Substrate binding and catalysis by the pseudouridine synthases RluA and TruB

Laura Carole Keffer-Wilkes

B.Sc. University of Guelph, 2008

A Thesis

Submitted to the School of Graduate Studies

of the University of Lethbridge

in Partial Fulfilment of the

Requirements for the Degree

MASTER OF SCIENCE

Department of Chemistry and Biochemistry

University of Lethbridge

LETHBRIDGE, ALBERTA, CANADA

Abstract

Pseudouridine is the most common RNA modification found in all forms of life. The exact role pseudouridines play in the cell is still relatively unknown. However, its extensive incorporation in functionally important areas of the ribosome and the fitness advantage provided to cells by pseudouridines implies that its presence is important for the cell. The enzymes responsible for this modification, pseudouridine synthases, vary greatly in substrate recognition mechanisms, but all enzymes supposedly share a universally conserved catalytic mechanism. Here, I analyze the kinetic mechanisms of pseudouridylation utilized by the exemplary pseudouridine synthase RluA in order to compare it with the previously determined rate of pseudouridylation of the pseudouridine synthase TruB. My results demonstrate that RluA has the same uniformly slow catalytic step as previously determined for TruB and TruA. This constitutes the first step towards identifying the catalytic mechanism of the pseudouridine synthase family. Additionally, it was my aim to identify the major determinants for RNA binding by pseudouridine synthases. By measuring the dissociation constants (K_D) for substrate and product tRNA by nitrocellulose filtration assays, I showed that both tRNA species could bind with similar affinities. These binding studies also revealed that TruB's interaction with the isolated T-arm is the major contact site contributing to the affinity of the enzyme to RNA. Finally, a new contact between tRNA and TruB's PUA domain was identified which was not observed in the crystal structure. In summary, my results provide new insight into the common catalytic step of pseudouridine synthases and the specific interactions contributing to substrate binding by the enzyme TruB. These results will enable future studies on the kinetic mechanism of pseudouridine synthases, in particular the kinetics of substrate and product binding and release, as well as on the chemical mechanism of pseudouridine formation.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Ute Wieden-Kothe, for the unending support and mentorship she has provided me throughout my degree. The opportunity to teach and learn as part of the Kothe lab has been a life-changing experience. Thank you for pushing me to be better at everything I do.

Thank you to Dr. Tony Russell and Dr. Stacey Wetmore, my committee members. The helpful advice and encouragement greatly contributed to my success.

Thank you to all the members of the Kothe and Wieden labs. I was welcomed with open arms to this tight-knit group and they have not only been amazing coworkers but have become life-long friends.

Finally, I would like to thank my friends and family. Being far away from home is very hard at times but knowing I have your support and love helped me to keep going. Thank you so much for everything you've done.

Table of Contents

List of Tables

List of Figures

Abbreviations

Ψ	pseudouridine
1,5-IAEDANS	5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid
2AP	deoxyribo-2-aminopurine
A	Alanine
ASL	anticodon stem loop (of tRNA)
C	Cysteine
CD	Circular dichroism
CIP	calf intestinal alkaline phosphatise
CMCT	*N*-cyclohexyl-*N'-β*-(4-methylmorpholinium)ethylcarbodiimide *p*-tosylate
D	aspartate
DEAE	diethylaminoethyl
DNA	deoxyribonucleic acid
Dpm	decays per minute
DTT	dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FL	fluorescein-5-thiosemicarbazide
FRET	Fluorescence resonance energy transfer
GFP	Green Fluorescent Protein
Gm18	2'-O-methylguanosine (position 18 in the D-loop of tRNA)
iPPase	inositol polyphosphate-1-phosphatase
IPTG	β-D-1-thiogalactopyranoside
LB	lysogeny broth
mRNA	messenger ribonucleic acid
MW	molecular weight
MWCO	molecular weight cutoff
MWM	molecular weight marker
NMR	Nuclear magnetic resonance

PAGE	polyacrylamide gel electrophoresis
PDB	Protein Data Bank
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
PTC	peptidyl transferase centre
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal ribonucleic acid
SDS	sodium dodecyl sulfate
snRNA	small nuclear ribonucleic acid
snRNP	small nuclear ribonucleoprotein
snoRNA	small nucleolar ribonucleic acid
tRNA	transfer ribonucleic acid
TSL	T-arm stem loop (of tRNA)
T	Threonine
T_m	melting temperature
U	Uracil
UV	Ultraviolet
wt	wild-type

Chapter 1 – Introduction

1.1 RNA in the Cell

Ribonucleic acids (RNA) play many essential roles within the cell. During gene expression, messenger RNA (mRNA) is transcribed from the DNA genome and then translated by the ribosome into a polypeptide sequence. Notably, ribosomal RNA (rRNA) is the catalytic component of the ribosome (Noller et al., 1992). Additionally, transfer RNAs (tRNA) transport amino acid residues to the ribosomal A site during protein synthesis. Gene regulation can be achieved through micro RNA (miRNA) or small interfering RNA (siRNA) as part of the RNA-induced silencing complex (RISC) in the RNA interference pathway. Small nuclear RNAs (snRNA) are highly conserved 60- to 300-nucleotide RNAs that are part of small nuclear ribonucleoproteins (snRNPs; see below) and are involved in splicing. Finally, small nucleolar RNAs (snoRNA) contribute to RNA biogenesis and are used as guide RNA templates in RNA modification. While some RNAs such as miRNA and siRNA require specific short ribonucleotide sequences for their roles in the cell, the common theme among many functional RNA species (e.g. tRNA and rRNA) is the requirement of a three-dimensional structure in order to complete their functions. This can be achieved through specific primary sequences, as well as through enzyme-catalyzed modifications to existing ribonucleotides or by interaction with RNA chaperones or RNA-binding proteins.

1.2 Pseudouridine

All forms of RNA are comprised of the same four ribonucleotide components. A nucleobase is attached to a ribose sugar at the C1’ position, and phosphate groups are found at the 3’ and 5’ carbons of the sugar connecting the individual nucleotides by phosphodiester bonds. The bases in RNA differ only slightly from those found in DNA.

RNA contains the typical bases adenine, guanine, and cytosine also found in DNA. Additionally, RNA is comprised of the unmethylated form of thymine, uracil (U), which base pairs with adenine. In addition to these four nucleotides, modifications to the base and sugar moieties are commonly found in noncoding, functional RNAs, such as tRNA and rRNA. Nearly 100 naturally occurring RNA modifications have been described (Lane 1998). Coding RNA (mRNA) can also be posttranscriptionally modified by the addition of the 5' cap and 3' poly-A tail. tRNA and rRNAs can have methyl groups attached to the endocyclic carbon (e.g. 2-methyladenosine), the endocyclic nitrogen (e.g. 1-methyladenosine), and the exocyclic nitrogen (N^6-methyladenosine)of the bases. Additionally, the exocyclic O^2-ogygen of the ribose (e.g. 2'-O-methylribose) can be methylated. The most common RNA modifications are pseudouridines (Lane, 1998).

Pseudouridine was first identified in the 1950s and has commonly been called the fifth nucleotide (Davis and Allen, 1957). Pseudouridine is the C-glycosyl isomer of the N-glycosyl nucleotide, uridine, and contains the only C-C glycosidic bond in RNA or DNA (Lane, 1998). The enhanced rotational freedom of the C-C bond over the N-C bond allows for greater conformational flexibility of the nucleoside (Figure 1; Davis, 1998). In pseudouridine, the N_1 position is available for additional hydrogen bonding, potentially forming novel base pairing interactions (Charette and Gray, 2000). Arnez and Steitz (1994) observed in the crystal structure of tRNA(Gln) in complex with glutaminyl-tRNA synthetase and ATP, that the N_1H imino group could coordinate a water molecule between the nucleobase and the phosphate backbone, resulting in an overall more rigid local tRNA structure.

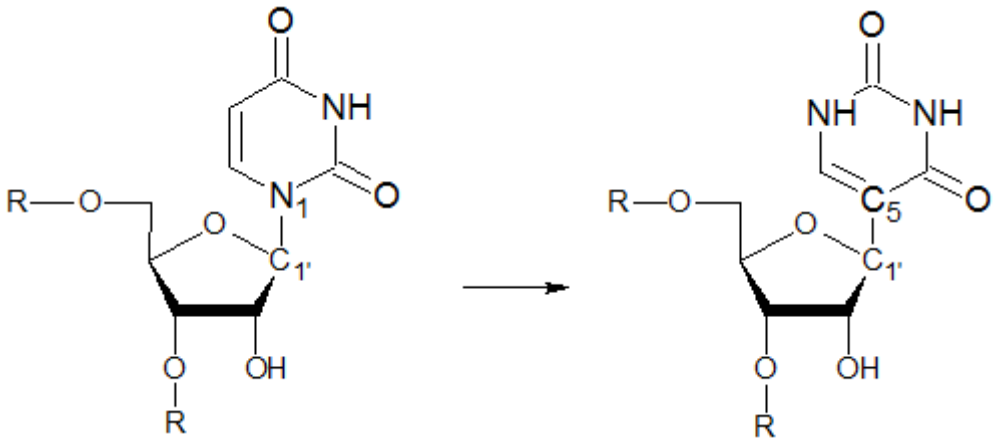


Figure 1.1: Conversion of uridine to pseudouridine. The $C_{1'}$ – N_1 glycosidic bond of uridine (left) is broken, the base rotated, and a new $C_{1'}$ – C_5 glycosidic bond forms in pseudouridine (right).

In solution, free pseudouridine tends to have a slight preference for the *syn* glycosyl conformation. This is in contrast to the *anti* conformation usually found in uridine and other nucleosides (Davis, 1998). When pseudouridine is incorporated into a polynucleotide chain however, only the *anti* configuration has been observed (Yarian et al., 1999). It is in this conformation that pseudouridine adopts the appropriate geometry to coordinate the water molecule between the N_1 position and its 5' phosphate connecting it to its neighbouring residue (Arnez and Steitz 1994). NMR, UV, and CD spectroscopy revealed that pseudouridine, as part of an oligoribonucleotide sequence, forms an A-form helical conformation with increased base stacking compared to the unmodified RNA (Davis, 1995). This improved base stacking has been proposed as the most important effect of pseudouridine on RNA structure stabilization (Davis, 1995).

1.3 Pseudouridine in RNA

Pseudouridines can be found in several functional RNAs. In eukaryotes, pseudouridines are an integral part of the spliceosomal machinery. The pre-mRNA initially transcribed from an organism's genome contains both protein-coding exons and non-coding introns.

During splicing, the introns are removed and the exons are spliced together to form the mature mRNA strand. The spliceosome is a large, dynamic RNA-protein complex. As part of spliceosome assembly, uridyl-rich snRNAs form RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs). Pseudouridines within these snRNAs are generally clustered in functionally important regions and are required for snRNP assembly and splicing. During spliceosome assembly, uridyl-rich snRNAs – U1, U2, U4, U5, and U6 associate with several protein components. Base-pairing between U1 and the 5' splice site of the pre-mRNA helps position the splicing machinery. The pseudouridine-containing U2 snRNP can then recognize the branch site through base-pairing and in that way helps to bulge out the branch point adenosine of the pre-mRNA (Yu et al., 2011). Specifically, Newby and Greenbaum (2002) showed that the presence of pseudouridine at U2 position 35 not only stabilizes the branch-site interactions but also changes the orientation of the bulged adenosine relative to the U2 snRNA-intron duplex. These findings suggest that pseudouridine better positions the branch-site adenosine for recognition and subsequent activity during splicing.

Ribosomal RNA can contain 1 – 100 pseudouridine modifications depending on the organism. There is a four- to eight-fold increase in the amount of pseudouridine incorporation in eukaryotes compared to prokaryotes. *Escherichia coli* contains only one pseudouridine in the small ribosomal subunit, while *Saccharomyces cerevisiae* 18S rRNA has up to 14 pseudouridines (Ofengand, 2002). When the pseudouridines of the small subunit were mapped to the three-dimensional ribosome structure, they appeared to be spread throughout the subunit. Conversely, within the large ribosomal subunit, most of the pseudouridine modifications can be mapped to functionally important regions of the ribosome. In *E. coli*, there are 10 pseudouridine modifications in the large ribosomal subunit, while in humans there are 55. The most highly conserved

pseudouridines are found in the loop of helix 69 (Ofengand, 2002). In *E. coli,* these residues, pseudouridine 1911, 1915, and 1917, help form an important bridge between the small and large ribosomal subunits, as well as interact with translation factors and tRNA during protein synthesis (O'Connor and Gregory, 2011). Pseudouridine residues are also frequently found to cluster around the peptidyl transferase centre. Bakin and Ofengand (1993) developed the *N*-cyclohexyl-*N'*-*β*-(4-methylmorpholinium)ethylcarbodiimide *p*-tosylate (CMCT) assay for detecting the presence of pseudouridine in large RNAs such as the 23S rRNA of *E. coli*. They found that pseudouridine residues were present at positions 746, 2457, 2504, 2580 and 2605 which are part of the peptidyl transferase centre and within close proximity to nucleotides that are directly involved in peptide bond formation (Bakin and Ofengand, 1993). This local clustering of pseudouridine residues may occur for structural reasons as pseudouridine can provide additional hydrogen bonds and confer an increased rigidity in the surrounding RNA structure. Additionally, Toh and Mankin (2008) demonstrated that *E. coli* strains lacking pseudouridines at positions 955, 2504, and 2580 show a strong increase in their susceptibility towards antibiotics targeting the peptidyl transferase centre of the ribosome, suggesting an intrinsic antibiotic resistance mechanism utilizing pseudouridine modifications. King and coworkers (2003) showed that in yeast, the deletion of individual pseudouridines from rRNA does not significantly affect the growth of the cells. However, when multiple pseudouridines were removed, synergistic effects were observed, resulting in impaired translation. In conclusion, the exact role pseudouridines play in ribosomal RNA is still under debate.

Transfer RNAs can have numerous modified nucleotides, and up to 25% of the nucleosides are post-transcriptionally modified or hypermodified. The universally conserved pseudouridine 55 is in the TΨC stem loop of all elongator tRNAs (Sprinzl et

al., 1999). The only tRNAs reported that do not contain pseudouridine 55 are eukaryotic cytoplasmic initiator tRNAs, tRNAs for Ala, Pro, Ser, Thr, and Val from *Mycoplasma mycoides* (Samuelsson et al., 1987), and two $tRNA^{Gly}$ species from *Staphylococcus epidermidis* that are involved in cell wall biosynthesis (Roberts, 1974). Additional pseudouridine modifications are commonly found in the D-arm and the anticodon stem loop of various tRNAs. Like the pseudouridines found in ribosomal RNA, those present in tRNA are not required for cell viability (Gutgsell et al., 2000; Raychaudhuri et al., 1999). An increase in local base stacking by pseudouridines was shown by Davis and coworkers (1995), where NMR analysis revealed that the N_1-H imino proton plays a role in stabilizing the conformation of the pseudouridine residue. This influence on the surrounding local RNA structure can be critical for tRNA binding to the ribosome and translational efficiency (Harrington et al., 1993; Yarian et al., 1999; Urbonavicius et al., 2002). Pseudouridine modifications, particularly pseudouridine 55, have also been proposed to aid in extreme temperature resistance in both *E. coli* and the thermophile *Thermus thermophilus* (Kinghorn et al., 2002; Ishida et al., 2011). Notably, Ishida and coworkers (2011) observed an increase in Gm18, m^5s^2U54, and m^1A58 levels when pseudouridine 55 was absent, signifying that pseudouridine 55 may play an important regulatory role in other tRNA modifications.

1.4 Pseudouridine Synthases

Pseudouridine synthases belong to six different families: TruA, TruB, TruD, RluA, RsuA, and Pus10. The first four families were characterized initially based on their sequence similarity (Koonin, 1996; Kaya and Ofengand, 2003). Later, the TruD family was identified through biochemical and gene sequencing techniques (Kaya and Ofengand, 2003). Eventually, Pus10 was identified as a representative of a sixth family found only

in archaea and eukaryotes based on its weak sequence similarity to Cbf5 (Watanabe and Gray, 2000; Roovers et al., 2006). Each family is named after the *E. coli* representative enzyme, whereas Pus10 (not found in bacteria) is named for the human protein. Crystal structures have been determined for proteins belonging to each family (Hoang and Ferré-D'Amaré, 2001; Hoang et al., 2006; Phannachet et al., 2004; Foster et al., 2000; Hoang et al., 2004; Sivaraman et al., 2002; McCleverty et al., 2007). The amino acid sequences of all pseudouridine synthases are poorly conserved; however, structural comparison reveals a common catalytic core (Figure 1.2). This conserved core is composed of an eight-stranded mixed β-sheet with adjacent helices and loops. The only absolutely conserved residue, the catalytic aspartate, resides on a loop that occupies part of the catalytic cleft. Outside the catalytic core, a variety of additional domains can be found. In members of the RluA (specifically RluC and RluD) and RsuA families, an N-terminal extension resembles ribosomal protein S4 (Mizutani et al., 2004; Sivaraman et al., 2002). The enzymes belonging to the TruB family contain a C-terminal extension termed the PUA domain, due to its presence in *p*seudo*u*ridine synthases and in some *a*rchaeosine-transglycosylases (Hoang and Ferré-D'Amaré, 2001; Phannachet et al., 2004; Sabina and Soll, 2006). The RluA and RsuA families are the most closely related pseudouridine synthases, sharing three conserved sequence motifs (motifs I, II, and III). The TruB family also resembles RluA and RsuA families but lacks motif III (Koonin, 1996; Hoang and Ferré-D'Amaré, 2001). Structural analysis also revealed that members of the TruD family have a circular permutation of the order of secondary structural elements within the catalytic core (Ericsson et al., 2004; Hoang et al., 2004). Interestingly, TruD is also found in all three domains of life (Kaya and Ofengand, 2003) andbased on these findings it is hypothesized that enzymes of the TruD family diverged first from other pseudouridine synthases (Ericsson et al., 2004). Members of the TruA

family are also unique among pseudouridine synthase enzymes, as they are the only pseudouridine synthase to function as a dimer (Foster et al., 2000).

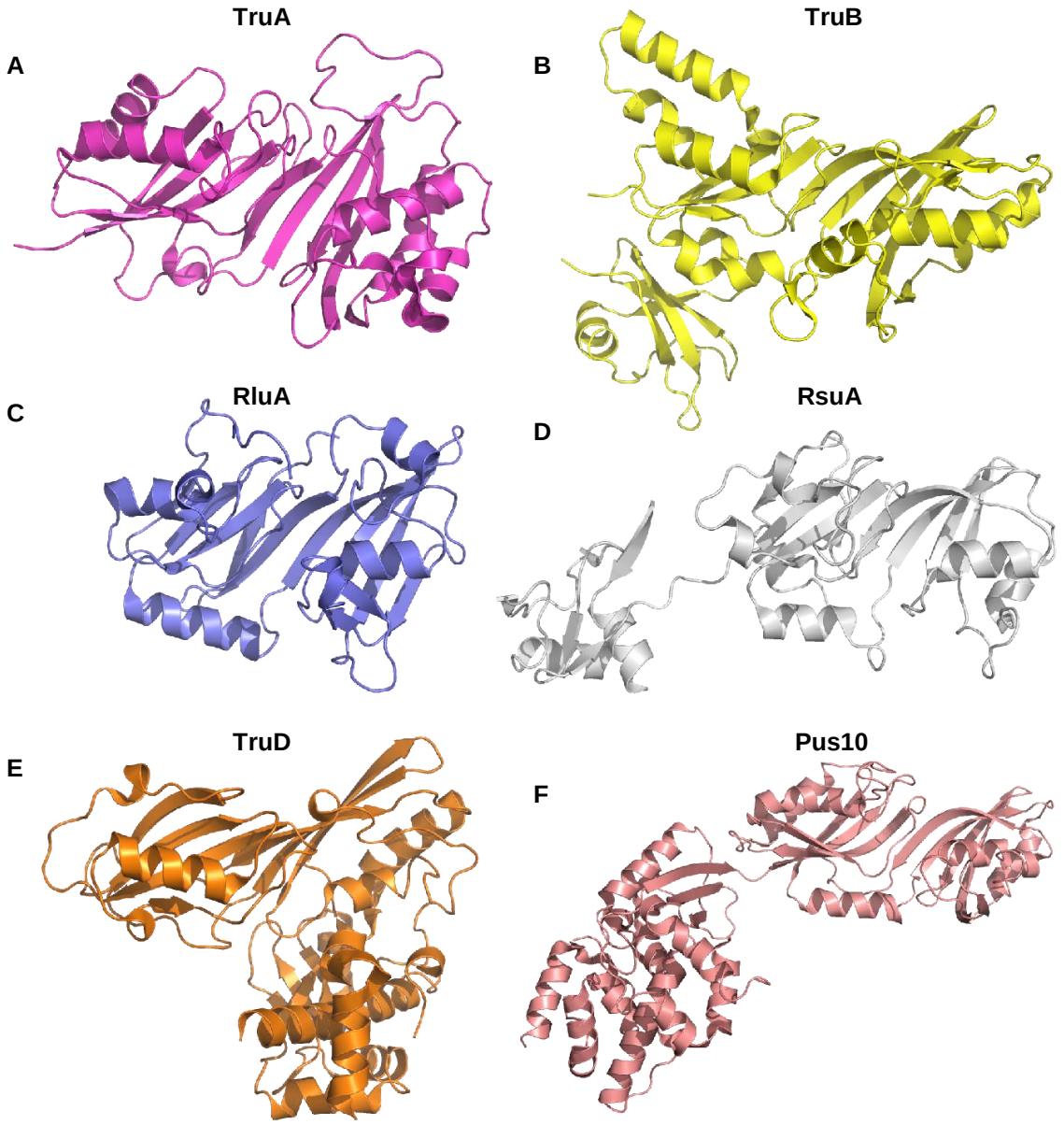


Figure 1.2: Structures of representative enzymes from all six pseudouridine synthase families. Comparison of ribbon representations of the structures of A. TruA (1DJ0; Foster et al., 2000), B. TruB (1K8W; Hoang and Ferré-D'Amaré, 2001), C. RluA (2I82; Hoang et al., 2006), D. RsuA (1KSK; Sivaraman et al., 2002), and E. TruD (1SZW; Ericsson et al., 2004), all from *E. coli*. F. Human Pus10 (2V9K; McCleverty et al., 2007). All structures show the catalytic domain in the same orientation.

Within all pseudouridine synthases, the strictly conserved catalytic aspartate residue is found within motif II in the active site cleft (Koonin, 1996). Biochemical and structural studies have confirmed the importance of the conserved aspartate residue in pseudouridine formation. Site-directed mutagenesis in members of five pseudouridine synthase families has demonstrated that this residue plays a critical role in pseudouridine catalysis, rather than in RNA binding or substrate recognition (Huang et al. 1998; Conrad et al. 1999; Ramamurthy et al. 1999; Raychaudhuri et al. 1999; Zebarjadian et al. 1999; Del Campo et al. 2001; Kaya and Ofengand 2003; Chan and Huang, 2009). Structurally, the catalytic aspartate residue has been shown to be within close proximity to the target uridine in RNA-protein co-crystal structures (Figure 1.3; Phannachet et al., 2004; Hoang et al., 2006). Motif II also contains a histidine (in TruB family members) or an arginine (in RluA, RsuA, and TruA enzymes) that intercalates into the RNA strand and ensures that the target uridine flips into the active site (Hoang and Ferré-D'Amaré, 2001; Hoang et al., 2006). Additionally, a tyrosine (or phenylalanine in TruD) is conserved in a K/RxY motif and helps maintain the structural integrity of the active site through hydrophobic interactions by stacking against the uracil , as well as acting as a general base through its hydroxyl group for the proton abstraction during the final step of catalysis (Phannachet et al., 2005). Meanwhile, the K/R residue in this K/RxY motif interacts with the phosphate of the target nucleotide through a salt bridge possibly positioning it for catalysis (Hoang and Ferré-D'Amaré, 2001; Pan et al., 2003; Phannachet et al., 2004). In contrast, motif I is not involved in catalysis but instead functions as a support to reinforce the active site loop in motif II (Hoang et al., 2005; Hamma et al., 2005; Spedaliere et al., 2000). In motif III, a conserved lysine or arginine residue makes a salt bridge with the catalytic aspartate (Hoang et al., 2006). Although TruB lacks motif III, it too has an arginine present that makes a salt bridge with the

catalytic residue. This salt bridge has been proposed to possibly activate the catalytic aspartate residue as a nucleophile by deprotonating the carboxylate (Hoang and Ferré-D'Amaré, 2001).

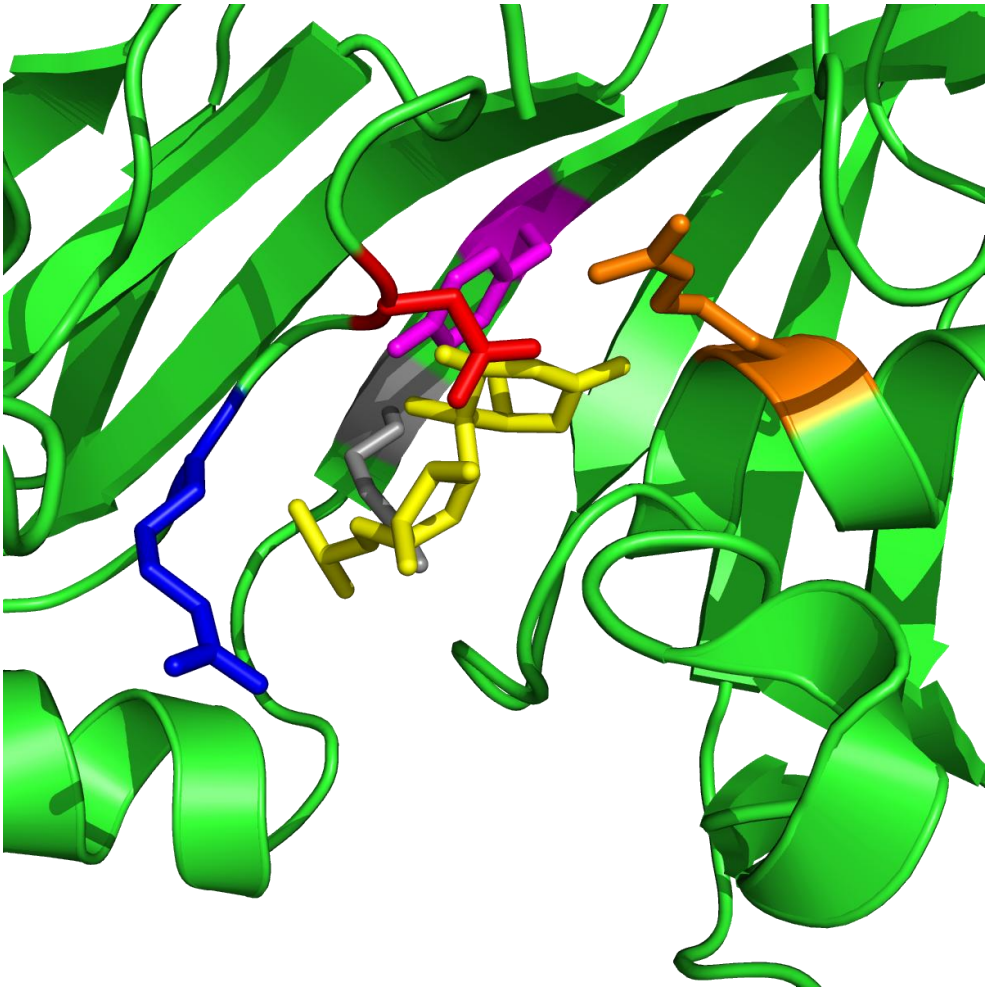


Figure 1.3: Active site of *E. coli* RluA. The conserved catalytic aspartate residue is shown in red. The arginine residue (R62) that intercalates into the RNA (not shown here) is depicted in blue. Both residues belong to motif II. The conserved KxY motif is shown in grey and fuchsia. Arginine 165 (orange) is proposed to form a salt bridge with the catalytic aspartate residue. The target uridine analog (5-fluorouracil) is shown in yellow.

Although much is understood about the residues involved in pseudouridine catalysis, the exact chemical mechanism is still not known. As all pseudouridine synthases share the same catalytic fold containing the conserved catalytic aspartate residue and have also

been proposed to have the same common evolutionary ancestor, it is very likely that they all share the same catalytic mechanism (Mueller, 2002). For all pseudouridine synthases the same chemistry is completed within their catalytic cleft. The first step is to disconnect the uridine base from the ribose sugar, breaking the N-C glycosidic bond. The base is then rotated or flipped still within the catalytic pocket and finally reattached to the ribose through the C5 position (Zhou et al., 2010). Gu and coworkers (1999) showed that TruA was able to form a covalent adduct with a 5-fluorouracil substituted tRNA substrate, suggesting a putative intermediate of the reaction pathway. From their findings it was proposed that the catalytic aspartate residue adds to the C6-position (Figure 1.4A) of the target uridine to form a stable Michael adduct, where the final product is released by hydrolysis of the ester linkage between the active-site aspartate and the pyrimidine ring (Gu et al., 1999). In contrast, co-crystal structures of RluA and TruB with 5-fluorouracil substituted RNA substrates showed that no covalent adduct formed and that the 5-fluoruracil was rearranged to a C-glycoside (like pseudouridine) (Figure 1.4B; Hoang and Ferré-D'Amaré, 2001; Hoang et al., 2006). These findings suggest an acylal intermediate where the catalytic aspartate attacks the C1 of the ribose ring. Subsequent studies revealed direct hydration of the 5-fluorouracil product rather than the proposed ester hydrolysis, indicating TruA, TruB, and RluA all handle 5-fluorouracil containing substrates in the same manner (McDonald et al., 2011). A third mechanism has been proposed by Miracco and Mueller (2011). Based on their studies they suggest a glycal intermediate reaction pathway (Figure 1.4C). They found that TruB converts 5-fluorouridine in RNA into two isomeric hydrated products. Here, as a minor product of the reaction, the pentose ring is epimerized to an arabinose and this inversion

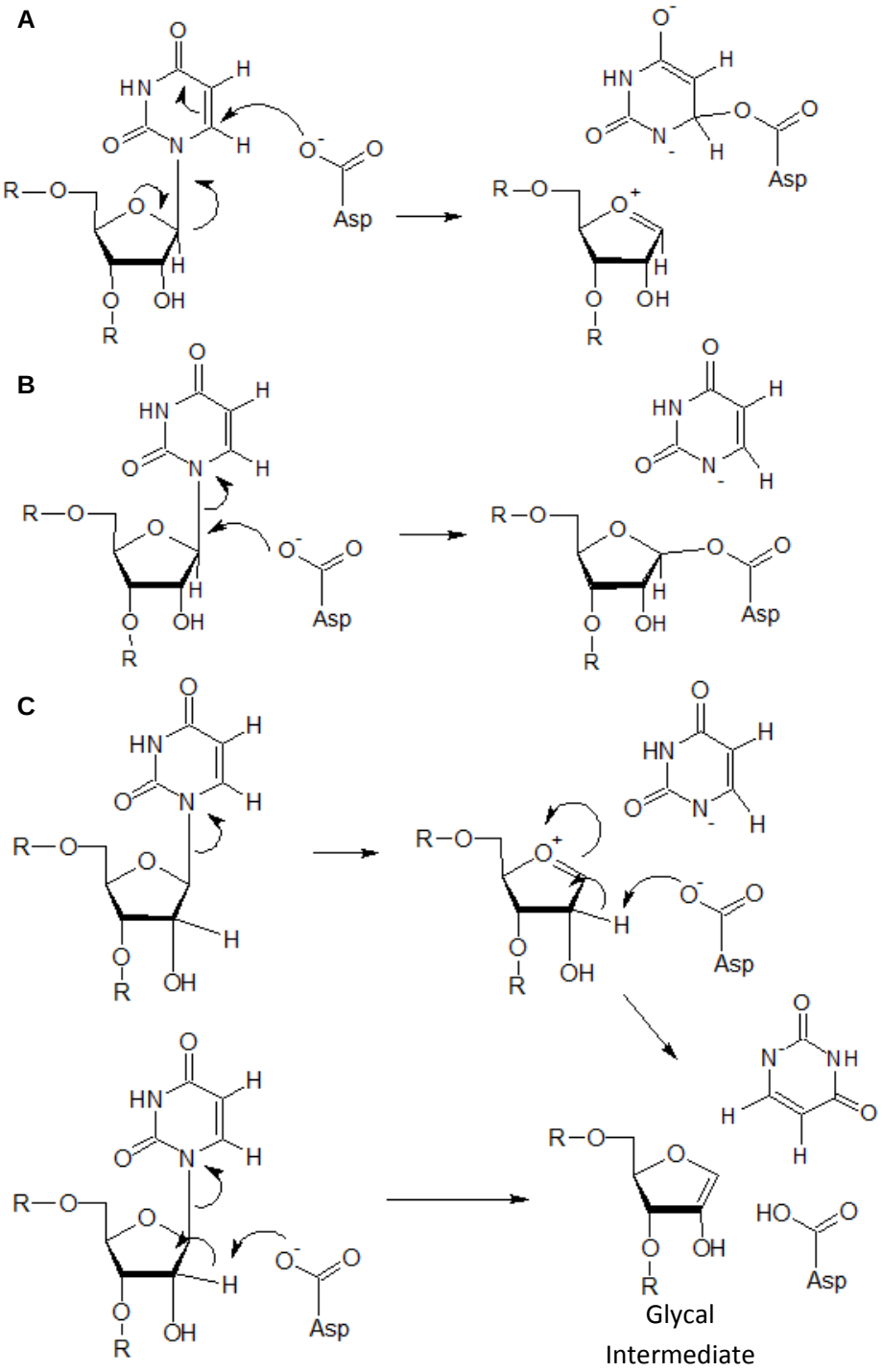


Figure 1.4: Proposed catalytic mechanisms of pseudouridine formation. A. According to the Michael addition hypothesis, the catalytic aspartate attacks the C_6 position of the uracil base. B. The aspartate attacks the C1' position of the ribose sugar as suggested in the acylal intermediate mechanism. Figure adapted from Zhou et al. (2010). C. Glycal mechanism for pseudouridine formation; uracil extraction can be stepwise (top) or concerted (bottom path). Figure adapted from Miracco and Mueller (2011).

suggests another chemical mechanism, but strongly disfavours the Michael addition (Miracco and Mueller, 2011). The proposed glycal mechanism is in accordance with their new findings, but still awaits experimental confirmation. Therefore, the exact mechanism of pseudouridine formation is still contested and more analysis is needed.

