molecular biology reviews : MMBR*, 71, 477–494.

Kammen, H.O., Marvel, C.C., Hardy, L. and Penhoet, E.E. (1988). Purification, structure, and properties of Escherichia coli tRNA pseudouridine synthase I. *Journal of Biological Chemistry*, 263, 2255–2263.

Karikó, K., Muramatsu, H., Welsh, F.A., Ludwig, J., Kato, H., Akira, S. and Weissman, D. (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic

vector with increased translational capacity and biological stability. *Molecular Therapy*, 1833–1840.

Karikó, K., Muramatsu, H., Keller, J.M. and Weissman, D (2012). Increased Erythropoiesis in Mice Injected With Submicrogram Quantities of Pseudouridine-containing mRNA Encoding Erythropoietin. *Molecular Therapy*, 20, 948–953.

Kaya, Y., Del Campo, M., Ofengand, J. and Malhotra, A. (2004). Crystal Structure of TruD, a Novel Pseudouridine Synthase with a New Protein Fold. *Journal of Biological Chemistry*, 279, 18107–18110.

Kaya, Y. and Ofengand, J. (2003). A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *Rna*, 9, 711–721.

Khaitovich, P., Tenson, T., Kloss, P and Mankin, A.S. (1999). Reconstitution of functionally active Thermus aquaticus large ribosomal subunits with in vitro transcribed rRNA. *Biochemistry*, 38, 1780–1788.

Kiss-László, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M. and Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: A novel function for small nucleolar RNAs. *Cell*, 85, 1077–1088.

Klein, D.J., Moore, P.B. and Steitz, T. A, (2004b). The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna*, (McCarthy 1962), 1366–1379.

Klein, D.J., Moore, P.B. and Steitz, T. A., (2004a). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *Journal of Molecular Biology*, 340, 141–177.

Koonin, E. V. (1996). Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic acids research*, 24, 2411–2415.

Korostelev, A., Trakhanov, S., Laurberg, M. and Noller, H.F. (2006). Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell*, 126, 1065–1077.

Krishnan, K. and Flower, A.M. (2008). Suppression of ΔbipA phenotypes in Escherichia coli by abolishment of pseudouridylation at specific sites on the 23S rRNA. *Journal of Bacteriology*, 190, 7675–7683.

Kurland, C.G. (1960). Molecular characterization of ribonucleic acid from Escherichia coli ribosomes. *Journal of Molecular Biology*, 2, 83–91.

Lake, J. A. (1985). Evolving ribosome structure: domains in archaebacteria, eubacteria, eocytes and eucaryotes. *Annual review of biochemistry*, 54, 507–530.

Lake, J. A. and Strycharz, W. A. (1981). Ribosomal proteins L1, L17 and L27 from Escherichia coli localized at single sites on the large subunit by immune electron microscopy. *Journal of molecular biology*, 153, 979–992.

Leppik, M., Ero, E., Liiv, A., Kipper, K. and Remme, J. (2012). Different sensitivity of H69 modification enzymes RluD and RlmH to mutations in Escherichia coli 23S rRNA. *Biochimie*, 94, 1080–1089.

Leppik, M. (2007a). *Pseudouridiini süntaasi RluD substraadi spetsiifilisus in vivo ja in vitro*. Tartu Ülikool.

Leppik, M., Peil, L., Kipper, K., Liiv, A. and Remme, J. (2007b). Substrate specificity of the pseudouridine synthase RluD in Escherichia coli. *FEBS Journal*, 274, 5759–5766.

Liang, X.-H., Liu, Q. and Fournier, M.J., 2009. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *Rna*, 15, 1716–1728.

Liiv, A., Karitkina, D., Maiväli, Ü. and Remme, J. (2005). Analysis of the function of E. coli 23S rRNA helix-loop 69 by mutagenesis. *BMC molecular biology*, 6, 18.

Lovejoy, A.F., Riordan, D.P. and Brown, P.O. (2014). Transcriptome-Wide Mapping of Pseudouridines: Pseudouridine Synthases Modify Specific mRNAs in S. cerevisiae. *PLoS ONE*, 9, e110799.

Lövgren, J.M. and Wikström, P.M. (2001). The rlmB gene is essential for formation of Gm2251 in 23S rRNA but not for ribosome maturation in Escherichia coli. *Journal of Bacteriology*, 183, 6957–6960.