1.5 H/ACA small ribonucleoproteins

In archaea and eukaryotes, most pseudouridine modifications are produced by RNA-protein complexes called H/ACA small ribonucleoproteins (Kiss et al., 2010). These H/ACA small ribonucleoproteins (H/ACA sRNPs) contain four protein subunits, Cbf5, L7Ae (Nhp2 in eukaryotes), Nop10, and Gar1, along with a guide RNA (Figure 1.5). Archaeal H/ACA RNAs are typically composed of 60 – 75 nucleotides that fold into a long hairpin (Dennis and Omer, 2005). In most eukaryotes such as yeast and humans, the H/ACA RNA forms two hairpins each containing a pseudouridylation pocket. These stem loops are connected by a single-stranded hinge (H) region (AnAnnA) and followed by the ACA box motif at the 3'end of the RNA (Balakin et al., 1996). Contained within the hairpin of the guide RNA is the pseudouridylation pocket. Here, the guide RNA can base pair with the target RNA sequence to position the uridine within the catalytic pocket. Cbf5 (dyskerin in humans), the catalytic subunit of these ribonucleoproteins, is a homolog of the *E. coli* TruB enzyme (Lafontaine et al., 1998).

Several studies have demonstrated that Cbf5 and L7Ae directly and independently bind to the H/ACA guide RNA (Baker et al., 2005; Charpentier et al., 2005; Li et al., 2006). Cbf5 requires the presence of the ACA motif to bind to the guide RNA, but will also independently interact with Nop10 and Gar1 to form a stable guide-RNA independent subcomplex (Rashid et al., 2006). Nop10 has been shown to interact with a conserved stretch of amino acids close to the catalytic centre of Cbf5. In other pseudouridine

synthases, this region (motif I in stand-alone pseudouridine synthases; see above) helps stabilize the catalytic cleft, and it was hypothesized Nop10 may contribute to the overall stability of Cbf5's catalytic core (Hamma et al., 2005).

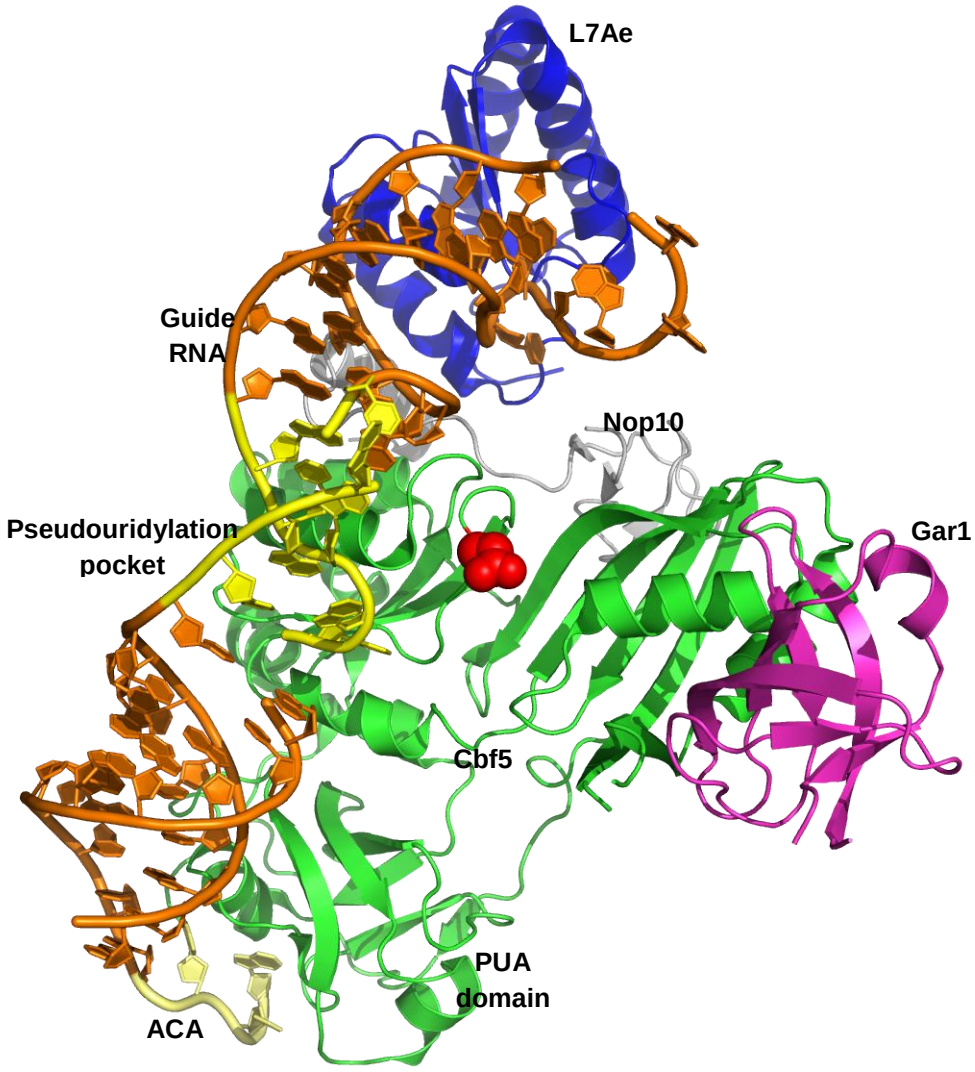


Figure 1.5: Organization of the H/ACA small ribonucleoprotein complex. The x-ray crystal structure (2HVY) of archael H/ACA small ribonucleoprotein (Li and Ye, 2006). The catalytic subunit Cbf5 is shown in complex with accessory proteins Gar1, Nop10, and L7Ae, as well as guide RNA. The catalytic aspartate (D95) is shown as red spheres. In yellow is the pseudouridylation pocket where substrate RNA binds to the guide RNA. The ACA motif is shown interacting with the PUA domain (light yellow).

Although Gar1 shares many structural similarities with the RNA-binding motifs of bacterial translation initiation (IF2) and elongation (EF-Tu) factors, it does not directly interact with the guide RNA. Instead Gar1 binds to the RNA-binding thumb loop of Cbf5 and regulates substrate turnover (Li et al., 2011; Rashid et al., 2006). Duan and coworkers (2009) showed that Gar1 binds to Cbf5's thumb loop when substrate RNA is not present. However, once substrate RNA binds, the thumb loop has extensive interactions with the substrate RNA. To release the product RNA, the thumb loop interactions must first be broken and it was proposed that Gar1 may provide a low-energy binding site for the thumb loop, therefore contributing to its dissociation from the RNA (Duan et al., 2009).

1.6 RluA

In prokaryotes, pseudouridine synthases typically function as stand-alone enzymes and consist of only a polypeptide chain without guide RNA. Two well studied examples of bacterial stand-alone pseudouridine synthases are *E. coli* TruB and RluA. RluA is responsible for the formation of Ψ746 in 23S rRNA and Ψ32 in $tRNA^{Phe}$, $tRNA^{Cys}$, $tRNA^{Leu4}$, and $tRNA^{Leu5}$. Wrzesinski and coworkers (1995) first identified RluA as a pseudouridine synthase responsible for the formation of Ψ746 in ribosomal RNA from *E. coli*. The surprising result was that RluA also forms Ψ32 in several tRNA species. *E. coli* RluA is a 219-amino acid protein with a molecular weight of approximately 25 kDa (Wrzesinski et al., 1995). The crystal structure of RluA shows a protein that adopts an elongated, mixed α/β fold bisected by the catalytic cleft (Figure 1.6; Hoang et al., 2006).

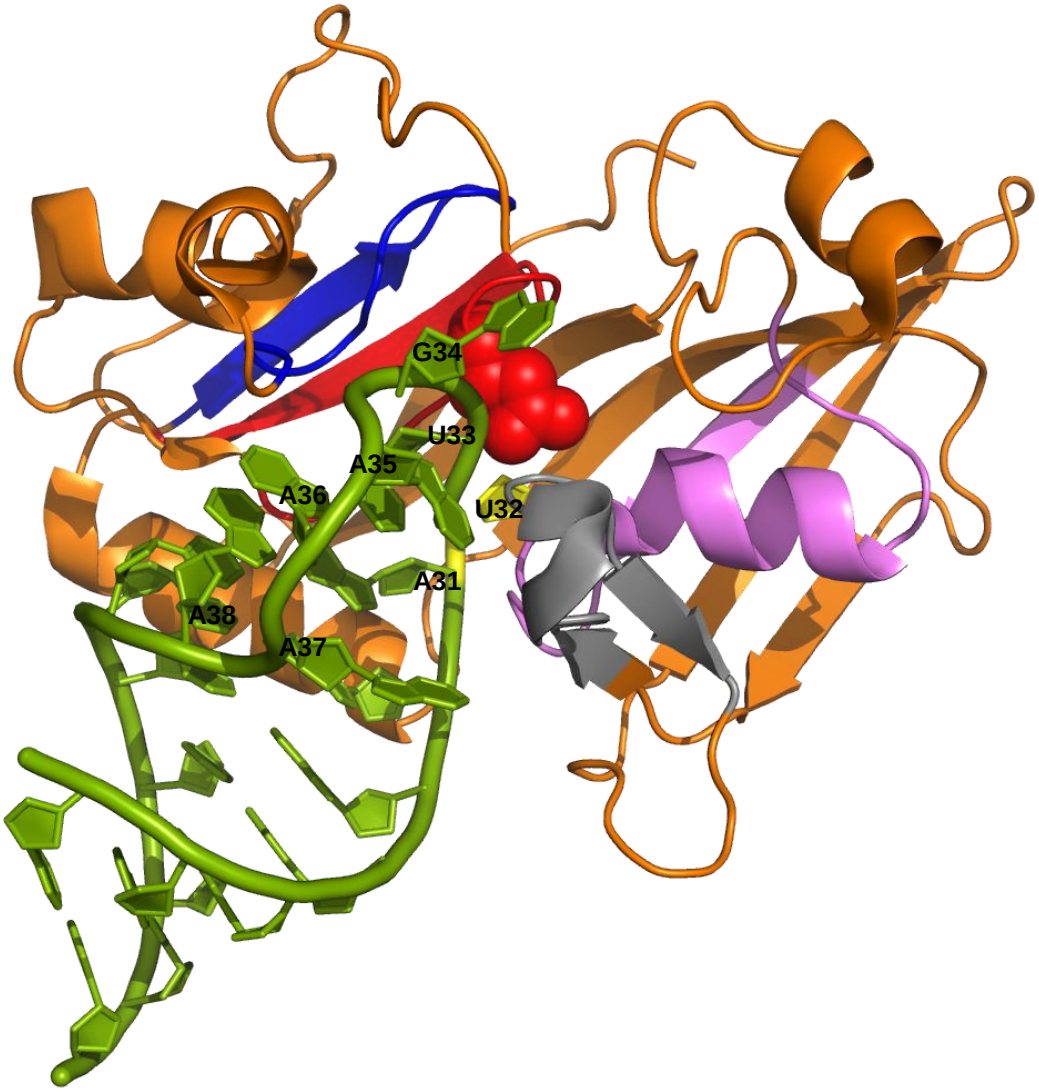


Figure 1.6: X-ray structure of *E. coli* RluA. The cartoon representation of RluA (2I82; Hoang et al., 2006) in complex with anticodon stem loop (green) shows the conserved catalytic cleft and the catalytic aspartate residue (red spheres). Motif I (blue), motif II (red), motif III (pink), and the thumb loop (grey) are also highlighted. The target uridine is shown in yellow.

As described above, RluA contains the conserved catalytic fold found in all pseudouridine synthases consisting of 8 β-strands with additional flanking helices and loops. The co-crystal structure of RluA in complex with the anticodon stem loop (ASL) of $tRNA^{Phe}$ shows that motif II, containing the catalytic aspartate, forms a protrusion that

packs against the minor groove face of the ASL. Motif III was also shown to interact with the RNA backbone of the ASL and additionally forms part of the active site cavity (Hoang et al., 2006). RluA has a compact "thumb" structure which lies between strands β6 and β9 in the catalytic cleft and interacts with the major groove of its RNA substrate (Hoang et al., 2006). Like all other pseudouridine synthases, RluA requires a conserved aspartate residue (D64) to complete catalysis (Raychaudhuri et al., 1999). In order to gain a better understanding of the role of the conserved aspartate in pseudouridine formation, Ramamurthy and coworkers (1999) mutated RluA's aspartate 64 to either alanine or cysteine and showed that this resulted in the loss of enzymatic activity. *In vitro* binding assays also demonstrated that these protein variants retainRNA binding abilities. Furthermore, Ramamurthy et al. (1999b) also found that substitution of conserved cysteine residues does not significantly alter the catalytic activity of either RluA or TruB. Additionally, Ramamurthy and coworkers (1999b) determined the catalytic constant (k_{cat}) of pseudouridine formation by RluA to be 0.099 ± 0.003 s^{-1} using a full-length $tRNA^{Phe}$ transcript. When a truncated RNA substrate corresponding to the anticodon stem loop was used in the assay, a similar catalytic constant, 0.068 s^{-1}, was determined (Hamilton et al., 2006).

The 2904 nucleotide 23S rRNA and ~76 nucleotide tRNA substrates of RluA share very little structural similarity. However, all five substrates share a conserved sequence surrounding the target nucleotide, ΨUXXAAA, where X can be any canonical ribonucleotide (Raychaudhuri et al., 1999; Hoang et al., 2006). As RluA binds to a tRNA anticodon stem loop substrate, several major changes occur within the RNA structure as seen in the co-crystal structure of RluA with RNA (Figure 1.6). Residues 32, 34, and 37 are flipped out from the helical stack of the stem loop. The empty space once occupied by residue 37 is filled by A36, thus making a non-sequential stacking interaction with

A38. Next, U33, which normally makes hydrogen bonds with backbone phosphates in the free tRNA structure, forms a reverse-Hoogsteen base pair with A36. Finally, A35 is rotated 180° and stacks underneath U33. As U32 is now flipped into the active-site of RluA, the gap between residues 31 and 33 is occupied with the guanidinium group from the side chain of arginine 62 from RluA. Arginine 62 is absolutely conserved in all members of the RluA, RsuA, and TruA families and most likely plays a similar role in substrate base-flipping. Numerous interactions between RluA's Arginine 62 and the RNA substrate are made, including a bidentate salt bridge with the phosphate group of U33 and two water-mediated hydrogen bonds with A38 (Hoang et al., 2006). This large structural rearrangement of the anticodon is unique to RluA. When tRNA is bound to the decoding centre of the ribosome, the canonical U turn conformation of the anticodon is maintained (Ogle et al., 2001). The phenylalaninyl-tRNA synthetase also recognizes the same anticodon stem loop as RluA, but only induces a small conformational distortion of the anticodon U turn (Goldgur et al., 1997). The U33•A36 (U747•A750 in 23S rRNA) reverse-Hoogsteen base pair has been shown to be critical in substrate recognition by RluA, and thus substrate specificity is achieved through an indirect readout of the protein-induced RNA structure (Hoang et al., 2006).

1.7 TruB

The universally conserved pseudouridine 55 in the TΨC arm of all elongator tRNAs is formed by TruB. Nurse and coworkers (1995) were the first to identify and characterize the 314-amino acid protein TruB (molecular weight of ~40 kDa) from *E. coli*. Similar to RluA, TruB adopts a mixed α/β fold in its catalytic domain, but TruB also has a distinct C-terminal domain called the PUA domain (Figure 1.7; Hoang and Ferré-D`Amaré, 2001). The catalytic domain is comprised of 11 antiparallel β-strands, decorated with

multiple helices and loops, while the catalytic cleft bisects this fold. The PUA domain, a roughly spherical domain consisting of a four-stranded β-sheet and one α-helix (Hoang and Ferré-D`Amaré, 2001). The co-crystal structure of TruB in complex with the TΨC arm of $tRNA^{Phe}$ depicts the RNA bound by a deep cleft on the surface of the enzyme. Residues from strands β8, β9, and helix α4, form a thumb-like structure that pinches the major groove of the RNA. Two insertions differentiate TruB from Cbf5 of the H/ACA small RNP complexes. Insertion 2 forms the majority of the thumb-loop as described above, while insertion 1 forms part of the RNA binding cleft (Hoang and Ferré-D`Amaré, 2001). In solution, the T loop of the TΨC stem loop was found to be very well ordered (Koshlap et al., 1999). When bound to TruB, the overall loop structure remained in this intact, folded conformation, indicating that TruB binds to a preorganized T-loop. However, when bound to TruB, nucleotides 55, 56, and 57 are flipped into the active site cleft, disrupting the normal helical stacking to the T arm. The role of arginine 62 in RluA is completed by histidine 43 in TruB which intercalates into the RNA to facilitate base-flipping into the active site. The imidazole ring of the histidine residue is inserted into the T loop and stacks underneath the U54•A58 reverse Hoogsteen base pair (Hoang and Ferré-D`Amaré, 2001). In the intact tRNA structure this position would normally be filled by G18 from the D loop. Hydrogen bonds between histidine's main chain amide and carbonyl groups help to stabilize its interaction with A58. U55 is forced out of the folded RNA structure to avoid steric clashes with the polypeptide chain C-terminal to histidine 43. The catalytic aspartate 48 residue is positioned near the target U55 by histidine 43 stacking under the reverse Hoogsteen base pair of U54•A58. Aspartate 48 also forms a salt bridge with arginine 181 which may activate the catalytic residue as a nucleophile by deprotonating the carboxylate (Hoang and Ferré-D`Amaré, 2001).

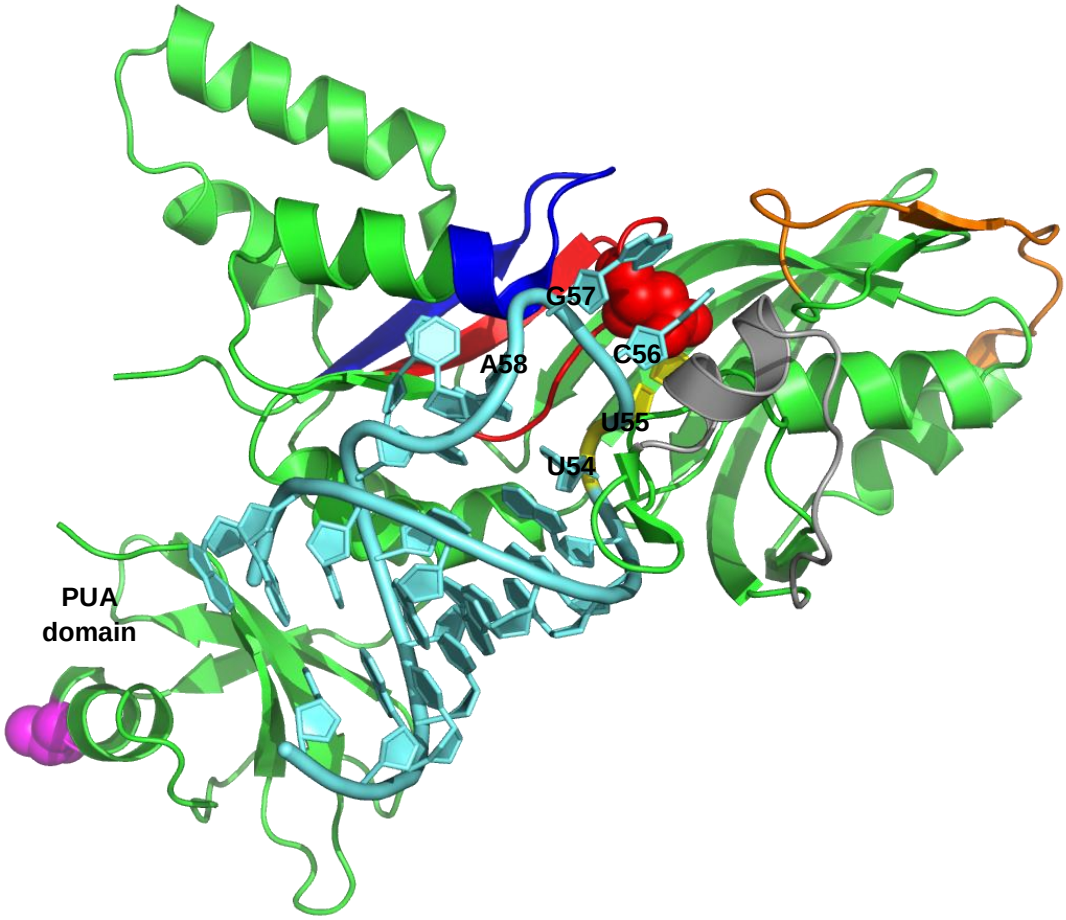


Figure 1.7: Co-crystal structure of TruB in complex with T-arm. This x-ray structure from Hoang and Ferré-D'Amaré (2001) (1K8W) depicts TruB (green) with truncated RNA (cyan) corresponding to the T-arm of $tRNA^{Phe}$. Motif I (dark blue), motif II (red), and insertions 1 (orange) and 2 (grey) are also shown. The catalytic aspartate residue is shown as red spheres. Threonine 259 is shown as pink spheres. U55 is in yellow. The C-terminal PUA domain and bases U54 – A58 are labelled.

U55 also forms a stacking interaction with tyrosine 76 in the active site of TruB. This same tyrosine residue is conserved in pseudouridine synthases from RluA, RsuA, TruA, TruB, and Pus10 families (TruD has a phenylalanine in its place). Phannachet and coworkers (2005) examined the role of tyrosine 76 in pseudouridine formation and showed that the tyrosine side chain may play a dual role within the active site. The phenyl ring helps to stabilize the active site by stacking against the target base and has

also been proposed to act as a general base through its hydroxyl group allowing for proton extraction during the last step of catalysis (Phannachet et al., 2005).

Several biochemical studies have been completed to investigate the roles of conserved amino acids in the chemical mechanism of pseudouridine formation (Ramamurthy et al., 1999; Ramamurthy et al., 1999b; Phannachet et al., 2005). Like in all other pseudouridine synthases, the catalytic aspartate is required for Ψ55 synthesis by TruB, but is not involved in RNA binding (Ramamurthy et al., 1999). When this aspartate was mutated to alanine or cysteine, TruB was still able to bind full-length tRNA substrate equally well as the wild-type enzyme. In an additional study, Ramamurthy et al. (1999b) determined the catalytic constant (k_{cat}) of pseudouridine formation to be 0.12 ± 0.01 s^{-1} using full-length $tRNA^{Phe}$ as the substrate. Additional kinetic studies demonstrated that a truncated RNA consisting of 17 nucleotides corresponding to the T-arm of yeast $tRNA^{Phe}$ could also be used as a catalytic substrate with a similar k_{cat} of 0.24 s^{-1} (Gu et al., 1998). These catalytic constants are a combination of several different steps along the reaction pathway. Wright and coworkers (2011) demonstrated that TruB has a single-round rate constant of pseudouridylation (k_{ψ}) of approximately 0.5 s^{-1}.

Unique to pseudouridine synthase members of the TruB family is the C-terminal PUA domain. This domain is also found in other RNA-modifying enzymes such as archaeosine transglycosylases and RNA methyltransferases, as well as in archaeal sulfate reductases and bacterial and yeast glutamate kinases (Ferré-D'Maré, 2003; Hur et al, 2006; Aravind and Koonin, 1999; Anantharaman et al., 2002; Hallberg et al., 2007). As discussed briefly above, PUA domains are typically composed of several β-strands coiled to form a pseudobarrel, which is closed on one side by a short α-helix (Perez-Arellano et al., 2007). In the TruB co-crystal structure from Pan et al. (2003), two RNA

stem-loops form a discontinuous and bulged duplex, reminiscent of the tRNA acceptor arm minus the CCA-3' terminus. This duplex binds to the PUA domain of TruB through minor grove interactions with small polar residues of the α-helix and β2 loop (Pan et al., 2003; Perez-Arellano et al., 2007). The crystal structure of *Pyrococcus furiosus* H/ACA small RNP revealed that Cbf5's PUA domain recognizes the 3'-ACA trinucleotide in a sequence specific manner and contributes to align the H/ACA RNA within the active site of the enzyme (Li and Ye, 2006). The deletion of the PUA domain of Cbf5 from *P. abyssi* demonstrated the critical role this domain plays in the function of H/ACA sRNPs (Manival et al., 2006). Interestingly, the majority of disease-related mutations in the human Cbf5 homolog, dyskerin, cluster in the PUA domain. These mutations cause the disease dyskeratosis congenita, characterized by nail dystrophy, abnormal skin pigmentation, bone marrow failure, and higher instances of epithelial cancers (Heiss et al., 1998; Knight et al., 1999; Mochizuki et al., 2004). Additionally, mutations within the PUA motif of the oncogene MCT-1 protein can significantly affect translation regulation (Reinert et al., 2006). In contrast, very little is known about the role of TruB's PUA domain in RNA binding or catalysis.

1.8 Objectives

Previous analysis of the chemical mechanism of pseudouridine formation has not yet revealed the exact steps of catalysis. Some studies have determined an overall rate of reaction, but no analysis of the individual steps has been conducted. As a first step towards gaining a better understanding of the reaction mechanism, we sought to analyse the kinetics of pseudouridylation by the model pseudouridine synthase RluA. The aim of this analysis was to determine the rate-limiting step for pseudouridine formation (catalysis vs. substrate binding or product release). Additionally, during catalysis the uracil base rotates within the catalytic pocket. By trying to exchange this base with a radiolabelled uracil, we could potentially provide direct evidence towards a chemical mechanism where no covalent adduct may be formed between the protein and nucleobase during catalysis.

Currently there are several co-crystal structures of pseudouridine synthases in complex with RNA; however, no crystal structure has been determined for TruB or RluA using the full-length substrate. Both TruB and RluA have been shown to interact with short stem loops corresponding to $tRNA^{Phe}$ T-arm and anticodon stem loop, respectively (Hoang and Ferré-D'Maré, 2001; Hoang et al., 2006). Potentially, there could be contacts that are missing in the crystal structures that could contribute to binding interactions between the enzyme and RNA. Furthermore, preliminary studies in the Kothe lab indicate that a stem loop might not interact with TruB in a similar manner as the full-length tRNA in contrast to previous reports (unpublished studies; Gu et al., 1998). Therefore, it was our aim to examine whether these truncated substrates behaved similarly to the full-length tRNA upon binding to these pseudouridine synthases. In addition, given that pseudouridine and uridine differ only in the glycosidic bond attachment, it was our goal

to gain a better understanding of how pseudouridine synthases differentiate between uridine- and pseudouridine-containing tRNAs.

The pseudouridine synthesis activity of H/ACA small RNPs requires the PUA domain of the catalytic subunit Cbf5 to interact with the ACA trinucleotide of the guide RNA (Li and Ye, 2006; Manival et al., 2006). The *E. coli* homolog TruB also contains a C-terminal PUA domain. Previously, this region has been shown to interact with an acceptor-arm mimic in a co-crystal structure (Pan et al., 2003; Hoang and Ferré-D'Amaré, 2001). No role has yet been described for TruB's PUA domain. Therefore, the function of TruB's PUA domain in substrate binding and catalysis was examined.

Chapter 2 - Materials and Methods

2.1 Buffers and Reagents

Buffer TAKEM$_4$: 50 mM Tris-HCl pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 1 mM EDTA, 4 mM $MgCl_2$. All nucleotide triphosphates and guanosine monophosphate for *in vitro* transcriptions were from Sigma. Radioactive [C5-^{3}H] UTP was purchased from Moraveck. Inorganic pyrophosphatase was from Sigma; all other enzymes were from Fermentas. The fluorescent dyes 5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid (1,5-IAEDANS) and fluorescein-5-thiosemicarbazide were purchased from LifeTechnologies (formerly Invitrogen). Chemicals were from VWR.

2.2 Mutagenesis

The QuikChange® method (Stratagene) was used for site-directed mutagenesis generating plasmids pET28a-TruBC58AC174AC193AT259C, pET28a-TruBΔPUA, pET28a-TruBD48NΔPUA, and pCA24N-RluAD64N. All reactions were carried out using a MyCycler™ thermo cycler (BioRad) and the conditions outlined in Table 2.1. Primers are given in Table 2.2. The melting temperature (T_m) for each primer was calculated using the Stratagene formula:

$$T_m = 81.5 + 0.41(\%GC) – 675/N - \%mismatch$$

Where N is the primer length in bases, and values for %GC and %mismatch are whole numbers.

Table 2.1: QuikChange® mutagenesis protocol for engineering TruBC58AC174AC193AT259C, TruBΔPUA, TruBD48NΔPUA, and RluAD64N

Step	Temperature (°C)	Time	Cycles
Initial denaturation	95	5 minutes	1 cycle
Denaturation	95	45 seconds	18 cycles
Annealing	15 below T_m	1 minute	
Extension	68	15 minutes	
Final extension	68	15 minutes	1 cycle

Table 2.2: Primers for mutagenesis reaction of TruB and RluA

Primer name	T_m (°C)	Sequence
TruBT259C sense	65.5	5’ – CCG GTG GTG AAT CTT CCG TTA TGC TCT TCT GTT TAC TTC AAA AAT GG – 3’
TruBT259C antisense	65.5	5’ – CCA TTT TTG AAG TAA ACA GAA GAG CAT AAC GGA AGA TTC ACC ACC GG – 3’
RluAD64N sense	68.7	5’ – GCA GGC AGA ATC AGT GCA CCG TCT GAA TAT GGC TAC CAG C – 3’
RluAD64N antisense	68.7	5’ – GCT GGT AGC CAT ATT CAG ACG GTG CAC TGA TTC TGC CTG C – 3’
TruBΔPUA sense	79.2	5‘ - GCC AAT GGA CAG TCC AGC TTA GTA GTA CCC GGG GGT GAA TCT TCC GTT - 3'
TruBΔPUA antisense	79.2	5' - AAC GGA AGA TTC ACC CCC GGG TAC TAC TAA GCT GGA CTG TCC ATT GGC - 3'

Each reaction was prepared using the same PCR conditions where the final concentrations were 1x Pfu buffer with $MgSO_4$, 0.4 mM dNTPs, 0.4 μM forward primer, 0.4 μM reverse primer, 0.12 U/μL Pfu DNA polymerase, and 0.5 to 1 μg template DNA. As template DNA, the plasmids pET28a-TruBC58AC174AC193A, pET28a-TruB,

pET28a-TruBD48N (Wright et al., 2011), and pCA24N-RluA (Keio collection, Kitagawa et al., 2005) were used which were previously generated and purified in the Kothe lab.

Following mutagenesis, the template DNA was digested using DpnI and samples were analyzed using agarose gel electrophoresis. The remaining DNA was used to transform *E. coli* DH5α cells. These cells were grown overnight at 37°C on kanamycin (pET28a plasmids) or chloramphenicol (pCA24N plasmids) containing LB plates. Colonies were picked the following day, grown overnight in the appropriate antibiotic-containing LB media, and plasmid DNA was isolated using a BioBasic MiniPrep kit. Plasmids were analysed for the correct mutation by restriction with KspAI (TruB) or Alw44I (RluA), followed by agarose gel electrophoresis. Sequencing by GeneWiz (South Plainfield. NJ) confirmed the correct mutations. These plasmids were then transformed into BL21(DE3) for pET28a variants and AG1 (ME5305) *E. coli* (Keio collection, Kitagawa et al., 2005) for pCA24N variants.

2.3 Protein expression

Recombinant hexahistidine-tagged RluA and RluAD64N were expressed from the pCA24N (GFP minus)-JW0057 plasmid provided by the National BioResource Project (NIG, Japan, Keio collection, Kitagawa et al., 2005) in AG1 (ME5305) *E. coli* cells in the presence of 50 µg/mL chloramphenicol. Kanamycin (50 µg/mL) was added to LB medium when TruBC58AC174AC193AT259C, TruBΔPUA, and TruBD48NΔPUA were overexpressed from pET28a plasmids in BL21(DE3) *E. coli* cells. Cultures were grown in 500 mL LB with 50 µg/mL chloramphenicol at 37°C starting at an optical density at 600 nm (OD_{600}) of 0.1. Protein expression was induced once cell growth had reached 0.6 OD_{600} by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1.0 mM. Cell cultures were grown for an additional 3 hours before being harvested by centrifugation at 5000×g for 15 minutes using a JA-14 rotor (Beckman).

Cells were flash frozen in liquid nitrogen and stored at -80°C for future use. Protein expression was monitored by removing 1 OD_{600} samples and resuspending the cell pellets in 0.1 M Tris-HCl pH 8.5 containing 5 M urea. These samples were then analyzed by 12% SDS-PAGE. Gels were stained using Coomassie blue, destained and scanned.

2.4 Protein purification

RluA and TruB proteins were purified using the same procedure. Frozen cell pellets were resuspended in Buffer A (20 mM Tris-HCl pH 8.1, 400 mM KCl, 5% (v/v) glycerol, 1 mM β-mercaptoethanol, 0.5 mM phenylmethanesulfonylfluoride (PMSF), 30 mM imidazole) at 5 mL/g of cells and thawed on ice while stirring. Lysozyme was added to the homogenous cell suspension to a final concentration of 1 mg/mL followed by incubation on ice for another 30 minutes. Next, sodium deoxycholate was added to the cell suspension at 12.5 mg/g cells. The cell suspension was incubated on ice for an additional 15 minutes. Cells were kept on ice and opened by sonication using 1 minute intervals of intensity level 6 and duty cycle at 60% for 10 minutes with a ½ inch probe. The suspension was then centrifuged at 30 000×g for 45 minutes in a JA-20 rotor (Beckman). The clear lysate was loaded onto a 5 mL Ni^{2+}-sepharose column (GE Healthcare) with a flow rate of 0.5 mL/min and washed with Buffer A for approximately 60 min at a flow of 1 mL/min until the absorbance at 280 nm (A_{280}) returned to baseline (BioLogic LP Chromatography system). Proteins were eluted using a linear gradient (50 mL at 1 mL/min) to 100% Buffer B (20 mM Tris-HCl pH 8.1, 400 mM KCl, 5% (v/v) glycerol, 1 mM β-mercaptoethanol, 500 mM imidazole). Peak fractions were analyzed by 12% SDS-PAGE, pooled and concentrated using ultrafiltration (Vivaspin MWCO 10 000). Additional purification and buffer exchange of each protein was completed using size exclusion chromatography. The concentrated (1.5 – 5 mL) Ni-sepharose purified fractions were injected onto a Superdex 75 column (XK26/100 column, GE Healthcare)

in Buffer C (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 1mM β-mercaptoethanol, 0.5 mM EDTA, 5 mM $MgCl_2$, 20% (v/v) glycerol) at a flow of 1 mL/min (BioLogic DuoFlow chromatography system). Peak fractions were analyzed and concentrated as before, aliquoted, flash frozen and stored at -80°C. The concentration of purified protein was determined photometrically at 280 nm using a molar extinction coefficient of 29 910 M^{-1} cm^{-1} for RluA variants and 20 860 M^{-1} cm^{-1} for TruB variants (calculated using ProtParam [Gill and von Hippel, 1989]). ImageJ (version 1.41o, NIH) analysis of SDS-PAGE samples confirmed A_{280} concentration results.

2.5 Fluorescent Labelling of TruBC58AC174AC193AT259C

Purified TruBC58AC174AC193AT259C was diluted to a concentration of 3 µM using 1× labelling buffer (25 mM Tris-HCl pH 8.1, 7 mM $MgCl_2$, 30 mM KCl). Five-fold excess of dye, 5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid (1,5–IEADANS) was added to the protein, followed by incubation at room temperature with rocking for one hour. Unbound dye was removed by overnight dialysis at 4°C in 1000-fold excess of 1× labelling buffer. The protein was concentrated using ultrafiltration (Vivaspin MWCO 10 000). Labelled protein was aliquoted, flash frozen and stored at -80°C. The concentration of labelled protein was estimated by absorbance readings at 280 nm. The final fluorescent label concentration was determined photometrically at 490 nm using an extinction coefficient of 5 700 M^{-1} cm^{-1} (Life Technologies, 2010).

2.6 tRNA Preparation

Template DNA of *E. coli* $tRNA^{Phe}$ was first generated from the plasmid pCFO (Sampson et al., 1989) through PCR amplification using the following primers:

T7 promoter sense 5’ – GCTGCAGTAATACGACTCACTATAG – 3’

EcotRNA^{Phe} antisense 5’ – mUmGGTGCCCGGACTCG – 3’

All PCR reactions were carried out using 1x Pfu buffer with $MgSO_4$, 0.4 mM dNTPs, 0.4 μM T7 promoter sense primer, 0.4 μM EcotRNA^{Phe} antisense reverse primer, 0.12 U/μL Pfu DNA polymerase, and 0.5 μg pCFO template DNA. The PCR conditions are outlined in Table 2.3.

Table 2.3: PCR amplification of tRNA^{Phe} gene from the pCFO DNA template

Step	Temperature (°C)	Time	Cycles
Initial denaturation	95	5 minutes	1 cycle
Denaturation	95	30 seconds	
Annealing	45 (increase 1°C each repeat)	30 seconds	
	72	20 seconds	6 cycles
Extension	95	30 seconds	
	50	30 seconds	
	72	20 seconds	29 cycles
Final extension	72	11 minutes	1 cycle

The template for the 17-mer T-arm of *E. coli* tRNA^{Phe} (5' – CUU GGU UCG AUU CCG AG – 3') was generated through annealing of the following primers: TSL sense 5' – GCGAATACGACTCACTATAGGGCTTGGTTCGATTCCGAG – 3'

TSL antisense 5' - mCmUCGGAATCGAACCAAGCCCTATAGTGAGTCGTATTCGC – 3'

For annealing, the primers (final concentration of 4 μM) were subjected to an initial denaturation step at 95°C for 2 minutes in 5x transcription buffer (200 mM Tris-HCl pH 7.5, 75 mM $MgCl_2$, 50 mM NaCl, and 10 mM spermidine), followed by annealing steps starting at 90°C and decreasing the temperature 0.1°C/second for 11 cycles, with a final incubation step at 37°C.

RNA corresponding to the anticodon stem loop of *E. coli* $tRNA^{Phe}$ with a deoxyribo-2-aminopurine (2AP) substitution (5' – GGG GAU U(2AP)A AAA UCC CC – 3') was purchased from Integrated DNA Technologies (IDT).

In vitro transcriptions were carried out using (PCR-generated) DNA template (10% v/v) in 5x transcription buffer (see above), 10 mM DTT, 3 mM NTPs (ATP, CTP, UTP, and GTP), 5 mM GMP, 0.01 U/uL iPPase, 0.3 µM T7-RNA-Polymerase, and 0.12 U/µL RNase inhibitor. The reaction mixture was incubated at 37°C for 16 hours. Tritium-labelled $tRNA^{Phe}$ was generated by the addition of 0.1 mM [C5-^{3}H] UTP (0.46 Ci/mmole), instead of 3 mM non-radioactive UTP, to the reaction mix. All radioactive *in vitro* transcriptions were carried out for 4 hours. Template DNA was digested by the addition of 0.002 U/µL DNase and the reaction mixture was incubated at 37°C for an additional 1 hour. The radiolabelled tRNA was purified using Nucleobond AX20 columns (Macherey-Nagel). Prior to purification the column was equilibrated with buffer R0 (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, pH 6.3). The *in vitro* transcription mixture was first diluted to a final KCl concentration of 0.2 M with buffer R0 and buffer R3 (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, 1150 mM KCl, pH 6.3) and was then loaded onto the column. The column was washed with buffer R1 (100 mM Tris/acetate, 10 mM $MgCl_2$, 15% ethanol, 300 mM KCl, pH 6.3), and the RNA was eluted with buffer R3. Nonradioactive tRNA was purified using a 5mL Bio-Scale Mini DEAE anion exchange column (Easton et al., 2010). The column was first washed with 60 mL 2 M NaOH prior to purification. The tRNA was eluted from the column using a gradient from 100% Buffer D (50 mM sodium phosphate pH 6.5, 150 mM sodium chloride, 0.2 mM EDTA) to 100% Buffer E (50 mM sodium phosphate pH 6.5, 2 M sodium chloride, 0.2 mM EDTA) as described in Easton et al. (2010). Peak fractions were analyzed by 15% urea-PAGE, pooled and ethanol precipitated. The tRNA concentration was determined photometrically by measuring the absorbance at 260 nm using the extinction coefficient

of 5 × 10^5 M^{-1} cm^{-1} (Peterson and Uhlenbeck, 1992). The specific activity of the radiolabelled tRNA was determined through scintillation counting (Tri-Carb 2800TR).

2.7 [^{32}P] labelling of RNA

400 pmol 2AP-anticodon stem loop or full-length tRNA was unfolded at 90°C for 2 minutes and flash cooled on ice. To dephosphorylate the 5' terminus, the RNA was incubated with 1× NEB3 buffer and 0.1 U/μL calf intestinal alkaline phosphatase (CIP; New England Biolabs). The reaction mixture was incubated at 50°C for 1 hour. A phenol:chloroform extraction was completed to remove the CIP, and the RNA was precipitated by the addition of 1 volume sodium-acetate and 2.5 volume ethanol. The RNA was resuspended in 26 μL water. The RNA was rephosphorylated by incubating with 1× reaction buffer A for T4 Polynucleotide Kinase, 15 U Polynucleotide Kinase (Fermentas) and 30 μCi [γ-^{32}P] ATP. The reaction mixture was incubated at 37°C for 30 minutes. The reaction was stopped by the addition of 0.5 M EDTA, pH 8.0 to a final concentration of 0.025 M and heat denaturation at 75°C for 10 minutes. The RNA was separated from unincorporated ATP by gel filtration on Sephadex G-25. A SigmaPrep spin column was prepared by adding 200 μL of Sephadex G-25, centrifuging for 3 minutes at 10 000 rpm and removing the supernatant. Prior to sample loading, the column was washed with water. The sample was then spread evenly over the sephadex gel, and the spin column was centrifuged for 3 min at 10,000 rpm to elute the RNA. The activity of the labeled RNA was determined by liquid scintillation counting.

2.8 Preparation of fluorescein-labelled tRNA

Purified $tRNA^{Phe}$ in water was oxidized with 2 mM potassium periodate for 30 minutes. The reaction was stopped by the addition of ethylene glycol to a final concentration of 10 mM. Following an ethanol precipitation, the tRNA was incubated with 10 mM fluorescein-

5-thiosemicarbazide in a 0.1 M sodium acetate solution for 16 hours in the dark. The RNA was precipitated by the addition of 3 M sodium acetate and ethanol. The unbound dye was removed via phenol extraction. The tRNA was concentrated by ethanol precipitation and was resuspended in water. RNA concentration was determined photometrically by measuring the absorbance at 260 nm using the extinction coefficient of 5×10^5 M^{-1} cm^{-1} (Peterson and Uhlenbeck, 1992). The dye concentration was determined by measuring the absorbance at 492 nm with the extinction coefficient of 85 000 $M^{-1}cm^{-1}$ (Life Technologies, 2010). The final labelling efficiency was estimated by comparing the concentration of tRNA and the dye.

2.9 Nitrocellulose filtration

$tRNA^{Phe}$ was allowed to fold by incubating the RNA at 65°C, followed by slow cooling at room temperature. A low constant concentration of $tRNA^{Phe}$ (10 nM) was incubated with increasing concentrations of enzyme (0 – 30 µM) for 10 minutes in $TAKEM_4$ at room temperature to allow for binding. The reaction mixture was filtered under vacuum through a nitrocellulose membrane (0.2 µm, Whatman, Maidstone, United Kingdom). The membrane was then washed immediately with 1 mL ice cold $TAKEM_4$ buffer and dissolved in 10 mL scintillation cocktail for 30 minutes. The level of tRNA binding was determined through scintillation counting. The dissociation constant (K_D) for tRNA binding was calculated by plotting the fraction of bound RNA against protein concentration and fitting the data to a hyperbolic function:

$$Bound = Bound_{max} \times [protein] / (K_D + [protein])$$

2.10 Tritium release assay

Previously folded [^{3}H]tRNA was incubated with enzyme at 37°C in $TAKEM_4$ buffer. Aliquots were removed at desired time points and quenched in 1 mL of 5% (v/v)

activated charcoal (Norit A) and 0.1 M HCl. Samples were centrifuged for 2 minutes at 10 000 × g, and 0.8 mL of the supernatant was added to 0.5 mL 5% Norit solution in 0.1 M HCl. Samples were centrifuged again, and 1 mL of the supernatant was filtered through glass wool. Finally 0.8 mL of the filtrate in 4 mL scintillation cocktail was used for scintillation counting to determine the amount of free tritium in solution corresponding to the amount of pseudouridine formed. Initial velocities were determined by completing a linear regression for the linear phase of pseudouridine formation.

2.11 Quench-flow measurements

A KinTek quench-flow apparatus was used to measure pre-steady-state kinetics, where 1 µM (final concentration) [^{3}H]-tRNA^{Phe} was rapidly mixed with RluA (final concentration 2.5 – 15 µM) in TAKEM$_4$ buffer at 37°C. The reaction was quenched by the addition of 0.1 M HCl after the desired reaction time (0.003 – 60 s). The amount of total [^{3}H]-tRNA^{Phe} in the quenched sample was determined by scintillation counting. The amount of free tritium was quantified by subjecting a defined volume (120 – 220 µL) of the quenched sample to the tritium release assay as described above. The percentage of pseudouridine formation (P) was determined as a fraction of the total radioactivity present for each time point. The resulting time courses were fit to a one-exponential function to determine the apparent rate, k_{app}:

$$P = P_\infty + A \times \exp(-k_{app} \times t)$$

Where P_∞ is the endlevel, A is the amplitude of the percentage change, t is time in seconds, and k_{app} is the apparent rate.

These apparent rates were then plotted against enzyme concentration and fitted to a hyperbolic function to determine the k_{max} for pseudouridine formation:

$k_{app} = k_{max} \times [RluA] / (K_{half} + [RluA])$ (Fersht, 1998)

2.12 Fluorescence Spectrometry and Stopped-flow measurements

A fluorescence spectrometer (Quanta Master, Photon Technology International) was used to examine equilibrium binding between fluorescently-labelled TruB and fluorescently-labelled $tRNA^{Phe}$ (FL-tRNA, see above). A sample of 3 µM TruB-1,5-IAEDANS in $TAKEM_4$ was excited alone and in the presence of fluorescein-tRNA at 336 nm, and the emission spectra from 510 – 650 nm was recorded. Additionally, a solution of 3 µM FL-tRNA was analyzed alone by exciting the fluorophore at both 336 and 491 nm and recording the emission spectra. All measurements were completed at 20°C with 1 nm slit width and 1 nm step size.

Pre-steady-state fluorescence stopped-flow measurements were completed using a KinTek SF-2004 stopped-flow apparatus. Here, 1 µM (final concentration) fluorescein-labelled tRNA was rapidly mixed with 2.5 – 15 µM TruB (final concentration) at 20°C to ensure single-turnover conditions. Control experiments were conducted where 1 µM (final concentration) fluorescein-labelled tRNA was rapidly mixed with $TAKEM_4$ buffer or itself. Fluorescein was excited at 480 nm and the emission was monitored using a LG-500 nm cutoff filter. Short 15 second time courses were analyzed by fitting to a one-exponential function:

$F = F_\infty + A \times \exp(-k_{app} \times t)$

Where F_∞ is the fluorescence endlevel, A is the amplitude of the fluorescence change, t is time in seconds, and k_{app} is the apparent rate.

Time courses longer than 30 seconds were fit to a one-exponential function followed by a linear photobleaching phase (with a slope of lin):

$F = F_\infty + A \times \exp(-k_{app} \times t) + (lin \times t)$

The apparent rates were plotted against enzyme concentration and fitted with a hyperbolic function:

$k_{app} = k_1 + k_{-1}\, K_s/([E] + K_s)$

Where K_s is the apparent equilibrium binding constant for tRNA interacting with TruB, [E] is enzyme concentration, and k_1 and k_{-1} are rate constants describing the conformational change in tRNA preceeding binding to TruB (see discussion).

2.13 [^{14}C]-Uracil Exchange

200 pmol non-radioactive tRNA (final concentration of 600 nM) was folded in 1× $TAKEM_4$ as described above prior to the start of the experiment. [^{14}C]-uracil (57 mCi/mmole) was added to the tRNA to a final concentration of 500 µM. TruB wild-type or the catalytically inactive variant TruBD48N were added to the reaction to a final concentration of 10 nM. The reaction mixture was incubated at 37°C for 30 minutes. A phenol-chloroform extraction was used to remove the proteins, and the tRNA was ethanol precipitated overnight at - 20°C. The tRNA was resuspended in water and purified using a urea-PAGE. Gel extraction of the RNA was completed by excising the RNA band from the gel under UV light and incubating overnight in gel extraction buffer (0.5 M Ammonium acetate, 10 mM Magnesium acetate, 1 mM EDTA, 0.1% SDS). The sample was centrifuged 14000 rpm for 5 minutes at 4°C. The supernatant was transferred to a new tube, and the RNA was ethanol precipitated and finally resuspended in water. Samples were analyzed via scintillation counting, absorbance measurements at 260 nm as described above, and urea-PAGE.

Chapter 3 - Results

3.1 Catalysis

The pre-steady-state kinetic data collected on RluA was completed by myself as part of a larger study published in RNA (Wright et al., 2011). The undergraduate students Jaden Wright and Selina Dobing as well as Dr. Ute Kothe completed all other experiments described in the publication.

3.1.1 Uniform Slow Catalysis

RluA containing an N-terminal histidine-tag was overexpressed from the pCA24N(GFP minus)-RluA plasmid in ASKA ME3505 *E. coli* cells (NIG, Japan; Kitagawa et al., 2005). Nickel-sepharose affinity chromatography was used to purify RluA from the majority of other contaminating cellular components (Figure 3.1 A). Two peaks were observed. The largest and first peak corresponds to a large number of cellular proteins from the cell lysate. The second, smaller peak is consistent with the elution of the his-tagged RluA protein. Size-exclusion chromatography further purified, as well as rebuffered, RluA (Figure 3.1 B). Here, a single peak eluting at 242 - 292 mL corresponds to RluA. Peak fractions matching the elution profile of RluA were analysed by SDS-PAGE (Figure 3.1 C). Bands corresponding to RluA (MW of ~ 26 kDa) were observed with no other bands visible. This indicates RluA was successfully purified to a purity of more than 95%. The concentration of RluA was determined photometrically at 280 nm. Additionally, ImageJ analysis of SDS-PAGE was used to confirm RluA concentration through comparison to a previously characterized sample of TruB.

One approach to studying an enzyme's mechanism is to determine the rate of the reaction and how it changes in response to changes in experimental parameters. At high

substrate concentrations, the enzyme typically becomes saturated (as all enzyme is now in the enzyme-substrate complex) and is limited by unimolecular steps of the reaction, which could correspond to conformational changes, chemical conversions, and/or product release. All of these steps contribute to the maximum velocity (v_{max}) of the reaction. The turnover number (k_{cat}) can be calculated by dividing v_{max} by the enzyme concentration (Voet and Voet, 2011; Cleland, 2009). These steady-state methods

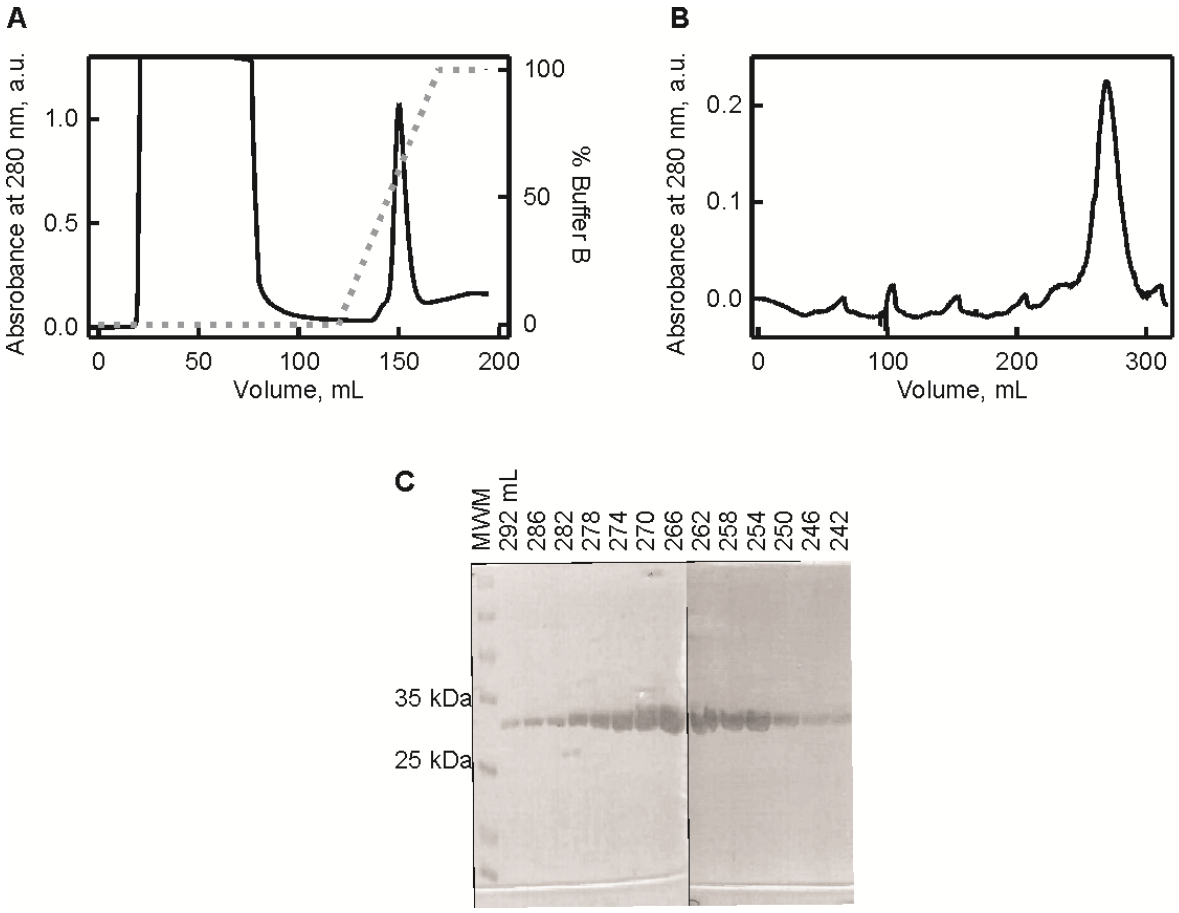


Figure 3.1: Purification of wild type RluA by Ni-Sepharose and size exclusion chromatography. A. Chromatogram of the Ni-sepharose affinity chromatography of 6× his-tagged RluA. Absorbance was measured at 280 nm (left y-axis), and a linear gradient of 30 to 500 mM imidazole (Buffer B, right y-axis, dashed line) was used to elute RluA. B. Chromatogram of the size exclusion chromatography of RluA using a Superdex 75 column (XK26/100) to remove any impurities. C. Coomassie-stained 12% SDS-PAGE of peak fractions containing RluA from size-exclusion chromatography. The elution volume of the analyzed fractions is indicated on top of the SDS-PAGE. MWM represents the protein molecular weight marker.

provide little direct information on the different steps of the enzyme's mechanism. k_{cat} is generally assumed to represent the chemistry step, but often represents a rate-limiting product release step or some conformational change within the enzyme (or substrate).

When studying steady-state kinetics the enzyme-substrate complex concentration is assumed to remain constant over the reaction progress. However, in the first few milliseconds of a reaction, the concentration of all reaction components are changing and this phase is termed the pre-steady-state. By studying pre-steady-state kinetics, the individual steps during the enzyme-substrate interaction and conversion can be dissected and measured directly. When excess enzyme is incubated with substrate, only one round of catalysis can occur; therefore, these conditions are called single turnover conditions. In a pseudo-first order reaction, at least 3-fold excess enzyme over substrate allows for the assumption that the enzyme concentration does not change during the reaction thus greatly simplifying kinetic analysis. Rapid-mixing devices, such as stopped-flow and quench-flow, allow for the analysis of these reactions on such a rapid timescale. Thereby, the rate-limiting steps can also be identified in order to establish the kinetic mechanism of an enzyme (Johnson, 2005).

In order to analyse the pseudouridylation kinetics of RluA, pre-steady-state experiments were completed. Here, the tritium release assay was adapted for use in quench-flow analysis and was completed under single-turnover conditions. The tritium release assay monitors the release of the radioactive tritium label from the C_5 position as the new C_1-C_5 glycosidic bond is formed in pseudouridine. Under these conditions, a single round of catalysis can occur where the tritium release assay monitors the formation of the enzyme-product complex as soon as it is formed. Given that the active site of RluA is

accessible to water, it is assumed that the released tritium can escape the active site cleft prior to product release (Hoang et al., 2006).

In the quench-flow apparatus, 3- to 15-fold excess protein was rapidly mixed with [^{3}H]-tRNA^{Phe} at increasing RluA concentrations for a specific length of time (0.1 to 90 seconds). The reaction was then quenched with 0.1 M HCl to denature the protein. Pseudouridine formation was observed to reach 100% within 10 seconds for all reactions, and each RluA concentration resulted in a time course that could be fit to a one-exponential function in order to determine the apparent rate (k_{app}) of pseudouridylation (Figure 3.2 A). The apparent rate of pseudouridylation, k_{app}, was then plotted against the protein concentration (Figure 3.2 B). Unlike for TruB or TruA (Wright et al., 2011), a dependence of the k_{app} for pseudouridine formation on the RluA concentration was observed. At low concentrations of RluA, the rate of pseudouridylation was approximately 0.2 s^{-1}; however this rate increased hyperbolically with RluA concentration reaching 0.49 ± 0.06 s^{-1} at a RluA concentration of 15 μM. This suggests that RluA is limited by tRNA binding under low-enzyme conditions. By fitting these k_{app} values to a hyperbolic equation ($k_{app} = k_{max} \times [RluA] / (K_{Half} + [RluA])$), a catalytic rate constant ($k_{max} = k_\psi$) could be extrapolated from the curve and was calculated to be 0.70 ± 0.15 s^{-1}. The K_{Half} was found to be 4.2 ± 2.5 μM. The previously determined catalytic constant, k_{cat}, measured under multiple turnover conditions, was reported as ~0.1 s^{-1} (Ramamurthy et al., 1999), which is significantly lower than k_ψ, indicating that product release may be limiting for RluA. Under multiple turnover conditions, each enzyme must release the product prior to catalyzing another reaction. If the rate of product release is slow, this can decrease the overall k_{cat} of the reaction. In the quench-flow experiments, only a single round of catalysis is occuring which can therefore not be limited by product release. Notably, the pseudouridylation rate of RluA is very similar to the k_ψ values

obtained for TruB and TruA, 0.5 ± 0.2 and 0.35 ± 0.2 s^{-1}, respectively (Wright et al., 2011).