Manival, X., Charron, C., Fourmann, J.-B., Godard, F., Charpentier, B. and Branlant, C. (2006). Crystal structure determination and site-directed mutagenesis of the Pyrococcus abyssi aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Research*, 34, 826–839.

McCleverty, C.J., Hornsby, M., Spraggon, G. and Freusch, A. (2007). Crystal Structure of Human Pus10, A Novel Pseudouridine Synthase. *Journal of Molecular Biology*, 373, 1243–1254.

McMahon, M., Bellodi, C. and Ruggero, D. (2013). The "fifth" RNA nucleotide: a role for ribosomal RNA pseudouridylation in control of gene expression at the translational level. In *Biophysics for the life sciences*.

Miracco, E.J. and Mueller, E.G. (2011). The products of 5-fluorouridine by the action of the pseudouridine synthase TruB disfavor one mechanism and suggest another. *Journal of the American Chemical Society*, 133, 11826–11829.

Mizutani, K., Machida, Y., Unzai, S., Park, S.Y. and Tame, J.R. (2004). Crystal Structures of the Catalytic Domains of Pseudouridine Synthases RluC and RluD from Escherichia coli. *Biochemistry*, 43, 4454–4463.

Nissen, P.,Hansen, J., Ban, N., Moore, P.B. and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289, 920–930.

Noeske, J.,Wasserman, M.R., Terry, D.S., Altman, R.B., Blanchard, S.C. and Cate, J.H.D. (2015). High-resolution structure of the Escherichia coli ribosome. *Nature Structural & Molecular Biology*, 22, 336–341.

Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R. and Waese, C.R. (1981). Secondary structure model for 23S ribosomal RNA. *Nucleic acids research*, 9, 6167–6189.

Noller, H.F. and Woese, C.R. (1981). Secondary structure of 16S ribosomal RNA. *Science*, 212, 403–411.

Nurse, K., Wrzesinski, J, Bakin, A., Lane, B.G. and Ofengand, J. (1995). Purification, cloning, and properties of the tRNA Ψ55 synthase from Escherichia coli. *Rna*,1, 102–112.

O'Connor, M. and Gregory, S.T. (2011). Inactivation of the RluD pseudouridine synthase has minimal effects on growth and ribosome function in wild-type Escherichia coli and Salmonella enterica. *Journal of Bacteriology*, 193, 154–162.

Ofengand, J., Malhotra, A., Remme, J., Gutgsell, N.S., Del Campo, M., Jean-Charles, S., Peil, L. and Kaya, Y. (2001a). Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harbor Symposia on Quantitative Biology*, 66, 147–159.

Ofengand, J. (2002). Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Letters*, 514, 17–25.

Ofengand, J. and Bakin, A. (1997). Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *Journal of molecular biology*, 266, 246–268.

Ofengand, J. and Del Campo, M. (2004). Modified Nucleosides of Escherichia coli Ribosomal RNA. *EcoSal Plus*.

Ofengand, J., Del Campo, M. and Kaya, Y. (2001b). Mapping pseudouridines in RNA molecules. *Methods*, 25, 365–373.

Ogle, J.M., Murphy, F.V., Tarry, M.J. and Ramakrishnan, V (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, 111, 721–732.

Ogle, J.M., Carter, A.P. and Ramakrishnan, V. (2003). Insights into the decoding mechanism from recent ribosome structures. *Trends in Biochemical Sciences*, 28, 259–266.

Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J. and Stroud, R.M. (2003). Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 12648–12653.

Petrov, A.S.,Bernier, C.R., Hershkovits, E., Hue, Y., Waterbury, C.C., Hsiao, C., Stepanov, V.G., Gaucher, E.A., Grover, M.A., Harvey, S.C., Hud, N.V., Wartell, R.M., Fox, G.E. and Williams, L.D. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Research*, 41, 7522–7535.

Pettersson, I., Hardy, S.J. and Liljas, A (1976). The ribosomal protein L8 is a complex L7/L12 and L10. *FEBS letters*, 64, 135–138.

Phannachet, K., Elias, Y. and Huang, R.H. (2005). Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry*, 44, 15488–15494.

Phannachet, K. and Huang, R.H. (2004). Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. *Nucleic Acids Research*, 32, 1422–1429.

Piekna-Przybylska, D., Decatur, W. A. and Fournier, M.J. (2008). The 3D rRNA modification maps database: With interactive tools for ribosome analysis. *Nucleic Acids Research*, 36, 178–183.

Powers, T. and Noller, H.F. (1995). Hydroxyl radical footprinting of ribosomal proteins on 16S rRNA. *Rna*, 1, 194–209.