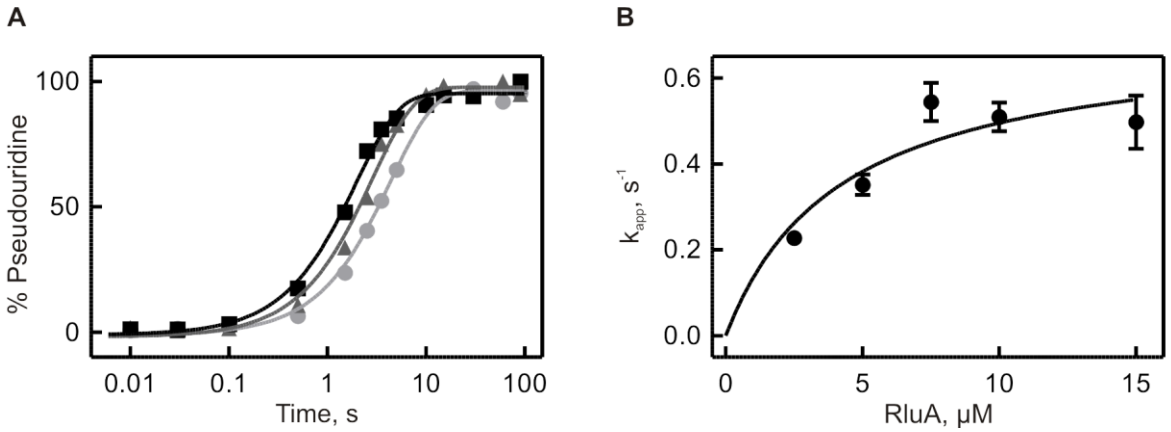


Figure 3.2: Rapid kinetic quench-flow analysis of pseudouridine formation by RluA. A. Time courses of pseudouridine formation by RluA under single turnover, pre-steady-state conditions. Increasing concentrations of RluA (circles: 2.5 μM, triangles: 5.0 μM, squares: 10 μM) were rapidly mixed with 1 μM [^{3}H]-tRNA^{Phe} in a quench-flow apparatus. A modified tritium release assay was used to determine the percentage of pseudouridine formed at each time point. Fitting with a one-exponential function (smooth lines) allowed for the determination of the apparent rate (k_{app}) of pseudouridine formation for each RluA concentration. B. Dependence of the apparent rate, k_{app}, of pseudouridine formation under single turnover conditions on RluA concentration. Data were fit to a hyperbolic function (smooth line) with a maximum rate of 0.7 ± 0.15 s^{-1}.

3.1.2 Investigating the Presence (or Absence) of a Free Uracil Base During Catalysis

During pseudouridine formation, the N – C glycosidic bond is broken, the base rotated and reattached to the ribose sugar, all within the catalytic pocket of the pseudouridine synthase. Three catalytic mechanisms have been suggested (Figure 1.4). The catalytic aspartate residue has been proposed to form either a Michael adduct through the attack at C6 of the uracil base or an acylal intermediate by attacking C1' of the ribose sugar. The third mechanism proposed by Miracco and Mueller (2011) implies deprotonation by the catalytic aspartate on the ribose C2' position to form a glycal intermediate in either a

step-wise or concerted process (Figure 1.4). Some studies have hypothesized that a covalent bond may form between the uracil and the aspartate residue pointing towards the Michael adduct pathway (Hoang et al., 1998; Gu et al., 1999). During catalysis it can be hypothesized that the uracil base, once deattached from the ribose, may escape from the active site cleft if no covalent bond is formed to the uracil as proposed in the glycal and acylal mechanism. To investigate whether the uracil base could be exchanged during the reaction, we first needed to examine whether free uracil may bind to TruB thereby preventing tRNA substrate binding. A nitrocellulose filtration assay was designed where 10 nM [^{3}H]-tRNA^{Phe} was incubated with increasing concentrations of uracil (0 – 1 mM). TruBD48N was added to the reaction at a concentration above the dissociation constant (5 µM, K_D of 1.4 µM; Wright et al., 2011) to ensure binding, and the reaction mixture was incubated at room temperature for 10 minutes. Proteins will bind to the nitrocellulose membrane and if RNA forms a complex with the protein, it too will be retained on the nitrocellulose membrane. The protein-substrate mixture is filtered under vacuum and the membrane washed with cold TAKEM$_4$ buffer. This washing step ensures that any unspecific interactions are broken and any unbound radioactive substrate is removed. Following filtration and scintillation counting, no significant changes were observed in the level of radioactivity retained on the filter. This constant level of radioactivity indicates that RNA binding is not inhibited by free uracil in solution and that TruB can still bind to its substrate tRNA (Figure 3.3 A).

Next, the exchange of the uracil base within the catalytic pocket with a [^{14}C]-labelled uracil from the solution was examined, which could ultimately lead to the incorporation of [^{14}C]-uracil into the modified tRNA. In addition to the wild-type enzyme, control experiments using the catalytically inactive TruB variant, TruBD48N, or no enzyme were also completed. Here, it was expected that no radioactive uracil would be incorporated

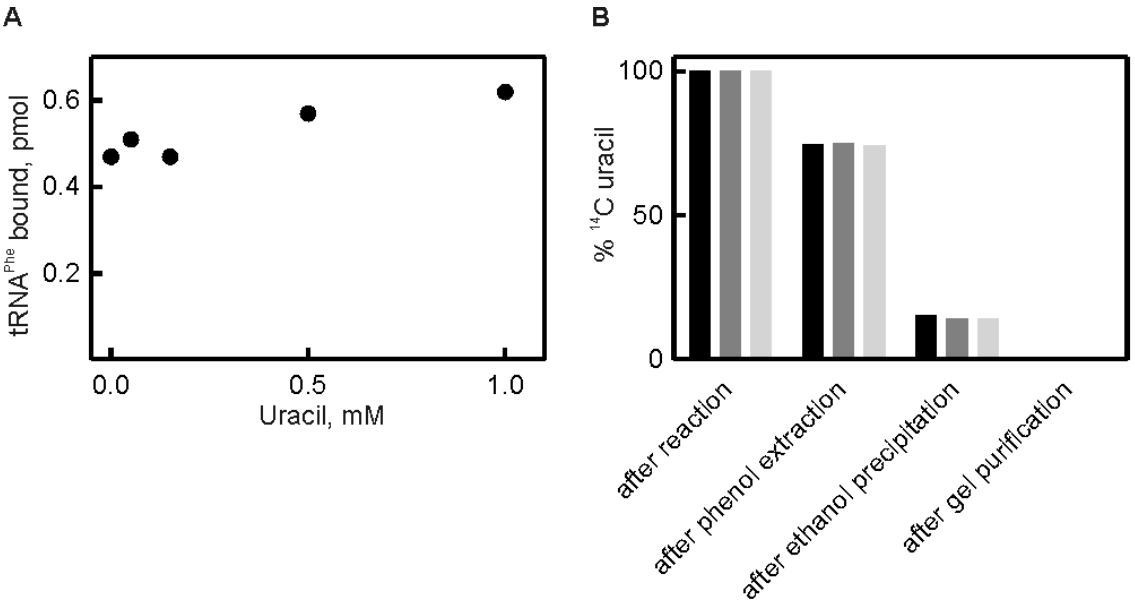


Figure 3.3: Analysis of free uracil base exchange during pseudouridine formation by TruB. A. Nitrocellulose filtration assay to examine tRNA binding by TruB in the presence of free uracil. Increasing concentrations of uracil (0 - 1 mM final concentration) were incubated with [^{3}H]-tRNA^{Phe} (10 nM final concentration) and TruBD48N (5 µM final concentration) at room temperature, filtered and the radioactivity remaining on the nitrocelulose filter was quantified. B. Assay to observe incorporation of [^{14}C]-uracil in tRNA during pseudouridylation by TruB. Unlabelled tRNA^{Phe} (1 µM final concentration) was incubated with 10 nM TruB wild-type (black), TruBD48N (dark grey), or without enzyme (light grey) in the presence of 500 µM [^{14}C]-uracil. Samples were collected following the 30 minute reaction, after phenol extraction to remove the protein, after ethanol precipitation of the RNA and after urea-PAGE purification of the RNA. The percentage of [^{14}C]-uracil remaining in the samples was determined by scintillation counting.

into the tRNA. Reactions were completed by incubating1 µM unlabelled tRNA^{Phe} with 5 µM [^{14}C]uracil and 10 nM TruB, TruBD48N or no enzyme at 37ºC for 30 min. The enzyme was removed by phenol/chloroform extraction, and the tRNA was precipitated, purified by urea-PAGE gel extraction and quantified photometrically at 260 nm. The level of [^{14}C]-uracil incorporation was determined through scintillation counting. The initial reaction contained 200 pmol tRNA and following phenol extraction, ethanol precipitation and urea-PAGE gel purification (to remove any free uracil) approximately 16 to 26% of tRNA was recovered. Notably, no [^{14}C]-uracil incorporation was detected in the purified

tRNA when the tRNA was incubated with TruB, the catalytically inactive TruBD48N or without enzyme (Figure 3.3 B).

3.2 RNA Binding

3.2.1 Interaction of modified and unmodified tRNA with Pseudouridine Synthases

A significant amount of research has gone into examining the substrate specificity requirements of different pseudouridine synthases. Here, the minimal structural and chemical differences between substrate tRNA (containing uridine) and product tRNA (containing pseudouridine), and their effects on binding to TruB and RluA, were investigated. In order to study binding interactions between RluA and tRNA, the catalytically inactive variant RluAD64N enzyme was prepared. Site-directed mutagenesis was used to create a point mutation in the pCA24N(GFP minus)-RluA plasmid, where the codon for the catalytic aspartate was replaced with an asparagine codon. Using the same conditions as for the wild-type enzyme (Fig. 3.1), RluAD64N was overexpressed, purified and quantified. The catalytically inactive TruBD48N variant was prepared previously (Wright et al., 2011). Nitrocellulose filtration assays were used to determine the dissociation constants (K_D) of tRNA binding to RluA and TruB. Following filtration of the reaction mixture, any [^{3}H]-tRNA^{Phe} bound to the enzyme remained on the nitrocellulose membrane filter. The extent of binding was then quantified through scintillation counting. The catalytically inactive variants of RluA and TruB were used to determine the enzymes' affinity for unmodified substrate, tRNA(U), while the catalytically active variants were used to determine the affinities for the pseudouridylated product, tRNA(Ψ). During incubation, the catalytically active TruB can bind and pseudouridylate the tRNA to form the product tRNA. Under the experimental conditions used, all the

tRNA will be modified by the enzyme. The RNA was melted and refolded as described in Materials and Methods prior to commencing the experiment. A low constant concentration of [^{3}H]-tRNA^{Phe} (10 nM) was incubated for 10 min at room temperature with increasing concentrations of enzyme (0 – 30 µM).

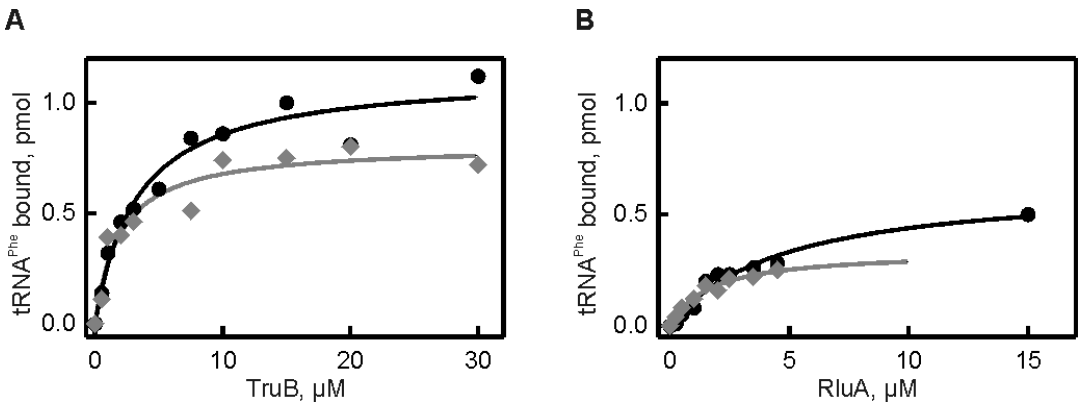


Figure 3.4: Measuring TruB and RluA affinities for uridine or pseudouridine containing tRNA. Nitrocellulose filtration assays were conducted by incubating 10 nM [^{3}H]-tRNA^{Phe} with 0 - 30 µM (final concentrations) of (A) TruB wild-type (grey diamonds) or TruBD48N (black circles) (B) RluA wild-type (grey diamonds) or RluAD64N (black circles). The amount of tRNA bound was determined through scintillation counting of the nitrocellulose filters. The dissociation constant, K_D, was determined by fitting the data to a hyperbolic function (smooth lines). Wild-type TruB and RluA were used to determine the affinity for product tRNA while the inactive variants were used to determine the affinity for substrate tRNA.

The percentage of RNA bound was plotted against the enzyme concentration and fitted to a hyperbolic function to determine K_D values (Figure 3.4). Increasing amounts of [^{3}H]-tRNA^{Phe} were retained on the filter as the protein concentration increased. The extent of binding reached a maximum of approximately 1.2 pmol for TruBD48N (substrate tRNA) and approximately 0.8 pmol for TruB wild type (product tRNA). Hyperbolic fitting of the binding curves revealed K_D values of 3.0 ± 0.7 and 2.0 ± 0.6 µM for substrate and product tRNA binding to TruB, respectively. The preparation of RluAD64N had a concentration of only 9 µM which limited the range of possible RluAD64N concentrations

that could be tested in the nitrocellulose filtration assay. The product tRNA(Ψ) bound to RluA wild-type to a higher extent than the substrate tRNA(U) bound to RluAD64N, approximately 0.5 pmol and 0.25 pmol, respectively. However, due to the low RluAD64N concentration, a greater extent of binding might be expected at concentrations higher than 4.5 µM. RluA had similar dissociation constants as TruB, where substrate tRNA had a K_D of 1.7 ± 0.4 µM and product tRNA had a K_D of 4.5 ± 0.9 µM for binding to RluA.

3.2.2 Interaction of Truncated tRNAs with Pseudouridine Synthases

According to previously published results, TruB is able to form a pseudouridine in a truncated tRNA substrate consisting of only the T-arm (Gu et al., 1998). Although biochemical studies have been completed using this truncated substrate with TruB, no dissociation constant has been reported. In order to verify this observation, the T-arm of *E. coli* $tRNA^{Phe}$ was *in vitro* transcribed using [5-^{3}H]-UTP to create a 17-mer radiolabelled substrate. Nitrocellulose filtration assays were completed to determine the extent of binding between TruB and the [^{3}H]-T-arm substrate. Prior to starting the experiment, the RNA was folded as described previously. Increasing concentrations of TruBD48N (0 – 30 µM) were incubated with the stem loop for 10 minutes at room temperature before filtration. The percentage of RNA bound was plotted against protein concentration and fitted to a hyperbolic function to determine K_D values (Figure 3.5 A). Under these conditions, the full-length tRNA substrate achieved an endlevel of approximately 1.2 pmol and had a K_D of 3.0 ± 0.7 µM. The truncated T-arm substrate, however, had a much lower level of binding (~ 0.3 pmol) and had a K_D of 2.2 ± 0.8 µM. This indicates that although TruB's affinity for the truncated substrate remains relatively unchanged, the amount of substrate that can effectively bind is much lower than for the full-length substrate under these experimental conditions.

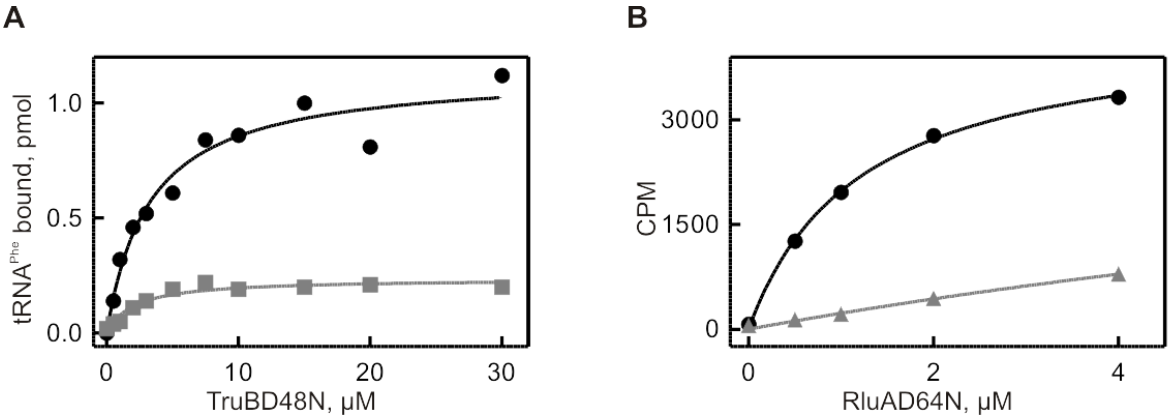


Figure 3.5: TruB and RluA interactions with truncated substrate tRNA. Nitrocellulose filtration assays of (A) TruBD48N (0 - 30 μM final concentrations) incubated with 10 nM full-length [^{3}H]-tRNA^{Phe} (black circles) or [^{3}H]-T-arm (grey squares) and (B) RluAD64N (0 - 4 μM final concentrations) incubated with 10 nM [^{32}P]-tRNA^{Phe} (black circles) or [^{32}P]-anticodon stem loop containing deoxyribose-2-aminopurine (grey triangles). The percentage of tRNA bound was determined through scintillation counting. The data were fit to a hyperbolic function to determine the dissociation constant, K_D.

RluA has also been shown to interact with truncated substrates comprised of the anticodon stem loop (ASL) of tRNA^{Phe} (Hamilton et al. 2006, Hoang et al. 2006). No dissociation constant for RluA binding to its ASL substrate has been described before. For fluorescence experiments, we had obtained a 2-aminopurine modified anticodon stem loop (IDT). The original base two positions downstream of the target uridine (U32) was replaced with a 2-aminopurine modification. Additionally, this RNA also contained a deoxyribose sugar as part of the 2-aminopurine modification. To measure binding to RluA, both full-length tRNA^{Phe} and the anticodon stem loop were radiolabelled with [^{32}P]. Again, due to protein concentration limitations, only 0 – 4.5 μM RluAD64N were incubated with 10 nM of [^{32}P]-labelled substrate. The reaction mixture was filtered and washed as described above. Following scintillation counting, the percentage of RNA bound was plotted against protein concentration and fitted to a hyperbolic function (Figure 3.5 B). Similar to TruB, RluA bound to the full-length substrate to a much higher

extent than the stem loop. The full-length tRNA substrate had a K_D of 1.2 ± 0.1 µM, while the truncated ASL had a dissociation constant of 17 ± 13 µM. These differences may be due to the presence of the 2-aminopurine modification and/or the deoxyribose sugar in the anticodon stem loop, missing protein-RNA interactions, or the presence of unproductive misfolded stem loop structures.

To examine the truncated T-arm as a possible substrate for pseudouridylation by TruB, Gu and coworkers (Gu et al., 1998) used a 17-mer corresponding to the T arm of wild-type yeast $tRNA^{Phe}$ in the standard tritium release assay. They were able to report K_M and k_{cat} values, 800 nM and 0.24 s^{-1} respectively, which were almost identical to those they determined for the full-length transcript, 780 nM and 0.24 s^{-1}. To examine the T-arm as a substrate for TruB and to test these previous results, a multiple turnover tritium release assay was completed using 10 nM TruB and 600 nM [^{3}H]-$tRNA^{Phe}$ or [^{3}H]-T-arm. Prior to starting the experiment, the RNA was folded as described previously. Using the full-length tRNA resulted in 100% pseudouridine formation within the first 10 minutes of the reaction. However, following 2 hours of incubation, no significant level of pseudouridine formation was observed when [^{3}H]-T-arm was included as substrate in the reaction (Figure 3.6 A). Next, the tritium release assay was repeated under single turnover conditions, where 1 µM tritium-labelled T-arm or full-length tRNA substrate was incubated with 5 or 15 µM TruB and the percentage of pseudouridine formation was determined over time (Figure 3.6 B). In contrast to the multiple turnover experiment, significant pseudouridine formation was observed after 1 minute for both enzyme concentrations, but approximately only 45% of the substrate was converted to pseudouridine after 90 minutes when TruB was incubated with the T-arm. The reaction

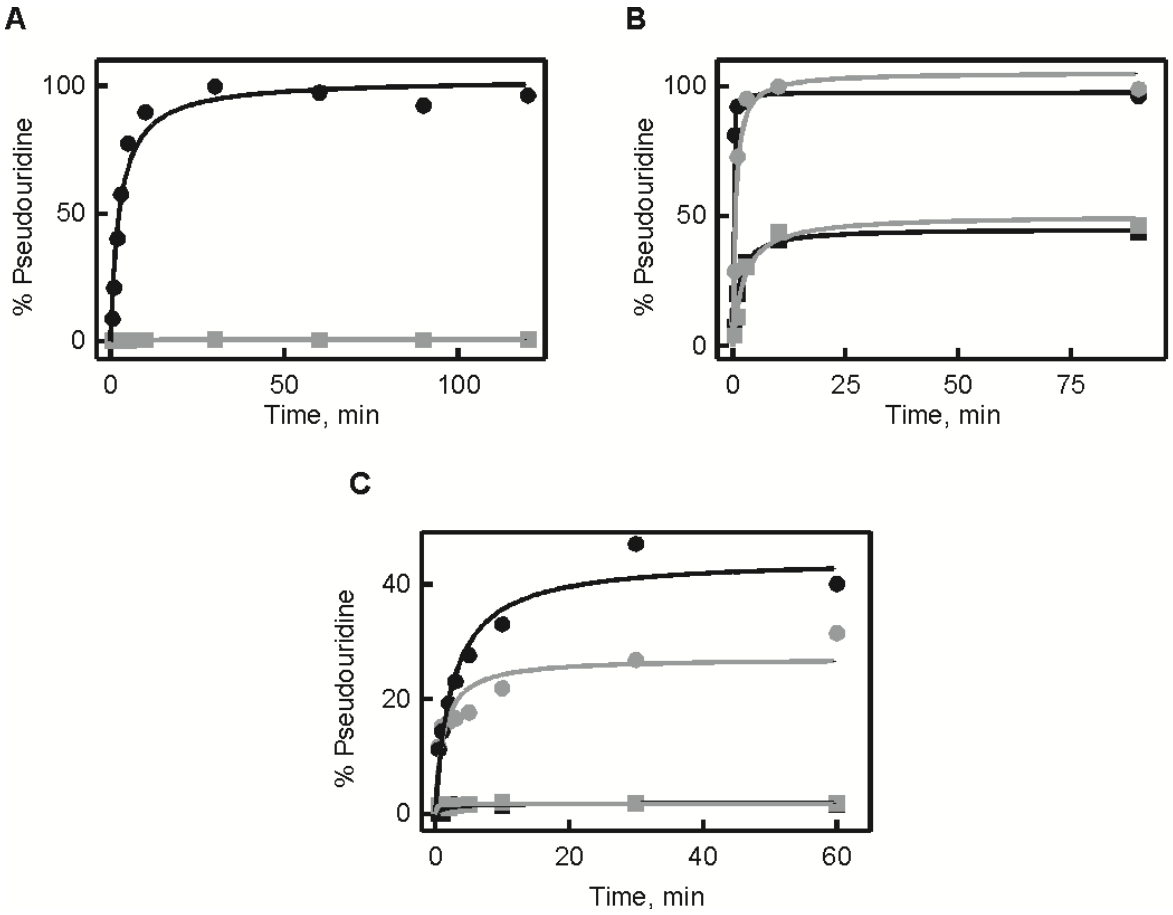


Figure 3.6: Tritium release assay utilizing [^{3}H]-T-arm as the substrate for TruB. A. Multiple turnover tritium release assay where 10 nM TruB was incubated at 37°C with 600 nM [^{3}H]-tRNA^{Phe} (black circles) or [^{3}H]-T-arm (grey squares). B. Single turnover tritium release assay where 1 µM [^{3}H]-T-arm was incubated with 5 µM (grey squares) or 15 µM (black squares) TruB. Control assays were also completed where 5 µM (grey circles) or 15 µM (black circles) TruB was incubated with 1 µM full-length [^{3}H]-tRNA^{Phe}. C. Investigating the effect of temperature on pseudouridine formation in T-arm substrate. 600 nM [^{3}H]-tRNA^{Phe} was incubated with 10 nM TruB at 15°C (black circles) and 20°C (grey circles). The truncated [^{3}H]-T-arm was also incubated with TruB at 15°C (grey squares) and 20°C (black squares). The percentage of pseudouridine formation was determined using the tritium release assay and scintillation counting.

was almost complete within the first minute of incubation when the full-length tRNA substrate was used. Notably, initial velocities of pseudouridine formation in the single turnover assays were quite different for the full-length and the truncated substrates. As expected, the full-length substrate was modified with a rate of ~5 nM•s^{-1}; with 10 nM

enzyme this corresponds to a single-turnover rate constant k_ψ of ~0.5 s^{-1} which is identical to the rate constant measured by quench-flow measurements; however, the truncated substrate had an initial rate of only ~0.6 nM•s^{-1}, almost 10-fold slower than the full-length tRNA. At the moment, we cannot distinguish whether this is due to an effect on binding or the catalytic step or a combination of both. In the nitrocellulose filtration assay approximately 10% of the T-arm substrate could bind to TruB at room temperature. Considering the T-arm has a melting temperature of approximately 18°C as determined by the oligonucleotide properties calculator (Kibbe, 2007), it was hypothesized that the RNA may unfold at 37°C during tritium release assays. Therefore multiple turnover tritium release assays were completed at 15 and 20°C using both the truncated and full-length substrates. Also, at these lower temperatures, the T-arm still did not achieve any significant pseudouridine formation; however, TruB was able to form pseudouridine using the full-length tRNA substrate, albeit to a lower extent (Figure 3.6 C).

3.3 PUA Domain

3.3.1 Analysis of RNA Binding Using Pre-Steady-State Stopped-Flow Kinetics

Wright and coworkers (2011) showed that as TruB binds an unlabeled tRNA, an increase in the abosorbance at 260 nm can be observed. From these data, they were able to determine an apparent rate of binding, k_{app}, which surprisingly lacked any concentration dependence. This suggested a two-step binding mechanism (Wright et al., 2011). To further dissect the two-step binding process, here binding interactions were investigated using the rapid-kinetic stopped-flow technique mixing a 3' fluorescein-tagged $tRNA^{Phe}$ (FL-$tRNA^{Phe}$) substrate (described in Materials and Methods) with

unlabelled TruB,. RNA and dye concentrations were determined photometrically at 260 and 492 nm, respectively. The RNA had a final concentration of 120 µM and a labelling efficiency of 70%. A constant low concentration of FL-tRNA^{Phe} (0.75 µM) was rapidly mixed with increasing amounts of wild-type (2.5 – 15 µM) or 5 µM catalytically inactive TruBD48N. Following excitation of the fluorescein group at 480 nm and passing of the emission light through a 500 nm cutoff filter, changes in the fluorescence signal were monitored over time (Figure 3.7 A). A linear photobleaching phase occurred after 15 seconds as observed in 300 s time course control experiments mixing FL-tRNA^{Phe} with buffer. No change in the fluorescent signal was observed in the first 10 s when FL-tRNA^{Phe} was mixed rapidly with buffer or itself, but when mixed with TruB or TruBD48N a significant decrease in fluorescence was detected (Figure 3.7 A). The resulting time

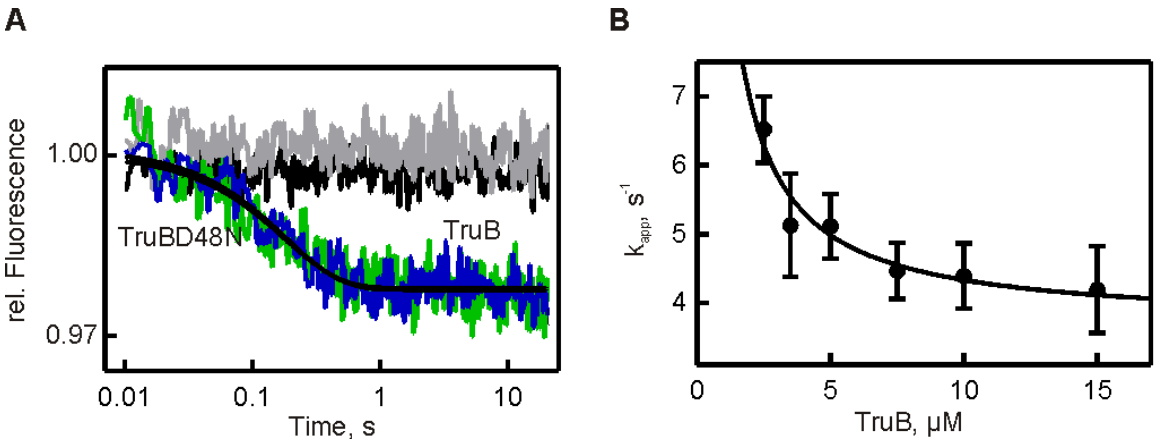


Figure 3.7: Fluorescence stopped-flow measurements to analyse TruB-tRNA interactions. A) Single-turnover, pre-steady-state conditions were used to analyse binding of fluorescein-labelled tRNA^{Phe} by TruB. 1 µM (final concentration) of fluorescent tRNA was rapidly mixed with buffer (light grey), fluorescein-tRNA^{Phe} (black), TruB wild-type (blue, 5 µM final concentration) or the catalytically inactive TruBD48N (green, 5 µM final concentration). Following excitation at 480 nm, the fluorescence signal was passed through a 500 nm cutoff filter and recorded. The time courses were fitted to a one-exponential function (smooth lines) to determine the apparent rate of binding, k_{app}. B) Dependence of the apparent rates of binding, k_{app}, of fluorescein-tRNA^{Phe} on the concentration of TruB wild-type. The data were fit with a hyperbolic function (smooth line).

courses were fit to a one-exponential function to determine an apparent rate of binding, k_{app}. As both active and catalytically inactive TruB show essentially the same time course, the change in fluorescence must be due to interaction of TruB with the substrate tRNA, but cannot be related to catalysis. More specifically, these results indicate that the 3' end of the tRNA interacts with TruB as the environment of the fluorophore changes.

When the resulting apparent rates (k_{app}) were plotted against TruB concentration, a decrease in k_{app} with increasing TruB concentration was observed (Figure 3.7 B). In a typical bimolecular binding reaction, the rate will increase linearly with increasing enzyme concentration. Instead, here the apparent rates decreased with increasing enzyme concentration, and this decrease could be fit with a hyperbolic function. These results suggest a model where binding occurs in two steps: a slow conformational change occurs within the tRNA, which is then followed by rapid binding to TruB. Before TruB can bind, the tRNA must first undergo a conformational change, resulting in the binding-competent form of tRNA. At low enzyme concentrations, all of the available binding-competent tRNA can rapidly bind to the enzyme, giving a fast k_{app}. However, at high enzyme concentrations, the enzyme must first wait for binding-competent tRNA to become available to bind, as the conformational change from binding-incompetent tRNA to binding-competent tRNA is much slower than the binding to TruB. This model correlates to a slow k_1 rate constant (binding-incompetent tRNA → binding-competent tRNA) and a fast k_2 rate constant (binding-competent tRNA + TruB → TruB•tRNA). Fitting to a hyperbolic function can provide information on the apparent binding constant (K_s) which is comparable to the K_D for substrate binding to TruB. Unfortunately, we found the K_s value could not be reliably determined as there are too few data points at low TruB concentrations. From the hyperbolic function, a k_{min} can be calculated, where k_{min} is the rate at which the tRNA conformational change (k_1) occurs. The k_{min} was found to be

3.7 ± 0.5 s^{-1} and is in good agreement with previous results where absorbance based studies showed that TruB binds to unlabelled full-length tRNA in a two step process with an average rate of 6.0 ± 1.8 s^{-1} (Wright et al., 2011).

3.3.2 Examining PUA Domain Deletion on Substrate Binding and Pseudouridylation Activity

Previous studies have indicated that the 3' ACA motif of the guide RNA in the H/ACA ribonucleoprotein particle interacts with the PUA domain of Cbf5 (Li & Ye, 2006). TruB, the *E. coli* homolog of Cbf5, also shares this PUA domain (amino acid residues 250 – 314) and therefore may interact with the 3' CCA terminus of tRNA molecules in a similar manner. To further investigate the role of TruB's PUA domain in RNA interactions, the PUA domain of TruB was removed by introducing two premature stop codons at positions 244 and 245 using quick-change site-directed mutagenesis. The deletion of the PUA domain was performed both on the background of active TruB as well as the catalytically inactive TruBD48N variant. The resulting TruBΔPUA and TruBD48NΔPUA variants were purified by Ni-Sepharose and size-exclusion chromatography as described for the wild-type protein.

First, the affinity of TruB lacking the PUA domain for substrate tRNA was determined. Towards this goal, the standard nitrocellulose filtration assay using increasing concentrations of TruBD48N or TruBD48NΔPUA (0 – 30 µM) and a low constant concentration (10 nM) of [^{3}H]-tRNA^{Phe} was used to determine the dissociation constant for the catalytically inactive ΔPUA variant protein binding to its substrate tRNA (Figure 3.8 A). Both TruB variants, full-length as well as TruB lacking the PUA domain, reached

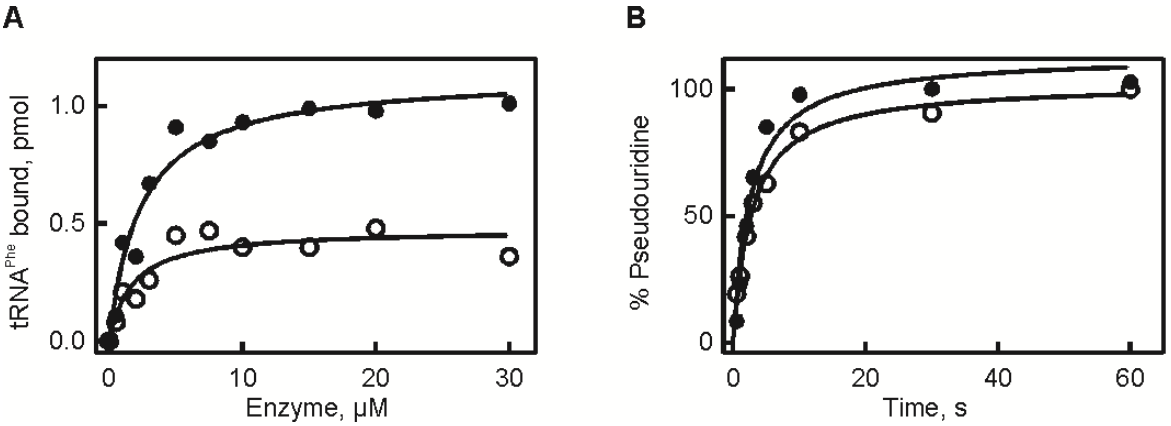


Figure 3.8: Effect of deleting TruB's PUA domain on tRNA binding and pseudouridine formation. A. Nitrocellulose filtration assay to assess the effect of PUA domain deletion on TruB`s ability to bind [^{3}H]-tRNA^{Phe}. Increasing concentrations of TruBD48N (closed circles) or TruBD48NΔPUA (open circles) were incubated with 10 nM (final concentration) of tRNA. The data were fit to a hyperbolic function (smooth line) to determine dissociation constants. B. Tritium release assays were used to measure pseudouridine formation by 10 nM TruB wild-type (closed circles) or TruBΔPUA (open circles) using 600 nM [^{3}H]-tRNA^{Phe} as substrate.

the final level of binding at about 10 µM of protein. In accordance with this observation, fitting to a hyperbolic function revealed a K_D value of 1.7 ± 0.7 µM for binding of substrate tRNA to the PUA deletion variant which is similar to that of the full-length protein (2.4 ± 0.5 µM). Interestingly, the extent of RNA binding at high protein concentrations for the truncated variant is approximately half of that for the wild-type enzyme. The exact reason for the differences in binding endlevels is not known, and these experiments will need to be repeated to confirm the results.

Second, a multiple turnover tritium release assay was used to assess the pseudouridylation activity of the TruB variant lacking the PUA domain. Here, 10 nM TruBΔPUA was incubated with 600 nM [^{3}H]-tRNA^{Phe}, and tritium release was monitored over time. If the PUA domain was involved in positioning the substrate in the catalytic cleft, very little or no pseudouridine formation would be expected. Interestingly, wild-type

endlevels of pseudouridylation were observed for TruBΔPUA (Figure 3.8 B). Additionally, when the initial velocities for both enzymes were calculated from the initial linear increase in pseudouridine formation, they were found to be nearly identical, 26 ± 2.6 and 26 ± 0.7 nM s^{-1} for TruB wild-type and TruBΔPUA, respectively. These results indicate that the PUA domain is not necessary for pseudouridine formation which is in accordance with the observation that the TruB variant lacking the PUA domain is capable of binding tRNA.

To examine whether a fluorescence decrease is still observed when the TruBΔPUA variant is rapidly mixed with 3' fluorescein-labelled tRNA, additional stopped-flow experiments were completed. Here, the same experimental conditions as above were repeated using 5 µM of the truncated TruB variants, TruBΔPUA and TruBD48NΔPUA, and 0.75 µM fluorescein-tRNA. Unlike for the wild-type enzyme, no change in the fluorescence signal was observed for the truncation variants (Figure 3.9). This indicates

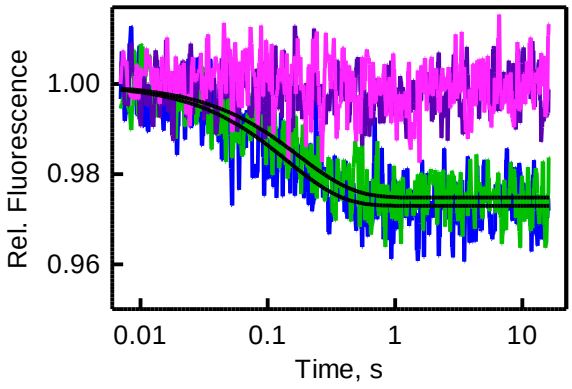


Figure 3.9: Fluorescence stopped-flow analysis of PUA deletion variants interacting with fluorescein-tRNA^{Phe}. Single turnover conditions were used to analyse substrate binding by TruB. Here, 1 µM fluorescein-tRNA^{Phe} was rapidly mixed with TruBΔPUA (pink) or TruBD48NΔPUA (purple; 5 µM final concentration). For comparison, time courses corresponding to wild-type TruB (green) and TruBD48N (blue) are also shown. Time courses were fitted with a one exponential function (black lines).

that the 3' acceptor stem of the fluorescein-tRNA was in fact interacting with the PUA domain of TruB; however based on the above results for the TruBΔPUA variant, this interaction is not essential for substrate binding or catalysis.

3.3.3 Analysing tRNA Interactions with TruB`s PUA domain via Fluorescence Spectrometry

Originally, we had planned to analyze the interaction of TruB with tRNA using fluorescence energy transfer between fluorescently labelled tRNA and fluorescently labelled TruB. As both labelled components were available, we tested here whether FRET could be observed indicating a proximity of the tRNA 3' end and the TruB PUA domain. Fluorescence resonance energy transfer (FRET) can be used to analyse the distance (r) between two dyes, donor and acceptor. These dyes can be attached to two different interacting species to examine intermolecular interactions, but can also be attached to a single molecule in order to analysis intramolecular conformational changes. Specifically, this technique is used to analyze binding interactions as changes in the fluorescence resonance energy transfer reflect changes in the distance between two dyes. Based on previous observations, it was proposed that the 3' terminus of a full-length tRNA substrate could potentially come within sufficient proximity to the PUA domain of TruB to induce FRET between a dye attached to the 3' end of the tRNA and a second dye located on the PUA domain. As described above, full-length $tRNA^{Phe}$ was labelled with a fluorescein dye at the 3' terminus (see Materials and Methods).

TruB was labelled using a thiol-specific dye, 5-[2-[(2-Iodo-1-oxoethyl)amino]ethylamino]-1-naphthalenesulfonic acid (1, 5 - IAEDANS). IAEDANS has an excitation maximum of 336 nm and an emission maximum of 490 nm. IAEDANS's emission spectra overlaps

well with the excitation spectrum of fluorescein, which has an excitation maximum of 492 nm and an emission maximum of 521 nm, giving the two dyes a Forster radius (R_0) of 46 Å (Life Technologies, 2010). First, the intrinsic cysteine residues in TruB were replaced with alanine to create the plasmid pET28a-EcoTruBC58A C174A C193A via site-directed mutagenesis (completed by Jaden Wright). These cysteine residues were replaced by alanine in Ramamurthy et al. (1999) without any effect on the pseudouridylation rates of TruB. Site-directed mutagenesis was again utilized to replace the non-conserved threonine 259 with a new cysteine residue in the PUA domain of TruB to create pET28a-EcoTruBC58AC174AC193AT259C (done by Ashley Taylor under my supervision; see Figure 1.7). This protein was overexpressed and purified using the standard conditions for TruB described previously. Next, the TruB variant was incubated with 5–fold excess of dye for 1 hour at room temperature, dialysed to remove excess dye and concentrated (completed by Nathan Dawson under my supervision). Protein and IEADANS concentrations were determined photometrically at 280 and 336 nm, respectively, resulting in TruB concentration of 79 µM and a 1, 5-IAEDANS concentration of 154 µM. The 2:1 dye:protein labelling ratio indicates that TruB could be labelled nonspecifically in other positions in addition to cysteine 259.

Using the TruB variant labelled in the PUA domain and the 3’ end-labelled tRNA, fluorescence experiments were conducted to detect the occurrence of FRET. First, $TAKEM_4$ buffer was excited at 336 nm to measure the intrinsic Raman scattering of the solution (Figure 3.10; buffer control). A very small peak (1 100 counts) was observed at 373 nm which can be seen in all other spectra. Next, a solution of 3 µM TruB-1,5-IAEDANS (donor dye) in $TAKEM_4$ buffer was excited at 336 nm, the excitation maximum for IAEDANS, and the emission spectra was recorded from 360 – 650 nm (Figure 3.10; donor only control).

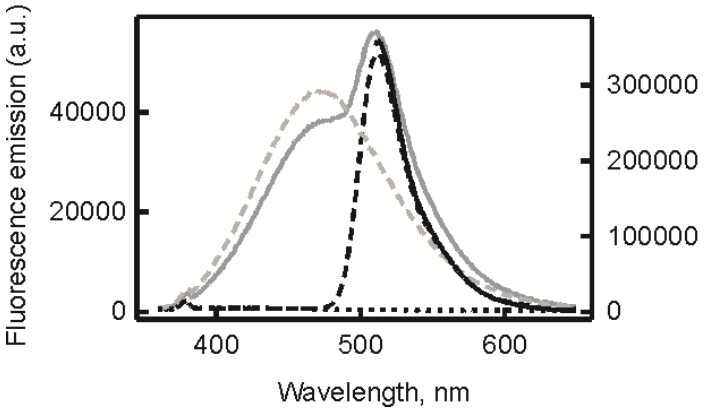


Figure 3.10: Fluorescence spectra of 1,5-IAEDANS-TruB and fluorescein-tRNA^{Phe} to detect Fluorescence Resonance Energy Transfer. Solutions of 3 µM 1,5 IAEDANS-TruB and 3 µM 3' labelled fluorescein-tRNA^{Phe} were analysed via fluorescence spectroscopy. Emission scans were completed from 360 to 650 nm by exciting 1,5-IAEDANS-TruB alone at 336 nm (grey dashes). Fluorescein-tRNA was excited alone at 336 nm (black dashes) and 491 nm (continuous black line; right y-axis). 1,5-IAEDANS-TruB and fluorescein-tRNA^{Phe} were mixed and 1,5-IAEDANS was excited at 336 nm and the emission spectra recorded from 360 to 650 nm (continuous grey line). TAKEM$_4$ buffer was excited at 336 nm, and the emission spectra was recorded in order to observe Raman scattering (black dots).

The peak emission of approximately 44 000 counts occurred at 467 nm. Separately, 3 µM of fluorescein-labelled tRNA (acceptor only control) was excited alone at 336 nm and the emission spectra recorded from 360 to 650 nm. This produced a peak with the highest emission occurring at 510 nm with counts of just over 50 000. The fluorescein-labelled tRNA was also excited at 491 nm, the excitation maximum for fluorescein, and the emission spectra recorded from 510 – 650 nm. This resulted in a large fluorescence emission peak of over 350 000 counts at 510 nm. Subsequently, TruB-1,5-IAEDANS was mixed with fluorescein-tRNA at final concentrations of 3 µM each. This concentration was chosen as it is above the dissociation constant established for TruB and full-length, pseudouridylated tRNA. By using active TruB, the enzyme will convert all substrate tRNA to pseudouridylated tRNA during the incubation resulting in the formation

of an enzyme-product complex which will be observed in the fluorescence spectrometer. Again, the donor dye (1,5-IAEDANS) was excited at 336 nm, and the fluorescence emission spectra were recorded from 360 to 650 nm (donor and acceptor). The 1,5-IAEDANS fluorescence emission decreases from 44 000 counts to 38 000 counts at 467 nm while the emission of fluorescein increases from 50 000 to 54 000 counts at 510 nm (Figure 3.10). Upon excitation of the donor dye in a FRET experiment, some of its energy is transferred to the acceptor dye resulting in acceptor fluorescence. However this energy transfer can only occur when the dyes are in close proximity to each other. Therefore a decrease in the donor dye emission and an increase in the acceptor dye emission reflects a decrease in the distance between the two dyes. This change in fluorescence emission indicates that the fluorescein at the 3' terminus of the tRNA was in close proximity to the 1,5-IAEDANS attached to C259 in the PUA domain of TruB.

Chapter 4 - Discussion

4.1 Uniform slow catalysis

Pre-steady-state quench-flow analysis was used to study the catalytic step of RluA in pseudouridine formation. These experiments revealed a dependence of the catalytic rate on RluA concentration and a k_{ψ} of 0.70 ± 0.15 s^{-1}. Together these results indicate that substrate binding limits the rate of catalysis at low RluA concentrations. However, at higher enzyme concentrations, this limitation is overcome and catalysis becomes rate-limiting. Wright and coworkers (Wright et al., 2011) reported the pseudouridylation kinetic data for two additional pseudouridine synthases: TruB and TruA. In contrast to RluA, both enzymes lacked any concentration dependence in their k_{app} values. This suggests that substrate binding is not limiting for TruB or TruA. However, the k_{ψ} values reported for TruB (0.5 ± 0.2 s^{-1}) and TruA (0.35 ± 0.2 s^{-1}) are very similar to that of RluA (0.70 ± 0.15 s^{-1}).

Multiple turnover, steady-state kinetics analysis have been completed previously for all three enzymes (Wright et al., 2011, Ramamurthy et al., 1999). RluA has been reported to have a k_{cat} of ~0.1 s^{-1}, which is significantly slower than the k_{ψ} value reported here. This means that the overall rate of the reaction as reflected by k_{cat}, has a rate-limiting step after catalysis. Since k_{ψ} was determined under single round conditions, only one round of catalysis occurs. Whereas under the steady-state conditions, multiple rounds of catalysis occurs. In the single round experiments, RluA only needs to bind tRNA and complete the catalytic step. The steady-state experiments require that RluA binds the substrate, completes catalysis, and then releases the product in order to complete another round of catalysis. Taken together, this suggests that under single round conditions catalysis is limiting above RluA concentrations of 5 μM, but under multiple

round conditions product release is limiting resulting in the lower k_{cat} value of the overall reaction. This is in contrast to TruB and TruA as Wright et al. (2011) found that both enzymes have k_{cat} values essentially equal to their k_{ψ} values. As discussed above, this indicates that product release is not rate limiting for TruB and TruA, and catalysis itself is the rate-limiting step.

Although pseudouridine synthases differ significantly in structure and sequence, they all share the same catalytic cleft containing the catalytic aspartate residue (Hoang et al., 1998; Ramamurthy et al., 1999; Hamilton et al., 2005). This implies that all pseudouridine synthases employ the same chemical processes during catalysis. Interestingly, the k_{ψ} values reported for all three enzymes are nearly identical (~0.5 s^{-1}) which supports this hypothesis. In most enzyme-catalyzed reactions, the k_{cat} value is much higher, typically 10^2 – 10^6 s^{-1} (Voet and Voet, 2011). Compared to these values, the rate constant of pseudouridylation is extremely slow. Several possibilities exist that could help to explain such a slow rate of catalysis. First, pseudouridine synthases may not have had any evolutionary pressure to increase the rate of pseudouridine formation. Previous reports have shown that pseudouridine synthases and the pseudouridine modifications they create are not essential for the cell (Gutgsell et al., 2000; O'Connor and Gregory, 2011). It can be envisioned that the majority of the long-lived tRNA in the cell is modified and only the small fraction of newly transcribed tRNA needs to be pseudouridylated at a given time. Therefore, the slow rate of pseudouridine formation may not cause a bottleneck in tRNA maturation. Furthermore, if tRNA lacks pseudouridine modifications, it can still function and do not affect cell growth as described by Gutgsell and coworkers (2000). Unmodified tRNAs should not be detrimental to the cell as they retain the same activity and kinetic parameters in aminoacylation as modified transcripts *in vitro* (Sampson and Uhlenbeck, 1988).

Second, pseudouridine formation may not be the primary function of pseudouridine synthases. Additional proposed roles in RNA folding may be more important for the cell; and therefore pseudouridylation may have been selected to be slow in order to allow for sufficient time for RNA folding to occur (Hoang and Ferré-D'Amaré, 2001). Finally, this rate constant may be the possible upper limit of pseudouridylation as the chemistry involved in pseudouridine formation is very complex. The catalytic step itself consists of at least three steps that occur all within the catalytic pocket. The enzyme must first break the C-N glycosidic bond between the sugar and base, rotate or flip the base, and reattach the base with a C-C glycosidic bond. It may be with such a complicated catalytic mechanism that pseudouridine synthases are not able to catalyze pseudouridylation faster than 0.5 s^{-1}. The first step of catalysis, cleavage of the N-C glycosidic bond, is very similar to the reaction catalyzed by uracil-DNA glycosylases (Stivers et al., 1999; Friedman and Stivers, 2010). These enzymes remove uracil from damaged DNA and have been reported to have k_{cat} values of 4 – 200 s^{-1} (Duraffour et al., 2007; Liu et al., 2007). This indicates that, in theory, the first step of pseudouridine formation could potentially be much faster and the second or third steps are more likely rate-limiting. Additional quench-flow experiments could be envisioned where transient intermediates along the reaction pathway could be isolated and characterized to analyse the individual steps of catalysis. Further investigations involving pseudouridine synthases from the other families and other organisms are also necessary in order to establish if catalysis is uniformly slow for all pseudouridine synthases.

In trying to characterize the reaction mechanism, it was found that [^{14}C] uracil could not be exchanged with the uracil within the catalytic pocket of TruB during catalysis. There are two possibilities that could explain these observations. First, the detached uracil in TruB's catalytic pocket may not be able to escape to the surrounding solution. The

crystal structure of TruB shows the catalytic cavity lined with several hydrophobic residues and the uracil base forming stacking interactions with the aromatic ring of Tyr76, which could render it impossible for the uracil base to dissociate from TruB (Hoang and Ferré-D'Amaré, 2001). The second possibility could be that the uracil base actually forms a covalent adduct with the catalytic aspartate residue through its C_6 position according to the Michael addition mechanism (Figure 1.4; Huang et al., 1998; Ramamurthy et al., 1999). If the radioactive uracil had been able to replace the unlabelled uracil in the tRNA strand, we would have provided evidence against this mechanism. Unfortunately, our current results do not support either catalytic mechanism as we cannot discriminate between these two possibilities.

4.2 Product versus Substrate RNA binding

Our binding assays clearly showed that TruB can bind to product and substrate tRNA equally well. The endlevels of binding in the nitrocellulose titration assay were very similar, as well as the determined dissociation constants, K_D. TruB had K_D values of 3.0 ± 0.7 and 2.0 ± 0.6 µM for substrate and product tRNA, respectively. Nitrocellulose filtration assays were also completed for TruB wild-type, TruBD48A, and TruBD48C by Ramamurthy and coworkers (1999). Here, TruB had a K_D of 8.8 ± 2.0 µM for product tRNA, while TruBD48A had a K_D of 1.6 ± 0.4 µM and TruBD48C had a K_D of 1.3 ± 0.2 µM for substrate tRNA. Furthermore, our group previously measured TruBD48N's affinity for substrate tRNA to be 1.4 ± 0.3 µM (Wright et al. 2011). Hence, all results for binding of inactive TruB to unmodified substrate RNA are similar. Only the reported affinity for modified product tRNA to wild-type TruB is significantly different between Ramamurthy and coworkers (1999) and the values reported here. It is noteworthy that the experiments were conducted under slightly different buffer conditions. Ramamurthy and coworkers used a buffer containing 50 mM HEPES, pH 7.5 and 100 mM NH_4Cl in their

nitrocellulose filtration binding experiments. Compared with the buffers used here (50 mM Tris-HCl, pH 7.5, 70 mM NH_4Cl, 30 mM KCl, 1 mM EDTA, and 4 mM $MgCl_2$) it is noteworthy that both buffers have the same pH and a similar ionic strength. These slight buffer differences may or may not account for the difference. In conclusion, TruB seems to bind either equally well or slightly weaker to product tRNA than substrate RNA.

Here, the dissociation constants for substrate and product tRNA binding to RluA were determined as 1.7 ± 0.4 and 4.5 ± 0.9 µM, respectively. The product tRNA appears to bind to the catalytically active RluA to a higher extent; however, the RluAD64N preparation limited the concentrations that could be tested with the substrate tRNA. Regardless, the substrate tRNA binds 2.5-fold tighter than the product tRNA (1.7 vs. 4.5 µM). Previously, RluA wild-type has been shown to have a K_D of 4.0 ± 1.1 µM when interacting with product tRNA (Ramamurthy et al., 1999), which is in agreement with the findings reported here. Additionally, the K_{Half} value determined here through quench-flow analysis was 4.2 ± 2.5 µM, which correlates to the K_D value for the substrate tRNA given the error of the measurement. Also, RluAD64A and RluAD64C were shown to have K_D's of 1.8 ± 0.9 and 1.2 ± 0.5 µM, respectively (Ramamurthy et al., 1999), which is again in agreement with our data. Taken together, these results indicate that RluA has a slightly higher affinity for substrate tRNA than product tRNA.

Notably, the K_D values for TruB and RluA were found to be rather similar for both substrate and product tRNA. These findings indicate that the majority of the contacts required for binding occur within the tRNA body, and the presence of uridine and pseudouridine does not affect this interaction significantly. Since the affinity for product and substrate tRNA are so similar, it raises an interesting question of how the product dissociates from these enzymes. In the H/ACA ribonucleoprotein complex, it has been suggested that additional accessory proteins such as Gar1 may help with product

dissociation (Duan et al., 2009); however, in the stand-alone enzymes, which have been shown to be fully active in the absence of other proteins, this would not be the case. Also, there is no indication that additional helicases are required in the cell to accelerate product release. As discussed above, the rate of product release could be rather slow for RluA; however, TruB has been shown to not be limited by product dissociation (Wright et al. 2011). A relatively fast dissociation rate, k_{off}, could be envisioned despite the relatively low K_D values if the association rate for product tRNA binding is also relatively fast ($K_D = k_{off} / k_{on}$). Hence, the product tRNA could be rapidly binding and dissociating from TruB in a very dynamic equilibrium. To fully understand product dissociation, further investigations into the mechanism of product release are necessary. Lastly, in spite of the near identical affinities for tRNA containing uridine or pseudouridine, within the cell, the modified tRNA would be in demand for use in other cellular processes such as protein synthesis and will bind to other factors such as amino-acyl synthetases, EF-Tu and the ribosome. At the same time, new unmodified tRNA will be constantly generated through transcription. Therefore, the concentration of available free pseudouridine-containing tRNA might be limited and would not necessarily compete with unmodified tRNA for binding to pseudouridine synthases.

4.3 Truncated RNA binding

Although truncated RNAs have been used extensively in studying pseudouridine synthases, particularly in structural analysis (Hoang and Ferré-D'Amaré, 2001; Phannachet and Huang, 2004; Hoang et al., 2006), few studies have been performed to understand the kinetics of these stem loops in pseudouridine formation (Gu et al. 1998, Hamilton et al. 2006). Gu and coworkers (1998) wanted to examine the molecular recognition of tRNA by *E. coli* TruB to determine the minimal structural requirements for tRNA substrates (Gu et al., 1998). They were able to show the majority of recognition

requirements for both binding and pseudouridylation lie within the T-arm of the tRNA. It was also reported that the 17-base oligoribonucleotide corresponding to the yeast $tRNA^{Phe}$ T-arm could serve as a substrate for pseudouridylation and results in K_m and k_{cat} values similar to the full-length substrate (Gu et al., 1998). So far, no data has been available regarding the affinity of TruB has for the truncated substrate with the *E.coli* nucleotide sequence. Therefore, a 17-mer consisting of the T-arm of *E. coli* $tRNA^{Phe}$ was analysed here as a substrate for TruB. Nitrocellulose filtration assays demonstrated that this truncated substrate could bind with a similar affinity as the full-length tRNA, 3.0 ± 0.7 µM for full length tRNA versus 2.2 ± 0.8 µM for the T-arm, although to a much lower extent, 1.2 pmol bound versus 0.2 pmol bound at high protein concentrations. The nearly identical dissociation constants indicate that the majority of important contacts for substrate binding occur within the T-arm stem loop structure. The differences in binding endlevels may be due to misfolding of the T-arm at the experimental temperatures (~20°C). Using an oligonucleotide properties calculator, the melting temperature of the 17-mer used in the above experiments was estimated to be approximately 18°C (Kibbe, 2007). Therefore, the T-arm may be partially unfolded at room temperature. This unfolded substrate might not bind to TruB or dissociates during the washing step due to a higher $k_{off,}$ thus contributing to the much lower extent of RNA binding for the truncated substrate. Another possible explanation for the lower extent of binding for the T-arm could be that during the 10 minute incubation step, the reaction has not yet reached equilibrium. This could be examined by incubating these samples for longer prior to filtering them.

Multiple turnover tritium release assays examined whether the T-arm could be used as a substrate for pseudouridylation. Unlike the full-length tRNA, which reached 100% pseudouridine formation within 30 minutes, the truncated T-arm was unable to function

as a substrate and no significant pseudouridine formation was observed in multiple turnover assays. As this assay was completed at 37°C, it can be hypothesized that the majority of the substrate would be unfolded, resulting in the low levels of pseudouridine formation. Next, the multiple turnover assay was completed using both the T-arm and full-length substrates at 15 and 20°C; here only the full-length tRNA showed any pseudouridine formation. This reveals that, even at temperatures at or below the estimated melting temperatures of the T-arm, the T-arm is still an inefficient substrate for pseudouridylation under multiple turnover conditions (10 nM TruB incubated with 600 nM RNA). As full-length tRNA with a comparable K_D is fully active in pseudouridine formation under multiple turnover conditions, it is quite surprising that not even low levels of modification were observed with the T-arm. One potential explanation could be that the T-arm binds in a non-productive fashion to TruB. Only under single turnover conditions could the T-arm be effectively used to form pseudouridine (Figure 3.6 B). Here, an endlevel of only 45% was achieved compared with 100% for the full-length substrate. Again, these assays were completed at a much higher temperature (37°C) than the reported melting temperature of the stem loop (18°C). Therefore, it is conceivable that a majority of the T-arm substrate would be unfolded at the assay temperatures. The higher level of pseudouridine formation (45%) than binding at 20°C (~ 10%) is mostly likely the result of the dynamic equilibrium of the folded T-arm with the unfolded T-arm. At these relatively high concentrations, all folded T-arm will bind to TruB and will become pseudouridylated. As soon as an unfolded T-arm spontaneously folds, it can therefore rapidly bind to TruB and can be modified, explaining a higher level of modification than actual binding. These findings are in stark contrast with Gu et al. (1998) where a 17-mer consisting of T-arm of yeast $tRNA^{Phe}$ was used to complete multiple turnover tritium release assays.

Our first indication that the T-arm behaved differently than the full-length substrate occurred during fluorescence stopped-flow with a 2-aminopurine substituted T-arm substrate and TruB. It was found that with increasing enzyme concentration, the amplitude of the fluorescent signal also increased (unpublished data). These results indicated TruB had a different affinity for the truncated RNA than for the full-length substrate. In fact, when radiolabelled with [^{32}P] and used in a nitrocellulose filtration assay, this 2-aminopurine T-arm had a K_D of ~ 30 µM (Ute Kothe, personal communication). These differences between the full-length substrate and the 2-aminopurine T-arm substrate could be due to the presence of the deoxyribo-2-aminopurine modification or the short T-arm itself. Upon investigating these possibilities, the unmodified truncated T-arm behaved significantly differently than the full-length tRNA in both the extent of binding and in pseudouridine formation. The discrepancies between our findings and those by Gu et al. (1998) may be due to several factors. The yeast tRNA^{Phe} was used to prepare the 17-mer T-arm in Gu et al. (1998), while we used the *E. coli* version. However, the nucleotide sequence in the T-arm for these two species is nearly identical, and differs only in positions that should not influence binding. Gu and coworkers (1998) used a range of 0.1 – 6 µM RNA and 0.02 – 0.2 µM TruB in their tritium release experiments. These values are similar to the conditions used in our multiple turnover experiments (10 nM TruB with 600 nM RNA), but had slightly different buffer conditions (as mentioned above, but pH and ionic strength were comparable). TruB has also been shown to utilize 5-fluorouracil substituted T-arm as a substrate, although the concentrations used in the assay were much higher (12.5 µM TruB and 125 µM RNA) (Spedaliere and Mueller, 2004). X-ray crystal structures clearly show TruB in complex with the T-arm of tRNA^{Phe}, but these structures were produced using significantly larger concentrations of RNA and protein than in the tritium release assays (Hoang and Ferré-D'Amaré, 2001; Phannachet and Huang, 2004). Zhou and coworkers

(2011) also demonstrated that TruB is able to form pseudouridine in a 22-mer corresponding to the T-arm of $tRNA^{Phe}$; however, these assays again were completed at RNA and TruB concentrations of 100 and 3 µM, respectively. Therefore, the data presented here agrees well with the majority of previously published data, but the differences seen with Gu et al. (1998) cannot be fully explained.