Purta, E., O'Connor, M., Bujnicki, J.M. and Duothwaite, S. (2009). YgdE is the 2'-O-ribose methyltransferase RlmM specific for nucleotide C2498 in bacterial 23S rRNA. *Molecular Microbiology*, 72, 1147–1158.

Ramakrishnan, V. (1986). Distribution of protein and RNA in the 30S ribosomal subunit. *Science*, 231, 1562–1564.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell*, 108, 557–572.

Ramamurthy, V., Swann, S.L., Paulson, J.L., Spedaliere, C.J. and Mueller, E.G. (1999). Critical aspartic acid residues in pseudouridine synthases. *Journal of Biological Chemistry*, 274, 22225–22230.

Rander, A. (2009). *111 aluspaarise deletsiooni mõju 23 S rRNA modifitseerimisele Escherichia coli rakkudes.* Tartu Ülikool.

Raychaudhuri, S., Conrad, J., Hall, B.G. and Ofengand, J. (1998). A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of Escherichia coli. *Rna*, 4, 1407–1417.

Raychaudhuri, S., Conrad, J. and Ofengand, J. (1999). Functional Effect of Deletion and Mutation of the. *Biochemistry*, 274, 18880 –18886.

Schaub, R.E. and Hayes, C.S. (2011). Deletion of the RluD pseudouridine synthase promotes SsrA peptide tagging of ribosomal protein S7. *Molecular Microbiology*, 79, 331–341.

Scheper, G.C., van der Knaap, M.S. and Proud, C.G. (2007). Translation matters: protein synthesis defects in inherited disease. *Nature reviews. Genetics*, 8, 711–723.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. and Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102, 615–623.

Schlünzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A. and Franceschi, F. (2001). Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, 413, 814–821.

Schmeing, T.M., Huang, K.S., Strobel, S.A. and Steitz, T.A. (2005). An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature*, 438, 520–524.

Schuwirth, B.S., Borovinskaya, M.A., Hau, C.W., Zhang, W., Vila-Sanjurjo, A., Holton, J.M. and Cate, J.H.D. (2005). Structures of the bacterial ribosome at 3.5 A resolution. *Science*, 310, 827–834.

Schwartz, S., Bemstein, D.A., Mumbach, M.R., Jovanovic, M., Herbst, R.H., Leon-Ricardo, B.X., Engreitz, J.M., Guttman, M., Satija, R., Lander, E.S., Fink, G. and Regev, A. (2014). Transcriptome-wide Mapping Reveals Widespread Dynamic-Regulated Pseudouridylation of ncRNA and mRNA. *Cell*, 159, 148-162.

Selmer, M., Dunham, C.M., Murphy, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R. and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, 313, 1935–1942.

Sergiev, P. V., Golovina, A.Y., Sergeeve, O.V., Osterman, I.A., Nesterchuk, M.V., Bogdanov, A.A. and Dontsova, O.A. (2012). How much can we learn about the function of bacterial rRNA modification by mining large-scale experimental datasets? *Nucleic Acids Research*, 40, 5694–5705.

Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, 71, 1342–1346.

Siibak, T. and Remme, J. (2010). Subribosomal particle analysis reveals the stages of bacterial ribosome assembly at which rRNA nucleotides are modified. *Rna*, 16, 2023–2032.

Sivaraman, J., Iannuzzi, P., Cygler, M. and Matte, A. (2004). Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23 S rRNA and is essential for normal cell growth of Escherichia coli. *Journal of Molecular Biology*, 335, 87–101.

Staker, B.L., Korber, P., Bardwell, J.C.A. and Saper, M.A. (2000). Structure of Hsp15 reveals a novel RNA-binding motif. *The EMBO journal*, 19, 749–757.

Stelzl, U., Connell, S., Nierhaus, K.H. and Wittmann-Liebold, B. (2001). Ribosomal proteins: role in ribosomal functions. *eLS*, 000, 1–12.

Strycharz, W. A, Nomura, M. and Lake, J. A. (1978). Ribosomal proteins L7/L12 localized at a single region of the large subunit by immune electron microscopy. *Journal of molecular biology*, 126, 123–140.

La Teana, A, Gualerzi, C.O. and Dahlberg, A. E. (2001). Initiation factor IF 2 binds to the alpha-sarcin loop and helix 89 of Escherichia coli 23S ribosomal RNA. *Rna*, 7, 1173–1179.