Similar to TruB, RluA has been shown to interact with the truncated anticodon stem loop of its substrate $tRNA^{Phe}$ (Spedaliere and Mueller, 2004; Hamilton et al., 2006; Hoang et al., 2006). A 2-aminopurine modified anticodon stem loop (ASL) from $tRNA^{Phe}$ had previously been purchased from IDT. This RNA, along with the 2-aminopurine substitution, also contained a deoxyribose sugar instead of the usual ribose at this modification site. Originally this RNA was to be used in fluorescence stopped-flow studies; however, upon completion of the stopped-flow experiments an increase in the fluorescence signal amplitude with increasing enzyme concentrations was noted (data not shown). This problem indicated that there were issues in RluA binding to the 2-aminopurine modified ASL most likely due to a lower affinity for the truncated substrate compared to full-length tRNA, as seen with TruB (see above). Therefore, both the ASL and full-length $tRNA^{Phe}$ were radiolabelled with [^{32}P] in order to complete nitrocellulose filtration assays to examine the affinities for these RNA species. The full-length tRNA could bind to a much higher extent than the short stem loop. However, this difference in endlevel values could be due to the result of different amounts of free [^{32}P] ATP contaminations in the RNA preparations. Furthermore, the dissociation constant determined for the full-length tRNA was 1.2 ± 0.1 µM, approximately 10-fold lower than the K_D for the ASL (17 ± 13 µM). Several reasons can be envisioned to explain the lower affinity of RluA for this truncated RNA. First, these differences could be due to possible missing protein-tRNA interactions. However, this seems unlikely as we know from the

crystal structure that RluA binds to just the anticodon stem loop with the remaining tRNA protruding from the protein into solution (Hoang et al., 2006). Another possible explanation for the high K_D of the ASL is the presence of unfolded stem loops in the solution as discussed above for TruB. These assays were completed at room temperature and may not have allowed for this short stem loop to remain folded properly (melting temperature of 18°C) (Kibbe, 2007). RluA also has a unique mechanism of substrate interaction where it will impose a protein-induced conformation onto possible RNA targets. If the RNA can adopt the correct conformation, catalysis can proceed (Hoang et al., 2006). It may be that with the short stem loop, important interactions within the tRNA itself are missing and the anticodon stem loop alone cannot form the correct RluA-induced conformation. Finally, the decreased affinity could also be caused by the 2-aminopurine deoxyribose modification itself. Interestingly, the same 2-aminopurine deoxyribose substitution was made in the T-arm of $tRNA^{Phe}$ (IDT) for use with TruB. In this case, the affinity of [^{32}P]-labelled 2-aminopurine T arm was also shown to be very low (see above). Hence, the presence of deoxyribo-2-aminopurine could in general affect the interaction of RNA with pseudouridine synthases.

4.4 The function of TruB's PUA Domain

The pseudouridine synthase and archaeosine transglycosylase (PUA) domain can be found in enzymes from all three domains of life. TruB is the only family of pseudouridine synthases containing a PUA domain. The importance of TruB's PUA domain was investigated using several biochemical techniques. First, pre-steady-state stopped-flow kinetics was used to examine the binding interaction between TruB and full-length tRNA. A 3' end fluorescein-tagged $tRNA^{Phe}$ substrate, was rapidly mixed with unlabelled TruB, the fluorescence emission was recorded, where a decrease in the fluorescence signal was observed. These results indicate that as TruB and tRNA interact, the 3' terminus of

the tRNA molecule undergoes a change in environment which causes the decrease in the fluorescence signal. This change could be due to interactions of the acceptor arm with the PUA domain as they come within close proximity to each other as seen when the TruB crystal structure is modeled in association with full-length tRNA (Figure 4.1; Hoang and Ferré-D'Amaré, 2001). In the tRNA guanine transglycosylase structure from Ishitani and coworkers (2003), they note that the basic residues in the PUA domain specifically interact with the backbone phosphates in the acceptor stem of the tRNA (Ishitani et al., 2003). Although little sequence similarity between TruB and archaeosine tRNA guanine transglycosylase is detectable, TruB's Arg307 is in a homologous position in its PUA domain to Arg578 in the transglycosylase structure and could have a similar role in binding the acceptor stem. It is interesting to note that the fluorescence signal decreased during the stopped-flow experiments. Typically, the fluorescence is expected to increase as the fluorophore forms new interactions with the protein and is removed from the solvent (Lakowicz, 2006). Instead the decrease may be caused by the fluorophore, which was initially interacting with the acceptor arm of the tRNA, is replaced by the PUA domain of TruB, causing the fluorescein to interact (and be quenched) by the surrounding solvent.

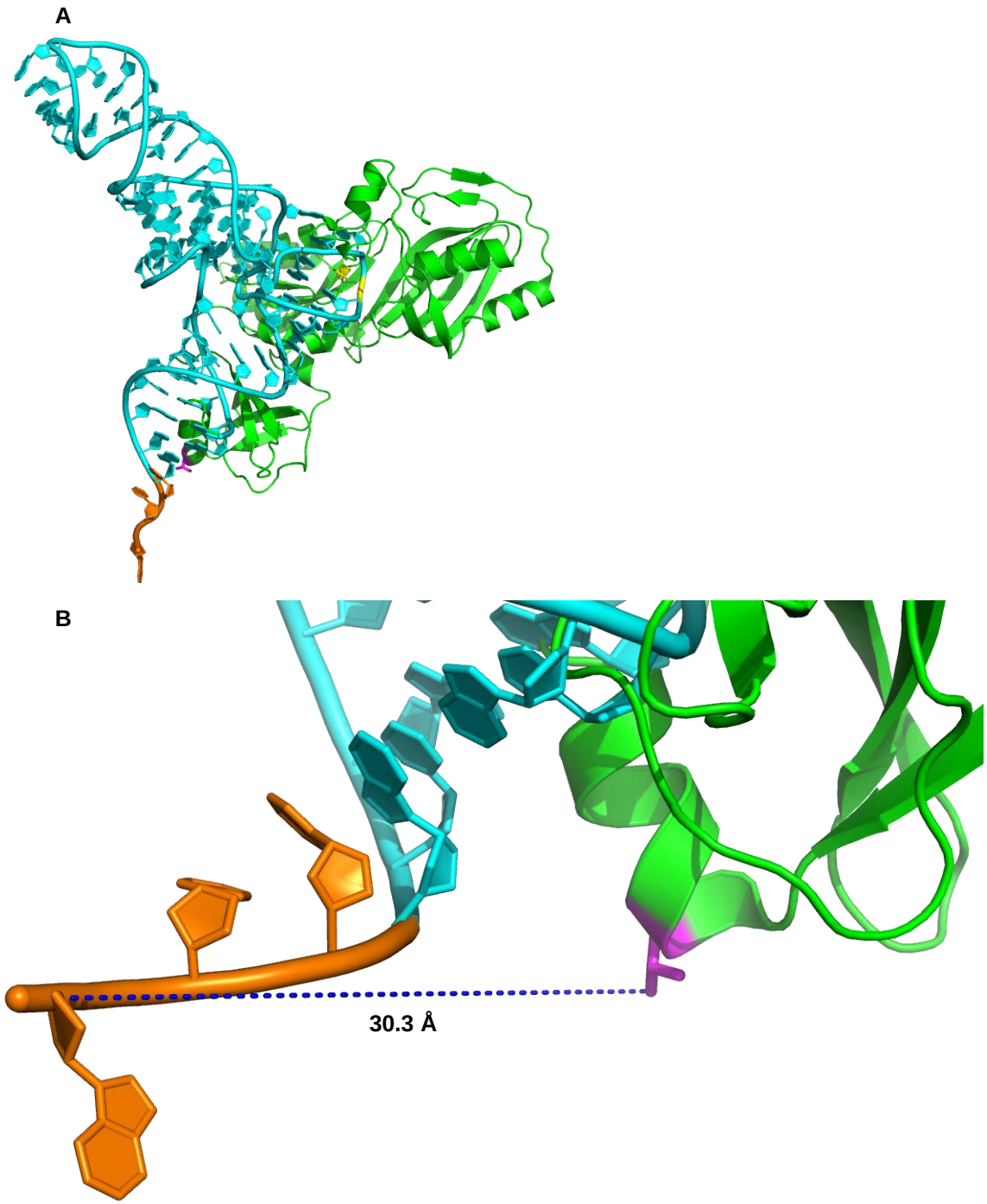


Figure 4.1: Model of TruB in complex with full-length tRNA^{Phe}. A. The model is based on the crystal structures of TruB (green; 1K8W; Hoang and Ferré-D'Amaré, 2001) and full-length yeast tRNA^{Phe} (blue; 4TNA; Hingerty et al., 1978). The crystal structure of TruB contained the structure of the T-arm (see Fig. 1.7) which was aligned here to the T-arm of full-length tRNA. Position 259 on the back surface of TruB's PUA domain is depicted in fuchsia and was used for fluorescent labelling. The 3' CCA terminus of tRNA used for labelling is shown in orange. The target uridine 55 is shown in yellow. B. According to this model and without assuming additional conformational changes, the distance between the 3' terminus of tRNA (orange) and position 259 in TruB (fuchsia) would be approximately 30 Å.

Interestingly, the k_{app} values for binding decrease as enzyme concentration increases. This dependence suggests a model where substrate binding occurs in two steps. Here, a pool of tRNA substrate undergoes a slow conformational rearrangement into binding-competent tRNA, which can rapidly bind to TruB. At low enzyme concentrations, only the available binding-competent tRNA binds to TruB resulting in a fast rate of binding. This occurs because at TruB concentrations below the K_D (~3 µM) only a small fraction of the tRNA will bind. However, with increasing enzyme concentrations, more tRNA must be converted to the binding-competent form before it can interact with TruB, as TruB will bind to all of the tRNA and not just a small fraction when the TruB concentration is above the K_D. This results in a decrease in the rate of binding as the reaction is now limited by the slow conformational change of the tRNA. As shown in Figure 4.2, this model is characterized by a slow conformational change (k_1) and a fast binding event (k_2). The rate constant of the tRNA conformational change (binding-incompetent tRNA to binding-competent tRNA) is k_1, whereas the reverse conformational change is described by the rate constant k_{-1}. Assuming the equilibrium (K_{eq}) between the two tRNA species favours the binding-incompetent tRNA (large $K_{eq} = k_{-1}/k_1$), only a small pool of binding-competent tRNA is present at the start of the reaction. As only the actual binding step is observed in our experiments, the apparent rate of binding at high TruB concentrations, i.e. complete binding, describes the rate-limiting tRNA conformational change (k_1). This k_{min} (= k_1), as determined by hyperbolic fitting, was found to be 3.7 ± 0.5 s^{-1}. This value is in good agreement with Wright et al. (2011) where it was shown that TruB could bind to tRNA at an average overall rate of 6.0 ± 1.8 s^{-1}. Substrate binding was also revealed to occur in two steps, as the k_{app} remained constant over the protein concentration range examined (Wright et al., 2011), but the signal was noisier than in the fluorescence experiments described here and the TruB concentrations were slightly higher. This could explain why the apparent rates of binding seemed to remain constant low enzyme concentrations.

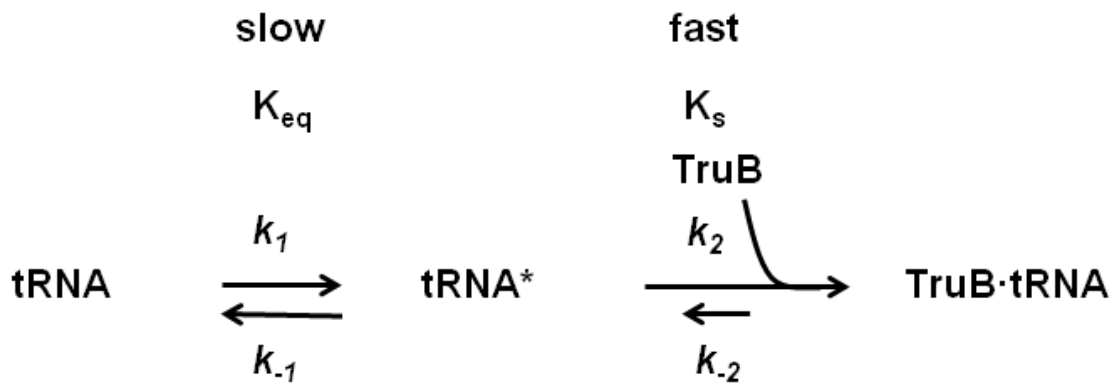


Figure 4.2: Model of TruB binding to substrate tRNA. Stopped-flow analysis of fluorescein-tRNA binding to TruB revealed a two-step binding mechanism. The first step consists of a slow conformational rearrangement within the tRNA from a binding-incompetent form (tRNA) to a binding-competent form (tRNA*) which is characterized by the rate constants k_1 and k_{-1} followed by the rapid second step where binding-competent tRNA (tRNA*) binds to TruB (rate constants k_2 and k_{-2} for association and dissociation, respectively).

In the canonical L-shape structure of tRNA, modifications within the T-arm and D-arm, specifically pseudouridine 55 and Gm18, help create the elbow of tRNA where the T-arm and D-arm interact thereby anchoring the acceptor side to the anticodon side (Bjork, 1995). Large conformational changes within the full-length tRNA have been shown to occur upon association with tRNA guanine transglycosylases which modifies G15 in the D-arm (Ishitani et al., 2003). However, the crystal structure of TruB in complex with the T-arm reveal only marginal changes between the enzyme bound and free stem-loop structures (Hoang and Ferré-D'Amaré, 2001). No crystal structure information is available for the full-length tRNA substrate in complex with TruB. Therefore, based on the findings presented here, it could be envisioned that the tRNA must first break the interactions between the T- and D-arms in order for uridine 55 to be flipped out into TruB's active site. Our data would suggest that tRNA spontaneously switches between

the binding-incompetent and binding-competent conformations, for example by loosening the interaction between D- and T-arm, and that only this form can bind to TruB. This also implies that TruB does not actively change the tRNA conformation, but rather relies on a dynamic equilibrium of conformations of the free tRNA to select the binding-competent form.

In order to establish the importance of TruB's PUA domain in tRNA substrate interactions, the PUA domain of TruB was removed via site-directed mutagenesis. The tRNA binding properties, as well as the pseudouridylation kinetics of the TruBΔPUA variants, were analysed. Nitrocellulose filtration assays showed that the determined K_D values were nearly identical to the wild-type values. Additionally, the endlevel and rate of pseudouridine formation by TruBΔPUA was essentially the same as for the wild-type protein. These results suggest that the PUA domain is dispensable for tRNA binding and catalysis. However, when stopped-flow experiments were repeated using the truncated TruB variants and 3'-labelled tRNA, no change in the fluorescence signal was observed (Figure 3.9). This indicates that the labelled 3' end of the tRNA interacts with the PUA domain of TruB, giving rise to the fluorescence changes observed with wild-type TruB. This is clear evidence for a role of the PUA domain in binding the acceptor stem of tRNA. However, although the fluorescence data suggest an interaction between TruB's PUA domain and the substrate tRNA, this interaction is not essential for tRNA binding or catalysis. As the dissociation constant reflects the free energy of binding, it seems counterintuitive that the deletion of the PUA domain, which removes contacts between the tRNA and protein, does not affect the K_D. However, as shown in the binding experiments with the truncated tRNA, the majority of contacts are found in the catalytic domain of the protein and the T-arm of the RNA. This suggests that the contacts between the tRNA's 3' end and the PUA domain are relatively weak. In this case,

removing the rather weak contacts with the PUA domain does not significantly affect the energy of binding and therefore the K_D. Notably, previous reports have indicated that the PUA domain is not essential for the function of glutamate 5-kinase or tRNA guanine transglycosylase (Perez-Arellano et al., 2005; Sabina and Soll, 2006). Furthermore a sequence alignment of 100 bacterial TruB amino acid sequences reveals very little sequence conservation within the PUA domain (Appendix 1). In fact, *Chlamydia trachomatis,* a gram-negative human pathogen, does not have a PUA domain in its TruB sequence. These findings further underline our finding that the PUA domain is not essential for TruB's pseudouridylation activity.

Finally, substrate tRNA binding to TruB was analysed using fluorescence spectrometry. As before, a fluorescein label was attached to the 3' end of $tRNA^{Phe}$, while TruB was fluorescently-labelled at position C259 in the PUA domain with 1,5-IAEDANS. When 1,5-IAEDANS was excited at 336 nm in the presence of fluorescein-tRNA, FRET occurred as shown by the decrease in 1,5-IAEDANS emission at 467 nm and an increase in the fluorescein emission at 510 nm. This is in particular remarkable as we know from previous experiments that the fluorescence of fluorescein-tRNA decreases when it binds to unlabelled TruB. However, this fact together with the fact tRNA labelling is incomplete (not 100%) prevents us from quantifying the FRET efficiency and thus determining the distance between the fluorophores. Therefore, these experiments merely demonstrate that the tRNA 3' end and the PUA domain of TruB come in proximity (distance comparable to the R_0). In future, additional control experiments are required to measure the fluorescence emission of labelled TruB in the presence of unlabelled tRNA as well as the fluorescence of labelled tRNA in the presence of unlabelled protein.

4.5 Future perspectives

To further understand the chemical mechanism involved in pseudouridine formation, the next step would be to isolate and quantify the abasic intermediate that occurs during catalysis. This will provide information on the first chemical step, N-C glycosidic bond breakage. If glycosidic bond cleavage is fast compared to the subsequent steps, an abasic intermediate should build up within a short time window. This intermediate could be isolated using quench-flow experiments and short time courses of reaction. Potentially, mass spectrometry could be used to identify such an intermediate. Furthermore, the nature of a potential covalent intermediate could be investigated by isolating it after quench-flow experiments to find out whether the catalytic aspartate attacks the uracil base or the ribose sugar. From the results provided here, we already know that all three steps in catalysis occur with an overall rate of approximately 0.5 s^{-1}. These future experiments could provide information on the individual steps of catalysis and hopefully determine the rate-limiting step. Additional analysis of the remaining pseudouridine synthase families will provide insight into whether the catalytic step is uniformly slow for all pseudouridine synthase enzymes.

This study has demonstrated that the pseudouridine synthases TruB and RluA have very similar affinities for their product and substrate tRNA targets. As they bind equally well to both uridine and pseudouridine containing tRNA, we could next analyse product release. The k_{off} rate constant of tRNA dissociation could be determined in stopped-flow experiments. Towards this aim, one could incubate the fluorescein-tagged tRNA with catalytically active TruB prior to rapidly mixing with an excess of unlabelled tRNA in the stopped-flow. Here, the dissociation of the labelled tRNA will be observed by a fluorescence increase while the unlabelled tRNA will bind to the free TruB preventing it from re-binding to labelled tRNA. Therefore, the change in the fluorescence signal over

time allows for the determination of the rate constant of dissociation. Similar experiments could be completed using the catalytically inactive TruB variant, which would provide information on the k_{off} for the substrate tRNA.

Unlike for Cbf5, TruB's PUA domain seems to be unnecessary for enzyme function. The obvious next question is: Why has this protein evolved to have such a domain? Previous studies have reported that TruB may function as a RNA chaperone within the cell (Gutgsell et al., 2000). Expression of the catalytically inactive TruBD48N from a plasmid was sufficient to restore the ability of the knockout TruB *E. coli* strain to compete with wild-type cells. Although no pseudouridylation occurred at position 55 in the tRNA of these cells, they were still able to grow as efficiently as the wild-type strain (Gutgsell et al., 2000). Mutations in the human homolog dyskerin occur in or around the PUA domain and result in the X-linked disease dysteratosis congenita (Knight et al., 1997). Additionally, these mutations in the PUA domain have been shown to interfere with the interaction of Cbf5 with the assembly factor Shq1 during RNP biogenesis (Grozdanov et al., 2009; Li et al., 2011). Lin and Momany (2003) also demonstrated that the fungus *Aspergillus nidulans* shows growth defects when a valine to phenylalanine mutation occurs within the PUA domain of the Cbf5 homolog SwoC1 revealing the PUA domain plays an important role in enzyme function which might be unrelated to its pseudouridylation activity. Therefore, the next step in studying the PUA domain would be to express the truncated TruB variant in the *E.coli* TruB knockout strain and to examine the fitness of the cells. If expression of the TruB∆PUA protein in the knockout strain cannot rescue this strain in competition with wild-type cells, this may indicate that the PUA domain is important for cellular fitness and could be involved in RNA-binding or protein-protein interactions within the cell. Immunoprecipitation assays could be

envisioned where an antibody specific for TruB could be used to “pull-down” additional interaction partners from wild-type cells.

4.6 Conclusion

This thesis presents the first pre-steady-state rapid kinetic analysis of RluA giving valuable insight into the pseudouridylation mechanism of *E. coli* pseudouridine synthases. We showed that the rate constant of pseudouridylation is approximately 0.7 s^{-1} for RluA. Together with analysis of TruB and TruA, we were able to show that all three pseudouridine synthases have a uniformly slow catalytic step (Wright et al., 2011). Investigations of the other pseudouridine synthases representing other families and organisms will determine if catalysis is uniformly slow for all pseudouridine synthases. We also demonstrated that truncated tRNAs may not function as well as biochemical substrates as previously thought. Precaution must be taken whenever using these substrates for kinetic measurements. Additionally, we found that TruB and RluA bind similarly to product and substrate RNA, therefore the kinetic analysis of product release is needed for these stand-alone pseudouridine synthases. Furthermore, the kinetic analysis of substrate binding revealed that TruB relies on a spontaneous and slow conformational change in free tRNA in order to rapidly bind to a binding-competent form of tRNA. Finally, we demonstrated that TruB’s PUA domain, like those of glutamate 5-kinase and archaeosine guanine transglycosylase, is not essential for substrate binding or the catalytic function. In summary, this study contributes significantly to our understanding of the molecular mechanisms and kinetics of tRNA interaction with the bacterial pseudouridine synthases TruB and RluA.

References

Alawi, F. and Lin, P. (2010) Loss of dyskerin reduces the accumulation of a subset of H/ACA snoRNA-derived miRNA. *Cell Cycle 9*, 2467-2469.

Anantharaman, V., Koonin, E.V. and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Research 30*, 1427-1464.

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., Buckle, M., and Grosjean, H. (1999) Pseudouridine synthetase Pus1 of *Saccharomyces cerevisiae*: Kinetic characterisation, tRNA structural requirement and real-time analysis of its complex with tRNA. *Journal of Molecular Biology 289*, 491-502.

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.

Arnez, J.G. and Steitz, T. A. (1994) Crystal structure of unmodified $tRNA^{Gln}$ complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure. *Biochemistry 33*, 7560-7567.

Bakin, A., and Ofengand, J. (1993) Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: Analysis by the application of a new sequencing technique. *Biochemistry 32*, 9754-9762.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996) The RNA world of the nucleolus: Two major families of small RNAs defined by different box elements with related functions. *Cell 86*, 823-834.

Becker, H.F., Motorin, Y., Sissler, M., Florentz, C., and Grosjean, H. (1997) Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the TΨ-loop of yeast tRNAs. *Journal of Molecular Biology 274*, 505-518.

Behm-Ansmant, I., Grosjean, H., Massenet, S., Motorin, Y., and Branlant, C. (2004) Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry 279*, 52998-53006.

Bjork, G.R. (1995). Biosynthesis and function of modified nucleosides. In *tRNA: Structure, Biosynthesis, and Function* (Söll, D. and RajBhandary, U., eds.) pp. 165-205, ASM Press, Washington, D.C.

Cabello-Villegas, J. and Nikonowicz E.P. (2005) Solution structure of Ψ32-modified anticodon stem-loop of *Escherichia coli* $tRNA^{Phe}$. *Nucleic Acids Research 33*, 6961-6971.

Chan, C.M. and Huang, R.H. (2009) Enzymatic characterization and mutational studies of TruD - the fifth family of pseudouridine synthases. *Archives of Biochemistry and Biophysics 489*, 15-19.

Charette, M. and Gray, M.W. (2000) Pseudouridine in RNA: What, where, how, and why*. IUBMB Life* 49, 341-351.

Cleland, W.W. (2009) Enzyme kinetics: Steady state. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd. Chichester, New Jersey.

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., Englund, 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 23S ribosomal RNA. *Journal of Biological Chemistry*, 273(29), 18562-18566.

Davis, D.R. (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Research 23*, 5020-5026.

Davis, D.R. (1998) Biophysical and conformational properties of modified nucleosides in RNA (nuclear magnetic resonance studies). In *Modification and Editing of RNA* (Grosjean, H. And Benne, R., eds.). pp. 103 – 112, ASM Press, Washington, D.C.

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

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.

Dennis, P.P. and Omer, A. (2005) Small non-coding RNAs in archaea. *Current Opinion in Microbiology 8*, 685-694.

Duan, J., Li, L., Lu, J., Wang, W., and Ye, K. (2009) Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Molecular Cell 34*, 427-439.

Duraffour, S., Ishchenko, A.A., Saparbaev, M., Crance, J.M., and Garin, D. (2007) Substrate specificity of homogeneous monkeypox virus uracil-DNA glycosylase. *Biochemistry 46*, 11874-11881.

Easton, L.E., Shibata, Y., and Lukavsky, P.J. (2010) Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA 16*, 647-653.

Ericsson, U.B., Nordlund, P.,and Hallberg, B.M. (2004) X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Letters 565*, 59-64.

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

Ferré-D'Amaré, A.R. (2003). RNA-modifying enzymes. *Current Opinion in Structural Biology 13*, 49-55.

Fersht, A. (1998) Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. W.H. Freeman and Co., New York.

Foster, P.G., Huang, L., Santi D.V., and Stroud, R.M. (2000) The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nature Structural & Molecular Biology 7*, 23-27.

Friedman, J.I. and Stivers, J.T. (2010) Detection of damaged DNA bases by DNA glycosylase enzymes. *Biochemistry 49*, 4957-4967.

Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry 182*, 319-326.

Goldgur, Y., Mosyak, L., Reshetnikoca, L., Ankilova, V., Lavrik, O., Khodyreva, S., and Safro, M. (1997) The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate $tRNA^{phe}$. *Structure 5*, 59-68.

Grozdanov, P.N., Fernandez-Fuentes, N., Fiser, A., and Meier, U.T. (2009) Pathogenic NAP57 mutations decrease ribonucleoprotein assembly in dyskeratosis congenita. *Human Molecular Genetics 18*, 4546-4551.

Gu, X., Liu, Y., and Santi, D.V. (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA. *PNAS 96*, 14270-14275.

Gu, X., Yu, M., Ivanetich, K.M., and Santi, D.V. (1998) Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry 37*, 339-343.

Gutgsell, N.S., 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.

Hallberg, B.M., Ericsson, U.B., Johnson, K.A., Andersen, N.M., Douthwaite, S., Nordlund, P., Beuscher IV, A.E., and Erlandsen, H. (2006) The structure of the RNA m5C methyltransferase YebU from *Escherichia coli* reveals a C-terminal RNA-recruiting PUA domain. *Journal of Molecular Biology 360*, 774-787.

Hamilton, C.S., Greco, T.M., Vizthum, C.A., Ginter, J.M., Johnston, M.V., and Mueller, E.G. (2006) Mechanistic investigations of the pseudouridine synthase RluA using RNA containing 5-fluorouridine. *Biochemistry 45*, 12029-12038.

Hamilton, C.S., Spedaliere, C.J., Ginter, J.M., Johnston, M.V., and Mueller, E.G. (2005) The roles of the essential Asp-48 and highly conserved His-43 elucidated by the pH dependence of the pseudouridine synthase TruB. *Archives of Biochemistry and Biophysics 433*, 322-334.

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

Hamma, T., Reichow, S.L., Varani, G., 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.

Harrington, K.M., Nazarenko, I.A., Dix, D.B., Thompson, R.C., and Uhlenbeck, O.C. (1993) In vitro analysis of translational rate and accuracy with an unmodified tRNA. *Biochemistry 32*, 7617-7622.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nature Genetics 19*, 32-38.

Hingerty, B., Brown, R.S., and Jack, A. (1978) Further refinement of the structure of yeast $tRNA^{Phe}$. *Journal of Molecular Biology 124*, 523-534.

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

Hoang, C. and Ferré-D'Amaré, 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., Chen, J., Vizthum, C.A., Kandel, J.M., Hamilton, C.S., Mueller, E.G., Ferré-D'Amaré, Adrian R. (2006) Crystal structure of pseudouridine synthase RluA: Indirect sequence readout through protein-induced RNA structure. *Molecular Cell* 24, 535-545.

Hoang, C. and Ferré-D'Amaré, A.R. (2001) Cocrystal structure of a tRNA pseudouridine 55 pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell 107*, 929-939.

Hoang, C., Hamilton, C.S., Mueller, E.G., Ferré-D'Amaré, A.R. (2005) Precursor complex structure of pseudouridine synthase TruB suggests coupling of active site perturbations to an RNA-sequestering peripheral protein domain. *Protein Science 14,* 2201-2206.

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

Hur, S., Stroud, R.M., and Finer-Moore, J. (2006). Substrate recognition by RNA 5-methyluridine methyltransferases and pseudouridine synthases: A structural perspective. *Journal of Biological Chemistry 281*, 38969-38973.

Hur, S. and Stroud, R.M. (2007) How U38, 39, and 40 of many tRNAs become the targets for pseudouridylation by TruA. *Molecular Cell 26*, 189-203.

Ishida, K., Kunibayashi, T., Tomikawa, C., Ochi, A., Kanai, T., Hirata, A., Iwahita, C., and Hori, H. (2011) Pseudouridine at position 55 in tRNA controls the contents of other modified nucleotides for low-temperature adaptation in the extreme-thermophilic eubacterium *Thermus thermophilus*. *Nucleic Acids Research 39*, 2304-2318.

Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell 113*, 383-394.

Johnson, K.A. (2005) Enzyme Kinetics: Transient Phase. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd. Chichester, New Jersey.

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.

Katunin, V., Soboleva, N., Mahkno, V., Sedelnikova, E., Zhenodarova, S., and Kirillov, S. (1994) Effect of the nucleotide-37 on the interaction of $tRNA^{Phe}$ with the P-site of *Escherichia coli* ribosomes. *Biochemie 76*, 51-57.

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.

Kibbe, W.A. (2007). Oligocalc: An online oligonucleotide properties calculator. *Nucleic Acids Research 35*, W43-W46.

King, T.H., Liu, B., McCully, R.R., and Fournier, M. (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Molecular Cell 11*, 425-4335.

Kinghorn, S.M., O'Byrne, C.P., Booth, I.R., and Stansfield, I. (2002) Physiological analysis of the role of TruB in *Escherichia coli*: A role for tRNA modification in extreme temperature resistance. *Microbiology 148*, 3511-3520.

Kiss, T., Fayet-Lebaron, E., and Jady, B.E. (2010) Box H/ACA small ribonucleoproteins. *Molecular Cell 37*, 597-606.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* aska library (a complete set of *E. coli* K-12 ORF archive): Unique resources for biological research. *DNA Research 12*, 291-299.

Knight, S.W., Heiss, N.S., Vulliamy, T.J., Greschner, S., Stavrides G., Pai, G.S., Lestringant, G., Varma, N., Mason, P.J., and Dokal, I. (1999) X-linked dyskeratosis

congenita is predominantly caused by missense mutations in the dkc1 gene. *American Journal of Human Genetics 65*, 50-58.

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*, 2411-2515.

Koshlap, K.M., Guenther, R., Sochacka, E., Malkiewicz, A., and Agris, P.F. (1999) A distinctive RNA fold: The solution structure of an analogue of the yeast $tRNA^{Phe}$ TΨC domain. *Biochemistry*, 38, 8647-8656.

Lakowicz, J.R. (2006) *Principles of Fluorescence Spectroscopy* (3 ed.) Springer, New York, NY.

Lancaster, L. and Noller, H.F. (2005) Involvement of 16S rRNA nucleotides G1338 and A1339 in discrimination of initiator tRNA. *Molecular Cell 20*, 623-632.

Lane, B.G. (1998) Historical perspectives on RNA nucleoside modifications. In *Modification and Editing of RNA* (Grosjean, H. And Benne, R., eds.). pp. 103 – 112, ASM Press, Washington, D.C.

Lane, B.G., Ofengand, J., and Gray, M.W. (1995) Pseudouridine and O2'-methylated nucleosides. Significance of their selective occurrence in rRNA domains that function in ribosome-catalyzed synthesis of the peptide bonds in proteins. *Biochemie 77*, 7-15.

Lane, B.G., Ofengand, J., and Gray, M. (1992). Pseudouridine in the large-subunit (23 s-like) ribosomal rna. The site of peptidyl transfer in the ribosome? *FEBS Letters, 302*, 1-4.

Lanfontaine, D.L., Bousquet-Antonelli, C., Henry, Y., Caizerques-Ferrer, M., and Tollervey, D. (1998) The box H/ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes and Development, 12*, 527-537.

Lecointe, F., Namy, O., Hatin, I., Simos, G., Rousset, J.-P., and Grosjean, H. (2002) Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases *in vivo* recoding efficiency. *Journal of Biological Chemistry 277*, 30445-30453.

Li, L. and Ye, K. (2006) Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature 443*, 302-307.

Li, S., Duan, J., Li, D., Yang, B., Dong, M., and Ye, K. (2011) Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes and Development 25*, 2409-2421.

Li, S., Duan, J. Li, D., Ma, S., and Ye, K. (2011) Structure of the Shq1-Cbf5-Nop10-Gar1 complex and implications for H/ACA RNP biogenesis and dyskeratosis congentia. *EMBO Journal 30*, 5010-5020.

Liang, B., Xue, S., Terns, R.M., Terns, M.P., and Li, H. (2007) Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nature Structural & Molecular Biology 14,* 1189-1195.

Life Technologies (2010) Molecular Probes Handbook, A guide to fluorescent probes and labeling technologies, 11th Edition. *Invitrogen*

Lin, X. and Momany, M. (2003) The *Aspergillus nidulans Swoc1* mutant shows defects in growth and development. *Genetics 165*, 543-554.

Liu, B., Yang, X., Wang, K., Tan, W., Li, H., and Tang, H. (2007) Real-time monitoring of uracil removal by uracil-DNA glycosylase using fluorescent resonance energy transfer probes. *Analytical Biochemistry 366*, 237-243.

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* aCbf5a-aNop10 complex reveal crucial roles of the Ct domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Research 34*, 826-839.

McCleverty, C.J., Hornsby, M., Spraggon, G., and Kreusch, A. (2007) Crystal structure of human Pus10, a novel pseudouridine synthase. *Journal of Molecular Biology 373*, 1243-1254.

McDonald, M.K., Miracco, E.J., Chen, J., Xie, Y., and Mueller, E.G. (2011) The handling of the mechanistic probe 5-fluorouridine by the pseudouridine synthase TruA and its consistency with the handling of the same probe by the pseudouridine synthases TruB and RluA. *Biochemistry 50*, 426-436.

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.

Mochizuki, Y., He, J., Kulkarni, S., Bessler, M., and Mason, P.J. (2004) Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing. *PNAS 101*, 10756-10761.

Motorin, Y. and Helm, M. (2010) tRNA stabilization by modified nucleotides. *Biochemistry 49,* 4934-4944.

Mueller, E.G. (2002). Chips off the old block. *Nature Structural & Molecular Biology 9*, 320-322.

Mueller, E.G. and Ferré-D'Amaré, A.R. (2009) Pseudouridine formation, the most common transglycosylation in RNA. In: DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution. Ed. Henri Grosjean. Landes Bioscience. Austin, TX.

Newby, M.I. and Geenbaum, N.L. (2001) A conserved pseudouridine modification in eukaryotic U2 snRNA induces a change in branch-site architecture. *RNA 7*, 833-845.

Nurse, K., Wrzesinski, J., Bakin, A., Lane, B.G., and Ofengand, J. (1995) Purification, cloning, and properties of the tRNA pseudouridine 55 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. (2002) Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Letters 512*, 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.

Ogle, J.M., Brodersen, D.E., Clemons, W.M., Jr., Tarry, M.J., Carter, A.P., and Ramakrishnan, V. (2001) Recognition of cognate transfer RNA by 30S ribosomal subunit. *Science 292*, 897-902.

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. *PNAS 100*, 12648-12653.

Perez-Arellano, I., Gallego, J., and Cervera, J. (2007) The PUA domain - a structural and functional overview. *FEBS Journal, 274*. 4972-4985.

Perez-Arellano, I., Rubio, V., and Cervera, J. (2005) Dissection of *Escherichia coli* glutamate 5-kinase: Functional impact of the deletion of the PUA domain. *FEBS Letters 579,* 6903-6908.

Peterson, E.T. and Uhlenbeck, O.C. (1992) Determination of recognition nucleotides for *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry 31,* 10380–10389.

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., Przybylski, P., Baudin-Baillieu, A., Rousset, J-P., and Fournier, M.J. (2008) Ribosome performance is enhanced by a rich cluster of pseudouridines in the A-site finger region of the large subunit. *Journal of Biological Chemistry 283*, 26026-26036.

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.

Ramamurthy, V., Swann, S.L., Spedaliere, C.J., and Mueller, E.G. (1999b) Role of cysteine residues in pseudouridine synthases of different families. *Biochemistry 38*, 13106-13111.

Rashid, R., Liang, B., Baker, D.L., Youssef, O.A., He, Y., Phipps, K., Terns, R.M., and Li, H. (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congentia. *Molecular Cell 21*, 249-260.

Rasmussen, L.C.V., Laursen, B.S., Mortensen, K.K., and, Sperling-Petersen, H.U. (2009) Initiator tRNAs in bacteria and eukaryotes. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd. Chichester, New Jersey.

Raychaudhuri, S., Niu, L., Conrad, J., Lane, B.G., and Ofengand, J. (1999) Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *Journal of Biological Chemistry 274*, 18880-18886.

Reinert, L.S., Shi, B., Nadi, S., Mazan-Mamczarz, K., Vitolo, M., Bachman, K.E., He, H., and Gartenhaus, R.B. (2006). MCT-1 protein interacts with the cap complex and modulates messenger RNA translational profiles. *Cancer Research 66*, 8994-9001.

Roberts, R.J. (1974) Staphylococcal transfer ribonucleic acids. II. Sequence analysis of isoaccepting glycine transfer ribonucleic acids 1A and 1B from *Staphylococcus epidermidis* Texas 26. *Journal of Biological Chemistry 249*, 4787-4796.

Roovers, M., Hale, C., Tricot, C., Terns, M.P. Terns, R.M., Grosjean, H. and Droogmans, L. (2006). Formation of the conserved pseudouridine at position 55 in archaeal tRNA. *Nucleic Acids Research 34*, 4293-4301.

Sabina, J. and Söll, D. (2006) The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation. *Journal of Biological Chemistry 281*, 6993-7001.

Sampson, J.R., DiRenzo, A.B., Behlen, L.S., and Uhlenbeck, O.C. (1989) Nucleotides in yeast $tRNA^{Phe}$ required for the specific recognition by its cognate synthetase. *Science 243,* 1363-1366.

Samuelsson, T., Guindy, YS., Lustig, F., Boren, T., and Lagerkvist, U. (1987) Apparent lack of discrimination in the reading of certain codons in *Mycoplasma mycoides*. *PNAS 84,* 3166-3170.

Samuelsson, T. and Olsson, M. (1990) Transfer RNA pseudouridine synthases in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry 265*, 8782-8787.

Sibert, B.S., Fischel-Ghodsian, N., and Patton, J.R. (2008) Partial activity is seen with many substitutions of highly conserved active site residues in human pseudouridine synthase 1. *RNA, 14*, 1895-1906.

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., Sauve, V., Larocque, R., Stura, E.A., Schrag, J.D., Cygler, M., and Matte, A. (2002) Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nature Structural & Molecular Biology 9*, 353-358.

Spedaliere, C.J. and Mueller, E.G. (2004) Not all pseudouridine synthases are potently inhibited by RNA containing 5-fluorouridine*. RNA 10,* 192-199.

Spedaliere, C.J., Hamilton, C.S., and Mueller, E.G. (2000) Functional importance of motif I of pseudouridine synthases: Mutagenesis of aligned lysine and proline residues. *Biochemistry 39*, 9459-9465.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998) Compliation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research 26*, 148-153.

Stivers, J.T., Rankiewicz, K.W., and Watanabe, K.A. (1999) Kinetic mechanism of damage site recognition and uracil flipping by *Escherichia coli* uracil DNA glycosylase. *Biochemistry 38*, 952-963.

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.

Tortoriello, G., de Celis, J.F., and Furia, M. (2010) Linking pseudouridine synthases to growth, development and cell competition. *FEBS Journal 277*, 3249-3263.

Tsui, H.C., Arps, P.J., Connolly, D.M., and Winkler, M.E. (1991) Absence of hisT-mediated tRNA pseudouridylation results in a uracil requirement that interferes with *Escherichia coli* K-12 cell division. *Journal of Bacteriology 173*, 7395-7400.

Urbonavicius, J., Durand, J.M.B., and Bjork, G.R. (2002) Three modifications in the D and T arms of tRNA influence translation in *Escherichia coli* and expression of virulence genes in *Shigella flexneri*. *Journal of Bacteriology 184*, 5348-5357.

Vendeix, F.A.P., Murphy IV, F.V., Cantara, W.A., Leszczynska, A., Gustilo, E.M., Sproat, B., Malkiewicz, A., and Agris, P.F. (2011) Human $tRNA^{Lys3}_{UUU}$ is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *Journal of Molecular Biology, 416*, 467-485.

Voet, D. and Voet, J. (2011). *Biochemistry* (4 ed.). John Wiley & Sons, Inc. Hoboken, New Jersey.

Watanabe, Y. and Gray, M.W. (2000). Evolutionary appearance of genes encoding proteins associated with box H/ACA snoRNAs: cbf5p in *Euglena gracilis*, an early diverging eukaryote, and candidate Gar1p and Nop10p homologs in archaebacteria. *Nucleic Acids Research 28*, 2342-2352.

Wright, J.R., Keffer-Wilkes, L.C., Dobing, S.R., and Kothe, U. (2011) Pre-steady-state kinetic analysis of the three *Escherichia coli* pseudouridine synthases TruB, TruA, and RluA reveals uniformly slow catalysis. *RNA 17*, 2074-2084.

Wrzenski, J., Bakin, A., Nurse, K., Lane, B.G., and Ofengand, J. (1995) Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli. Biochemistry 34*, 8904-8913.

Wrzenski, 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 pseudouridine746 in 23S RNA is also specific for pseudouridine32 in $tRNA^{Phe}$. *RNA 1,* 437-448.

Wu, H. and Feigon, J. (2007) H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *PNAS 104*, 6655-6660.

Yarian, C.S., Basti, M.M., Cain, R.J., Ansari, G., Guenther, R.H., Sochacka, E., Czerwinska, G., Malkiewicz, A., and Agris, P.F. (1999) Structural and functional roles of the N1- and N3-protons of pseudouridine at tRNA's position 39. *Nucleic Acids Research 27*, 3543-3549.

Yi, C. and Pan, T. (2011) Cellular dynamics of RNA modification. *Accounts of Chemical Research 44,* 1380-1388.

Youssef, O.A., Terns, R.M., and Terns, M.P. (2007) Dynamic interactions within sub-complexes of the H/ACA pseudouridylation guide RNP. *Nucleic Acids Research 35*, 6196-6206.

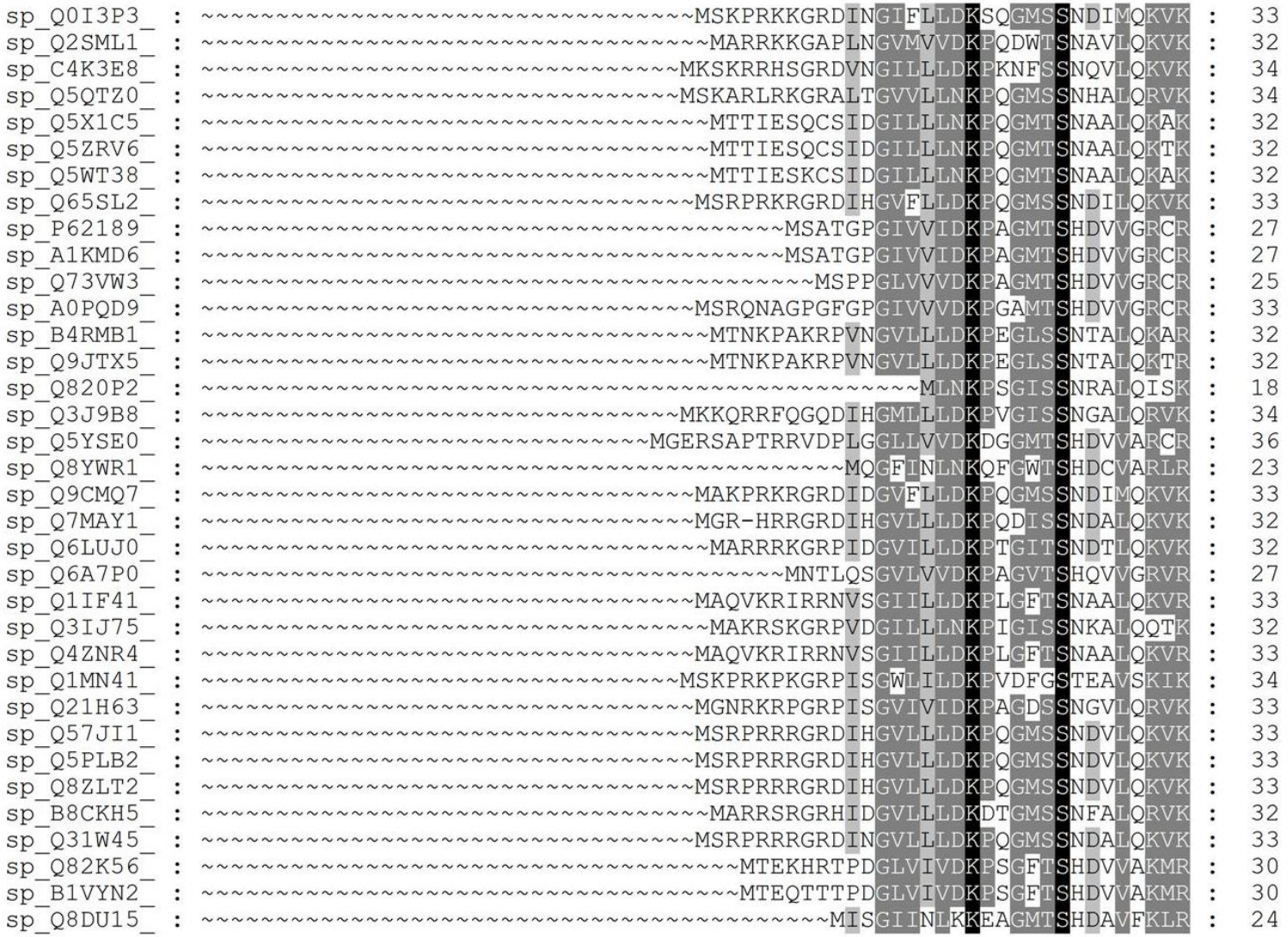
Yu, A.T., Ge, J., and Yu, T.-T. (2011) Pseudouridines in spliceosomal snRNAs. *Protein Cell 2*, 712-725.

Zebarjadian, Y., King, T., Fournier, M.J., Clarke, L., and Carbon, J. (1999) Point mutations in yeast Cbf5 can abolish *in vivo* pseudouridylation of rRNA. *Molecular and Cellular Biology 19*, 7461-7472.

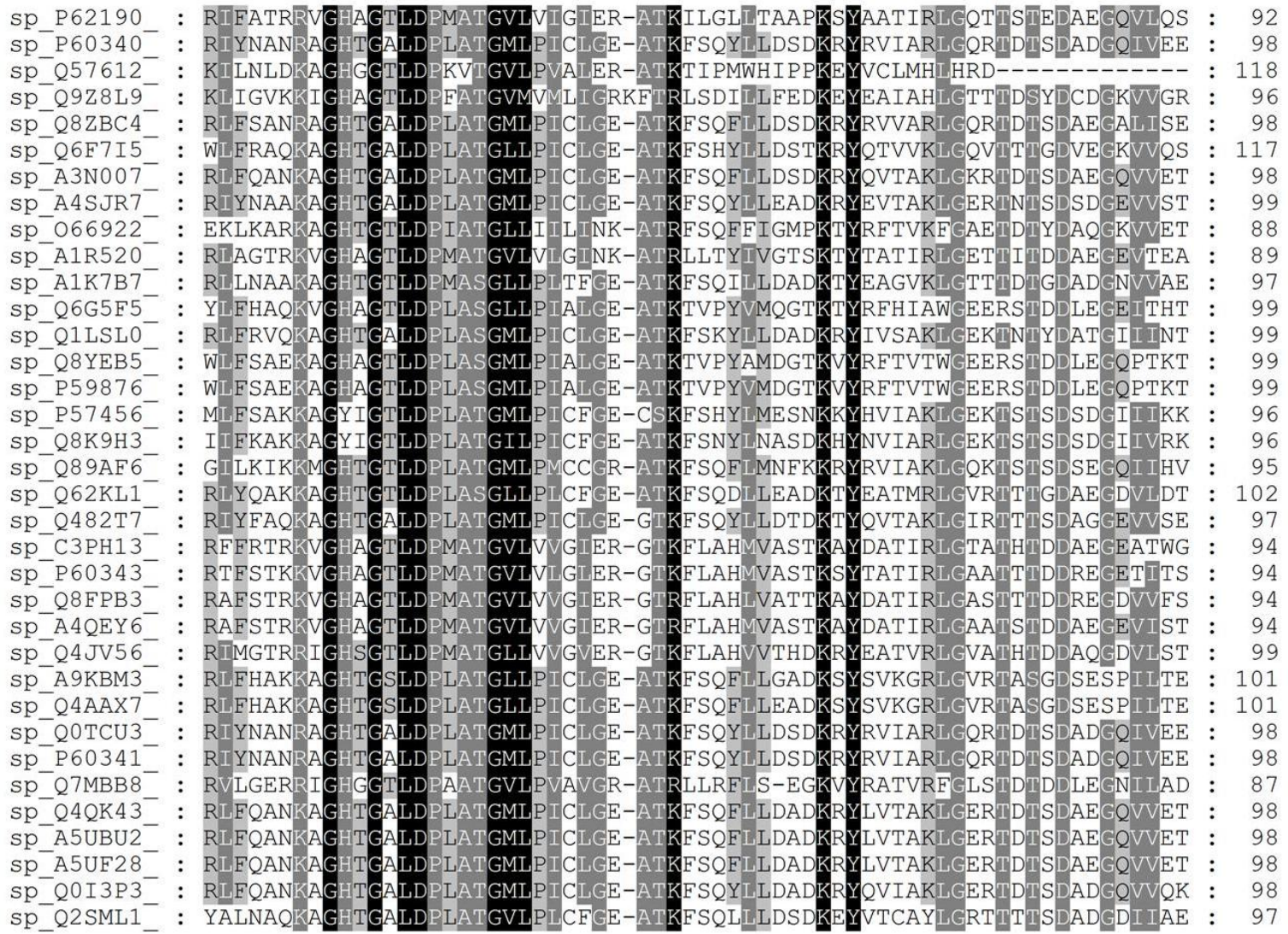
Zhou, J., Lv, C., Liang, B., Chen, M., Yang, W., and Li, H. (2010). Glycosidic bond conformation preference plays a pivotal role in catalysis of RNA pseudouridylation: A combined simulation and structural study. *Journal of Molecular Biology 401*, 690-695.

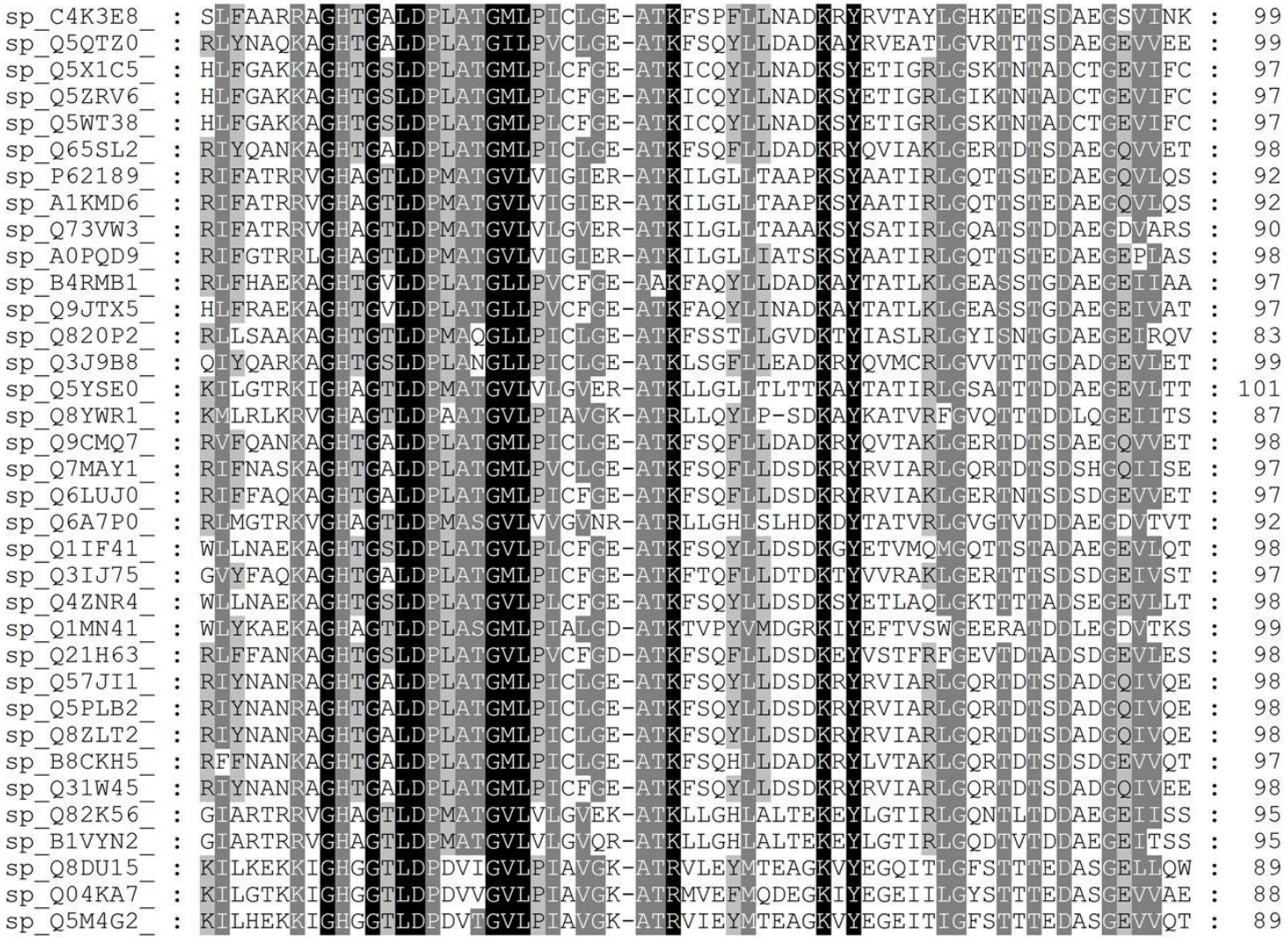
Zhou, J., Liang, B., and Li, H. (2011) Structural and functional evidence of high specificity of Cbf5 for ACA trinucleotide. *RNA 17*, 244-250.

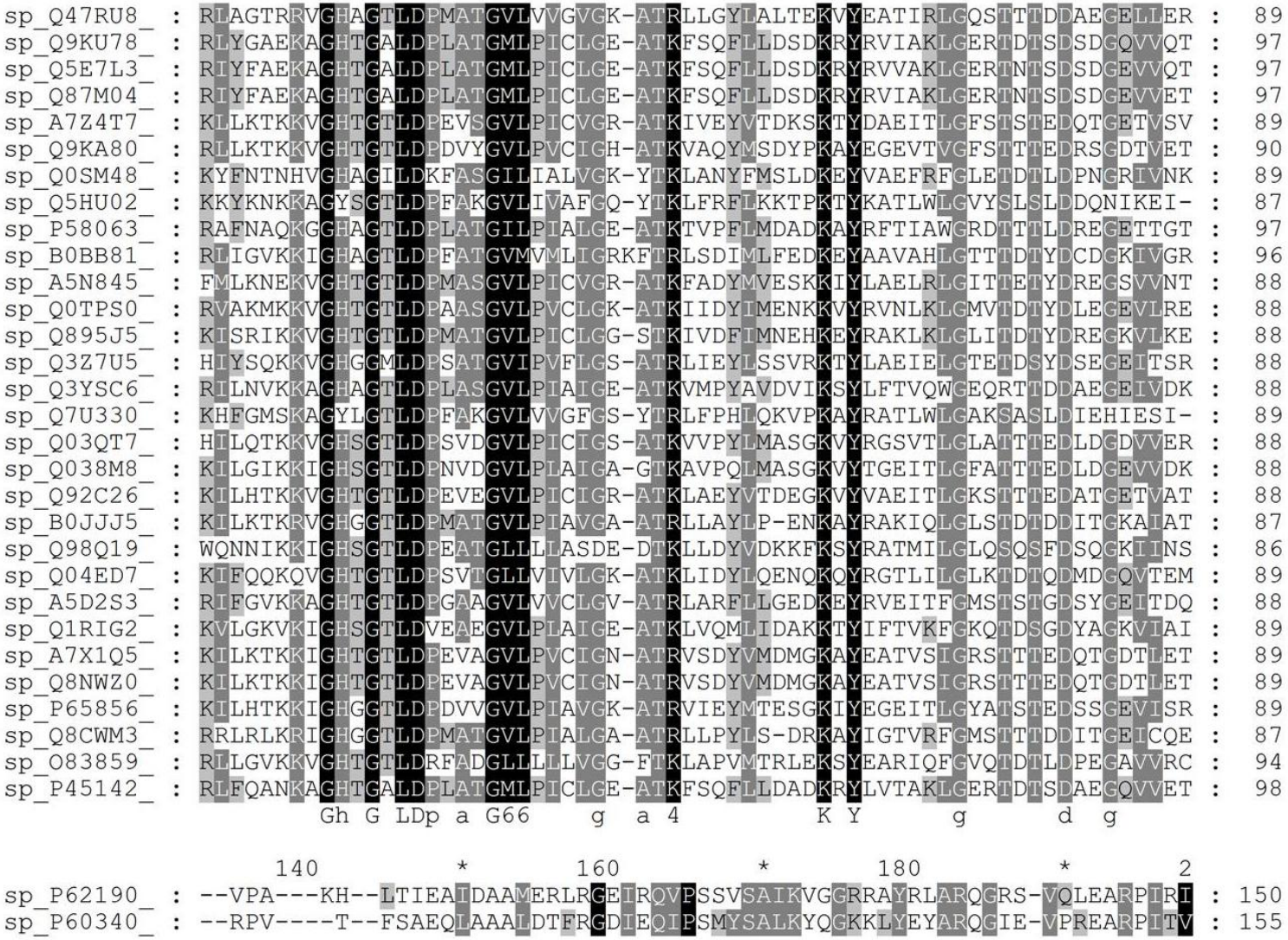
```
                      *        20         *        40         *        60
sp_P62190_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSATGPGIVVIDKPAGMTSHDVVGRCR : 27
sp_P60340_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDINGVLLLDKPQGMSSNDALQKVK : 33
sp_Q57612_ : MILLEKTQEKKINDKEELIVKEEVETNWDYGCNPYERKIEDLIKYGVVVVDKPRGPTSHEVSTWVK : 66
sp_Q9Z8L9_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MDLAVELKEGILLVDKPQGRTSFSLIRALT : 30
sp_Q8ZBC4_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MGRPRRRGRDINGVLLLDKPLGLSSNDVLQKVK : 33
sp_Q6F7I5_ : ~~~~~~~~~~~MNEFFNIIFNL-F--DVEHIMTAKSSKISRRPVSGVFLLNKPLGLSSNGVLQKVR : 52
sp_A3N007_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRKKGRDVHGVFLLDKPQGMSSNDILQKVK : 33
sp_A4SJR7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRRRRFKGRDVHGILLLDKPTGLTSNDVLQKVK : 34
sp_O66922_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MDGALLIDKPKGITSTEVVERVK : 23
sp_A1R520_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MLSGLVIVDKPQGWTSHDVVGRMR : 24
sp_A1K7B7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MQRKIPRRIVDGVLLLDKPSGMTSNGALQTAR : 32
sp_Q6G5F5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARQSKKKGRPVSGWIILDKPKGMRSTEAVSQIK : 34
sp_Q1LSL0_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MCFSLNINSHNINGIILLDKQEGLSSNYLLHKVK : 34
sp_Q8YEB5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRGKKKGRPISGWVIFDKPKGMGSTEAVSKIK : 34
sp_P59876_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRGKKKGRPISGWVIFDKPKGMGSTEAVSKIK : 34
sp_P57456_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MFFHKKRDVHGLLLLDKPQGISSNNALQKVK : 31
sp_Q8K9H3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MFFHKKRNVNGFLLLDKPKGMTSNNVLQKVK : 31
sp_Q89AF6_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MYSEFRSIDGIILIDKPYGLSSHETLQKVK : 30
sp_Q62KL1_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTTVSPRPRMARRALDGVLLLDKPVGLSSNDALMRAK : 37
sp_Q482T7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MAKRRKGRQVNGVLLLDKPHGLSSNHALQTVK : 32
sp_C3PH13_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTDPLERSGLVVVDKPAGMTSHDVVGKLR : 29
sp_P60343_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNDALADSGLVIVDKPQGMTSHDVVSKIR : 29
sp_Q8FPB3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNDPVQNSGLVVVDKPAGMTSHDVVSKLR : 29
sp_A4QEY6_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNAPAPKPGLVIVDKPAGMTSHDVVSKLR : 29
sp_Q4JV56_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNQRSARSVLSRSGVVLVDKPSGPTSHDMVAKLR : 34
sp_A9KBM3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTATNHPKLFKRPVDGVLLLDKPGGMTSNEALQRVK : 36
sp_Q4AAX7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTATNHPKLFKRPVDGVLLLDKPGGMTSNEALQRVK : 36
sp_Q0TCU3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDINGVLLLDKPQGMSSNDALQKVK : 33
sp_P60341_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDINGVLLLDKPQGMSSNDALQKVK : 33
sp_Q7MBB8_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MEGFLNLDKPAGLTSHDCIARLR : 23
sp_Q4QK43_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRKRWRDVDGVFLLDKPQGMSSNDIMQKVK : 33
sp_A5UBU2_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRKRWRDVDGVFLLDKPQGMSSNDIMQKVK : 33
sp_A5UF28_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRKRWRDVDGVFLLDKPQGMSSNDIMQKVK : 33
```