Tissieres, A. and Watson, J. (1958). Ribonucleoprotein particles from Escherichia coli. *Nature*, 182, 778–780.

Tissières, A., Watson, J.D., Sohlessinger, D. and Hollingwoth, B.R. (1959). Ribonucleoprotein particles from Escherichia coli. *Journal of Molecular Biology*, 1, 221–233.

Toh, S.M. and Mankin, A.S. (2008). An Indigenous Posttranscriptional Modification in the Ribosomal Peptidyl Transferase Center Confers Resistance to an Array of Protein Synthesis Inhibitors. *Journal of Molecular Biology*, 380, 593–597.

Tycowski, K.T., Smith, C.M., Shu, M.-D. and Steitz, J.A. (1996). A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 14480–14485.

Vaidyanathan, P.P., Deutscher, M.P. and Malhotra, A. (2007). RluD, a highly conserved pseudouridine synthase, modifies 50S subunits more specifically and efficiently than free 23S rRNA. *Rna*, 13, 1868–1876.

Voorhees, R.M., Weixlbaumer, A., Loakes, D., Kelley, A.C. and Ramakrishnan, V. (2009). Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nature structural & molecular biology*, 16, 528–533.

Waller, J.P. and Harris, J.I. (1961). Srudies on the composition of the protein from Escherichia coli ribosomes. *Biochemistry*, 47, 18–23.

Watkins, N.J.,Gottschalk, A., Neubauer, G., Kastner, B., Fabrizio, P., Mann, M. and Luhrmann, R. (1998). Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *Rna*, 4, 1549–1568.

Wilson, D.N. and Nierhaus, K.H. (2005). Ribosomal proteins in the spotlight. *Critical reviews in biochemistry and molecular biology*, 40, 243–267.

Wimberly, B.T., Brodersen, D.E., Clemons Jr, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407, 327–339.

Wriggers, W., Agrawal, R.K., Drew, D.L., McCammon, A. and Frank, J. (2000). Domain motions of EF-G bound to the 70S ribosome: insights from a hand-shaking between multi-resolution structures. *Biophysical journal*, 79, 1670–1678.

Wrzesinski, J., Nurse, K., Bakin, A., Lane, B.G. and Ofengand, J. (1995). A dual-specificity pseudouridine synthase: an Escherichia coli synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). *Rna*, 1, 437–448.

Wrzesinski, J., Bakin, A., Ofengand, J. and Lane, B.G. (2000). Isolation and properties of Escherichia coli 23S-RNA pseudouridine 1911, 1915, 1917 synthase (RluD). *IUBMB life*, 50, 33–37.

Yusupov, M.M., Yusupova, G. Zh., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D. and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 A resolution. *Science*, 292, 883–896.

Yusupov, M.M. and Spirin, A S. (1986). Are there proteins between the ribosomal subunits? Hot tritium bombardment experiments. *FEBS letters*, 197, 229–233.

Zhao, Q., Öfverstedt, L.-G., Skoglund, U. and Isaksson, L.A.(2004). Morphological variation of individual Escherichia coli 50S ribosomal subunits in situ, as revealed by cryo-electron tomography. *Experimental Cell Research*, 300, 190–201.

Web pages list

http://apollo.chemistry.gatech.edu/RibosomeGallery/

http://www.rna.icmb.utexas.edu/

Supplements

Supplement 1

Pseudouridine positions modified by RluCD in different 23S rRNA domains

Pseudouridine position	Domains of 23S rRNA
67	Domain I
1199	Domain II
1201	Domain II
1203	Domain II
1219	Domain II
1222	Domain II
1258	Domain 0
1263	Domain 0
1267	Domain 0
1940	Domain IV
1944	Domain IV
1963	Domain IV
2130	Domain V
2139	Domain V
2500	Domain V
2514	Domain V
2519	Domain V
2552	Domain V
2561	Domain V
2613	Domain 0
2617	Domain 0
2629	Domain VI
2637	Domain VI
2647	Domain VI
2650	Domain VI
2656	Domain VI
2898	Domain I

Non-exclusive licence to reproduce thesis and make thesis public

I, Jekaterina Aid
(date of birth: 28.04.1984),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2.make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

The role of S4-like RNA-binding and catalytic domains in the activity and specificity of bacterial pseudouridine synthases RluC and RluD *in vitro*,

supervised by Prof. Jaanus Remme, PhD and Margus Leppik, PhD

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 26.05.2015