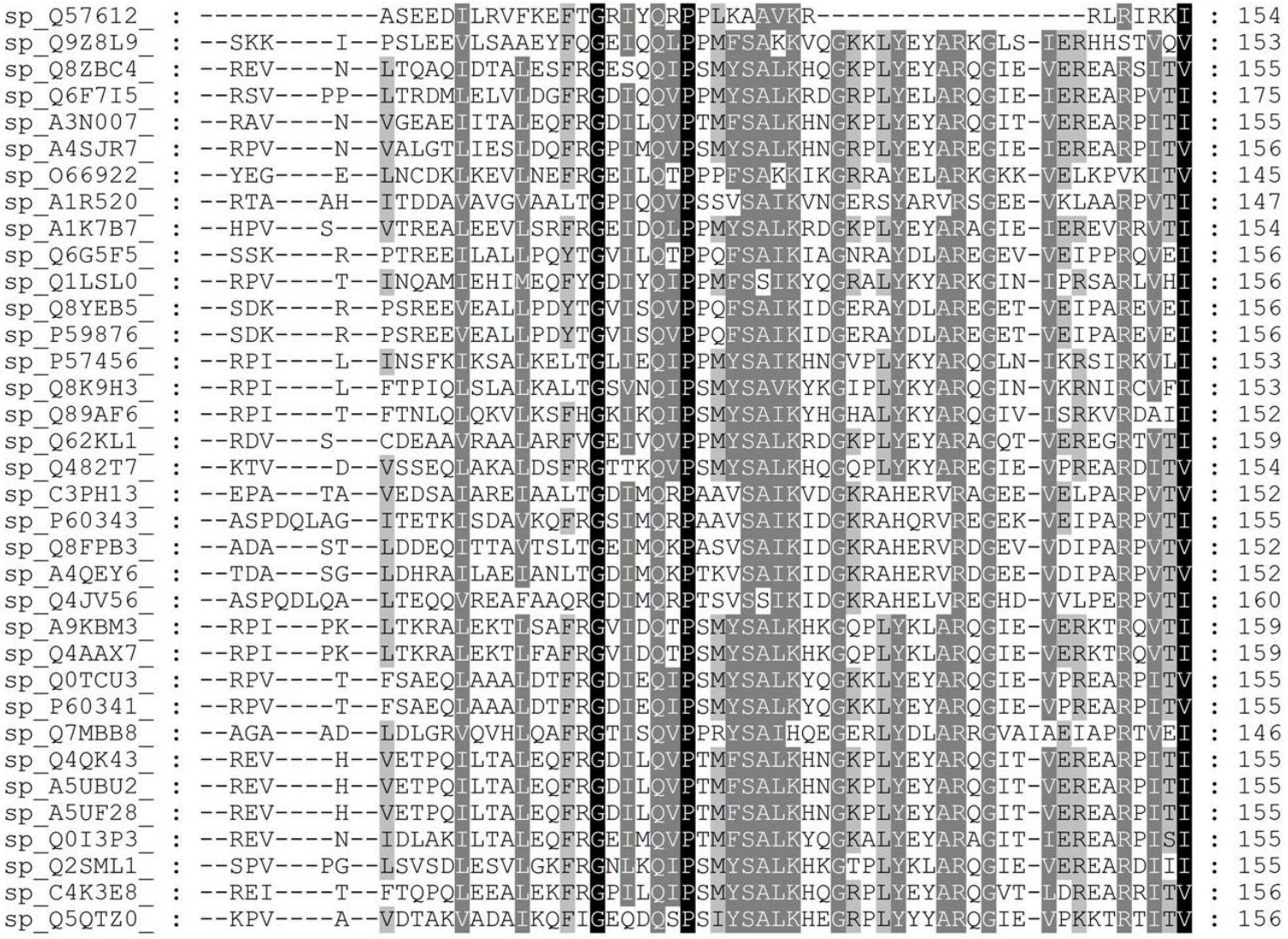

sp_Q0I3P3_ : ~~~~~~MSKPRKKGRDINGIFLLDKSQGMSSNDIMQKVK : 33
sp_Q2SML1_ : ~~~~~~MARRKKGAPLNGVMVVDKPQDWTSNAVLQKVK : 32
sp_C4K3E8_ : ~~~~~~MKSKRRHSGRDVNGILLLDKPKNFSSNQVLQKVK : 34
sp_Q5QTZ0_ : ~~~~~~MSKARLRKGRALTGVVLLNKPQGMSSNHALQRVK : 34
sp_Q5X1C5_ : ~~~~~~MTTIESQCSIDGILLLNKPQGMTSNAALQKAK : 32
sp_Q5ZRV6_ : ~~~~~~MTTIESQCSIDGILLLNKPQGMTSNAALQKTK : 32
sp_Q5WT38_ : ~~~~~~MTTIESKCSIDGILLLNKPQGMTSNAALQKAK : 32
sp_Q65SL2_ : ~~~~~~MSRPRKRGRDIHGVFLLDKPQGMSSNDILQKVK : 33
sp_P62189_ : ~~~~~~MSATGPGIVVIDKPAGMTSHDVVGRCR : 27
sp_A1KMD6_ : ~~~~~~MSATGPGIVVIDKPAGMTSHDVVGRCR : 27
sp_Q73VW3_ : ~~~~~~MSPPGLVVVDKPAGMTSHDVVGRCR : 25
sp_A0PQD9_ : ~~~~~~MSRQNAGPGFGPGIVVVDKPGAMTSHDVVGRCR : 33
sp_B4RMB1_ : ~~~~~~MTNKPAKRPVNGVLLLDKPEGLSSNTALQKAR : 32
sp_Q9JTX5_ : ~~~~~~MTNKPAKRPVNGVLLLDKPEGLSSNTALQKTR : 32
sp_Q820P2_ : ~~~~~~MLNKPSGISSNRALQISK : 18
sp_Q3J9B8_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MKKQRRFQGQDIHGMLLLDKPVGISSNGALQRVK : 34
sp_Q5YSE0_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MGERSAPTRRVDPLGGLLVVDKDGGMTSHDVVARCR : 36
sp_Q8YWR1_ : ~~~~~~MQGFINLNKQFGWTSHDCVARLR : 23
sp_Q9CMQ7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MAKPRKRGRDIDGVFLLDKPQGMSSNDIMQKVK : 33
sp_Q7MAY1_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MGR-HRRGRDIHGVLLLDKPQDISSNDALQKVK : 32
sp_Q6LUJ0_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRRKGRPIDGVILLDKPTGITSNDTLQKVK : 32
sp_Q6A7P0_ : ~~~~~~MNTLQSGVLVVDKPAGVTSHQVVGRVR : 27
sp_Q1IF41_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MAQVKRIRRNVSGIILLDKPLGFTSNAALQKVR : 33
sp_Q3IJ75_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MAKRSKGRPVDGILLLNKPIGISSNKALQQTK : 32
sp_Q4ZNR4_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MAQVKRIRRNVSGIILLDKPLGFTSNAALQKVR : 33
sp_Q1MN41_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSKPRKPKGRPISGWLILDKPVDFGSTEAVSKIK : 34
sp_Q21H63_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MGNRKRPGRPISGVIVIDKPAGDSSNGVLQRVK : 33
sp_Q57JI1_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDIHGVLLLDKPQGMSSNDVLQKVK : 33
sp_Q5PLB2_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDIHGVLLLDKPQGMSSNDVLQKVK : 33
sp_Q8ZLT2_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDIHGVLLLDKPQGMSSNDVLQKVK : 33
sp_B8CKH5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRSRGRHIDGVLLLDKDTGMSSNFALQRVK : 32
sp_Q31W45_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRRRGRDINGVLLLDKPQGMSSNDALQKVK : 33
sp_Q82K56_ : ~~~~~~MTEKHRTPDGLVIVDKPSGFTSHDVVAKMR : 30
sp_B1VYN2_ : ~~~~~~MTEQTTTPDGLVIVDKPSGFTSHDVVAKMR : 30
sp_Q8DU15_ : ~~~~~~MISGIINLKKEAGMTSHDAVFKLR : 24

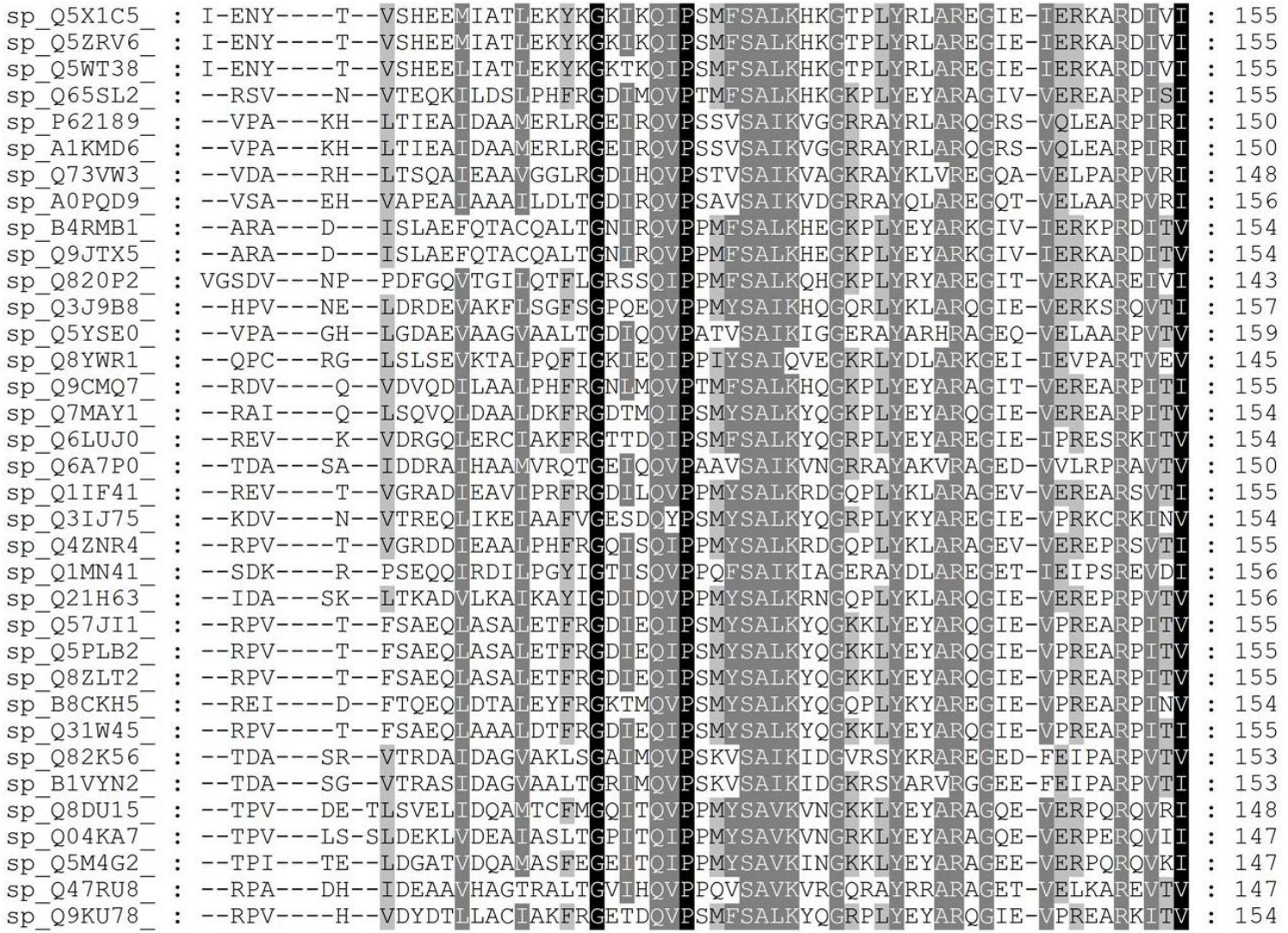
```
sp_Q04KA7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGIINLKKEAGMTSHDAVFKLR : 23
sp_Q5M4G2_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MISGIINLKKEAGMTSHDAVFKLR : 24
sp_Q47RU8_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MADGVVIVDKPAGWTSHDVVARVR : 24
sp_Q9KU78_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRRKGRVIHGVILLDKPTGISSNDALQKVK : 32
sp_Q5E7L3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRRKGRPVNGVILIDKPTGITSNDTLQKVK : 32
sp_Q87M04_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRRKGRPINGVILLDKPTGISSNDALQKVK : 32
sp_A7Z4T7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MVNGVLLLHKPVGMTSHDCVMKIR : 24
sp_Q9KA80_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MDMTGILPLAKPRGMTSHDCVAKLR : 25
sp_Q0SM48_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MENGFLLINKEQGKTSFETLFPIK : 24
sp_Q5HU02_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNKIFAAFKPKGLSSNAFLSTLK : 23
sp_P58063_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MARRKKGDAVSGWLCLDKPYDLTSTTAVSRVR : 32
sp_B0BB81_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MELATESIEGVLLVDKPQGRTSFSLIRSLV : 30
sp_A5N845_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MDGILNINKPEGMTSFDVVRKVR : 23
sp_Q0TPS0_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGVINIYKNTGMTSFDVVAIVR : 23
sp_Q895J5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGVINVYKPQNITSFDVVRIIK : 23
sp_Q3Z7U5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGILNINKPPGLTSFGVVSKVR : 23
sp_Q3YSC6_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MYGWVNLDKPCGMSSALAVNLVK : 23
sp_Q7U330_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MANALLVAAYKPPFLSSNACLSRLK : 25
sp_Q03QT7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGIIPLYKERGLTSFDCVAKLR : 23
sp_Q038M8_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGILPLYKPTGMTSADAVYHAR : 23
sp_Q92C26_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGIIPLWKERGMTSHDCVFKLR : 23
sp_B0JJJ5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MFGFLNLNKPPDWTSHDCVAKVR : 23
sp_Q98Q19_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MFYLIYKEKGISSFKAIKDFA : 21
sp_Q04ED7_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MYNGIVLVDKPAGLTSFDVVAKLR : 24
sp_A5D2S3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNGIVNVLKPPGMSSHDVVDRIR : 23
sp_Q1RIG2_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MNSYWLNVYKPRGISSAKLVSIIK : 24
sp_A7X1Q5_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MYNGILPVYKERGLTSHDVVFKLR : 24
sp_Q8NWZ0_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MYNGILPVYKERGLTSHDVVFKLR : 24
sp_P65856_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MITGIINLKKEAGMTSHDAVFKLR : 24
sp_Q8CWM3_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MFGFLNLNKPAGCTSHDCINELR : 23
sp_O83859_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MCRLSMPDAIVPFAKVSGLTSFAALAQVR : 29
sp_P45142_ : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MSRPRKRWRDVDGVFLLDKPQGMSSNDIMQKVK : 33
                                                           g     Kp g  S

                *        80         *        100        *        120        *
```

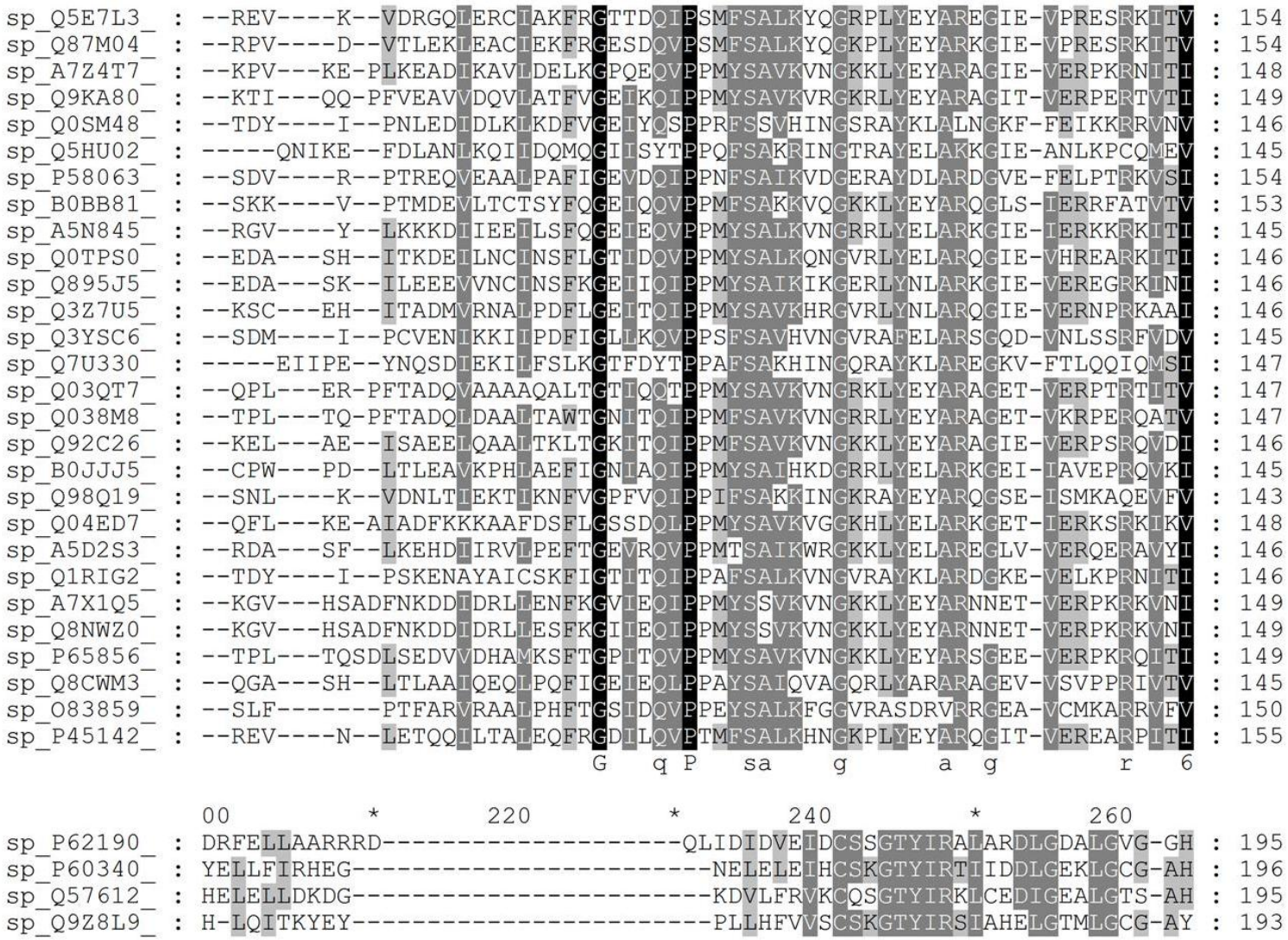

sp_P62190_ : RIFATRRVGHAGTLDPMATGVLVIGIER-ATKILGLLTAAPKSYAATIRLGQTTSTEDAEGQVLQS : 92
sp_P60340_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVEE : 98
sp_Q57612_ : KILNLDKAGHGGTLDPKVTGVLPVALER-ATKTIPMWHIPPKEYVCLMHLHRD---------- : 118
sp_Q9Z8L9_ : KLIGVKKIGHAGTLDPFATGVMVMLIGRKFTRLSDILLFEDKEYEAIAHLGTTTDSYDCDGKVVGR : 96
sp_Q8ZBC4_ : RLFSANRAGHTGALDPLATGMLPICLGE-ATKFSQFLLDSDKRYRVVARLGQRTDTSDAEGALISE : 98
sp_Q6F7I5_ : WLFRAQKAGHTGALDPLATGLLPICLGE-ATKFSHYLLDSTKRYQTVVKLGQVTTTGDVEGKVVQS : 117
sp_A3N007_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDSDKRYQVTAKLGKRTDTSDAEGQVVET : 98
sp_A4SJR7_ : RIYNAAKAGHTGALDPLATGMLPICLGE-ATKFSQYLLEADKRYEVTAKLGERTNTSDSDGEVVST : 99
sp_O66922_ : EKLKARKAGHTGTLDPIATGLLIILINK-ATRFSQFFIGMPKTYRFTVKFGAETDTYDAQGKVVET : 88
sp_A1R520_ : RLAGTRKVGHAGTLDPMATGVLVLGINK-ATRLLTYIVGTSKTYTATIRLGETTITDDAEGEVTEA : 89
sp_A1K7B7_ : RLLNAAKAGHTGTLDPMASGLLPLTFGE-ATKFSQILLDADKTYEAGVKLGTTTDTGDADGNVVAE : 97
sp_Q6G5F5_ : YLFHAQKVGHAGTLDPLASGLLPIALGE-ATKTVPYVMQGTKTYRFHIAWGEERSTDDLEGEITHT : 99
sp_Q1LSL0_ : RLFRVQKAGHTGALDPLASGMLPICLGE-ATKFSKYLLDADKRYIVSAKLGEKTNTYDATGIIINT : 99
sp_Q8YEB5_ : WLFSAEKAGHAGTLDPLASGMLPIALGE-ATKTVPYAMDGTKVYRFTVTWGEERSTDDLEGQPTKT : 99
sp_P59876_ : WLFSAEKAGHAGTLDPLASGMLPIALGE-ATKTVPYVMDGTKVYRFTVTWGEERSTDDLEGQPTKT : 99
sp_P57456_ : MLFSAKKAGYIGTLDPLATGMLPICFGE-CSKFSHYLMESNKKYHVIAKLGEKTSTSDSDGIIIKK : 96
sp_Q8K9H3_ : IIFKAKKAGYIGTLDPLATGILPICFGE-ATKFSNYLNASDKHYNVIARLGEKTSTSDSDGIIVRK : 96
sp_Q89AF6_ : GILKIKKMGHTGTLDPLATGMLPMCCGR-ATKFSQFLMNFKKRYRVIAKLGQKTSTSDSEGQIIHV : 95
sp_Q62KL1_ : RLYQAKKAGHTGTLDPLASGLLPLCFGE-ATKFSQDLLEADKTYEATMRLGVRTTTGDAEGDVLDT : 102
sp_Q482T7_ : RIYFAQKAGHTGALDPLATGMLPICLGE-GTKFSQYLLDTDKTYQVTAKLGIRTTTSDAGGEVVSE : 97
sp_C3PH13_ : RFFRTRKVGHAGTLDPMATGVLVVGIER-GTKFLAHMVASTKAYDATIRLGTATHTDDAEGEATWG : 94
sp_P60343_ : RTFSTKKVGHAGTLDPMATGVLVLGLER-GTKFLAHMVASTKSYTATIRLGAATTTDDREGETITS : 94
sp_Q8FPB3_ : RAFSTRKVGHAGTLDPMATGVLVVGIER-GTRFLAHLVATTKAYDATIRLGASTTTDDREGDVVFS : 94
sp_A4QEY6_ : RAFSTRKVGHAGTLDPMATGVLVVGIER-GTRFLAHMVASTKAYDATIRLGAATSTDDAEGEVIST : 94
sp_Q4JV56_ : RIMGTRRIGHSGTLDPMATGLLVVGVER-GTKFLAHVVTHDKRYEATVRLGVATHTDDAQGDVLST : 99
sp_A9KBM3_ : RLFHAKKAGHTGSLDPLATGLLPICLGE-ATKFSQFLLGADKSYSVKGRLGVRTASGDSESPILTE : 101
sp_Q4AAX7_ : RLFHAKKAGHTGSLDPLATGLLPICLGE-ATKFSQFLLEADKSYSVKGRLGVRTASGDSESPILTE : 101
sp_Q0TCU3_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVEE : 98
sp_P60341_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVEE : 98
sp_Q7MBB8_ : RVLGERRIGHGGTLDPAATGVLPVAVGR-ATRLLRFLS-EGKVYRATVRFGLSTDTDDLEGNILAD : 87
sp_Q4QK43_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYLVTAKLGERTDTSDAEGQVVET : 98
sp_A5UBU2_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYLVTAKLGERTDTSDAEGQVVET : 98
sp_A5UF28_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYLVTAKLGERTDTSDAEGQVVET : 98
sp_Q0I3P3_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQYLLDADKRYQVIAKLGERTDTSDADGQVVQK : 98
sp_Q2SML1_ : YALNAQKAGHTGALDPLATGVLPLCFGE-ATKFSQLLLDSDKEYVTCAYLGRTTTSDADGDIIAE : 97

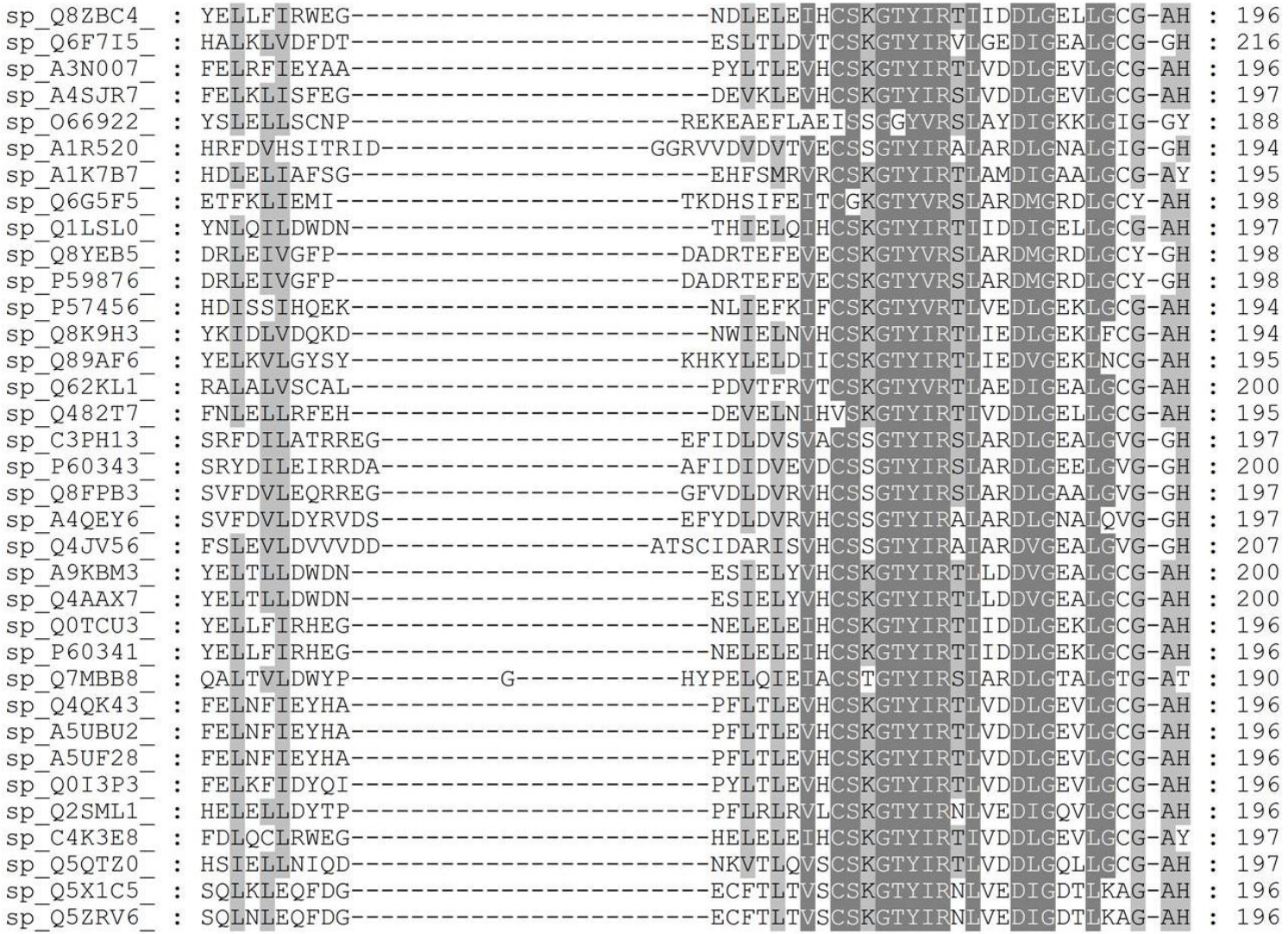

sp_C4K3E8_ : SLFAARRAGHTGALDPLATGMLPICLGE-ATKFSPFLLNADKRYRVTAYLGHKTETSDAEGSVINK : 99
sp_Q5QTZ0_ : RLYNAQKAGHTGALDPLATGILPVCLGE-ATKFSQYLLDADKAYRVEATLGVRTTTSDAEGEVVEE : 99
sp_Q5X1C5_ : HLFGAKKAGHTGSLDPLATGMLPLCFGE-ATKICQYLLNADKSYETIGRLGSKTNTADCTGEVIFC : 97
sp_Q5ZRV6_ : HLFGAKKAGHTGSLDPLATGMLPLCFGE-ATKICQYLLNADKSYETIGRLGIKTNTADCTGEVIFC : 97
sp_Q5WT38_ : HLFGAKKAGHTGSLDPLATGMLPLCFGE-ATKICQYLLNADKSYETIGRLGSKTNTADCTGEVIFC : 97
sp_Q65SL2_ : RIYQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYQVIAKLGERTDTSDAEGQVVET : 98
sp_P62189_ : RIFATRRVGHAGTLDPMATGVLVIGIER-ATKILGLLTAAPKSYAATIRLGQTTSTEDAEGQVLQS : 92
sp_A1KMD6_ : RIFATRRVGHAGTLDPMATGVLVIGIER-ATKILGLLTAAPKSYAATIRLGQTTSTEDAEGQVLQS : 92
sp_Q73VW3_ : RIFATRRVGHAGTLDPMATGVLVLGVER-ATKILGLLTAAAKSYSATIRLGQATSTDDAEGDVARS : 90
sp_A0PQD9_ : RIFGTRRLGHAGTLDPMATGVLVIGIER-ATKILGLLIATSKSYAATIRLGQTTSTEDAEGEPLAS : 98
sp_B4RMB1_ : RLFHAEKAGHTGVLDPLATGLLPVCFGE-AAKFAQYLLDADKAYTATLKLGEASSTGDAEGEIIAA : 97
sp_Q9JTX5_ : HLFRAEKAGHTGVLDPLATGLLPVCFGE-ATKFAQYLINADKAYTATLKLGEASSTGDAEGEIVAT : 97
sp_Q820P2_ : RLLSAAKAGHTGTLDPMAQGLLPICLGE-ATKFSSTLLGVDKTYIASLRLGYISNTGDAEGEIRQV : 83
sp_Q3J9B8_ : QIYQARKAGHTGSLDPLANGLLPICLGE-ATKLSGFLLEADKRYQVMCRLGVVTTTGDADGEVIET : 99
sp_Q5YSE0_ : KILGTRKIGHAGTLDPMATGVLVLGVER-ATKLLGLLTLTTKAYTATIRLGSATTTDDAEGEVLTT : 101
sp_Q8YWR1_ : KMLRLKRVGHAGTLDPAATGVLPIAVGK-ATRLLQYLP-SDKAYKATVRFGVQTTTDDLQGEIITS : 87
sp_Q9CMQ7_ : RVFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYQVTAKLGERTDTSDAEGQVVET : 98
sp_Q7MAY1_ : RIFNASKAGHTGALDPLATGMLPVCLGE-ATKFSQFLLDSDKRYRVIARLGQRTDTSDSHGQIISE : 97
sp_Q6LUJ0_ : RIFFAQKAGHTGALDPLATGMLPICFGE-ATKFSQFLLDSDKRYRVIAKLGERTNTSDSDGEVVET : 97
sp_Q6A7P0_ : RLMGTRKVGHAGTLDPMASGVLVVGVNR-ATRLLGHLSLHDKDYTATVRLGVGTVTDDAEGDVTVT : 92
sp_Q1IF41_ : WLLNAEKAGHTGSLDPLATGVLPLCFGE-ATKFSQYLLDSDKGYETVMQMGQTTSTADAEGEVLQT : 98
sp_Q3IJ75_ : GVYFAQKAGHTGALDPLATGMLPICFGE-ATKFTQFLLDTDKTYVVRAKLGERTTTSDSDGEIVST : 97
sp_Q4ZNR4_ : WLLNAEKAGHTGSLDPLATGVLPLCFGE-ATKFSQYLLDSDKSYETLAQLGKTTTTADSEGEVLLT : 98
sp_Q1MN41_ : WLYKAEKAGHAGTLDPLASGMLPIALGD-ATKTVPYVMDGRKIYEFTVSWGEERATDDLEGDVTKS : 99
sp_Q21H63_ : RLFFANKAGHTGSLDPLATGVLPVCFGD-ATKFSQFLLDSDKEYVSTFRFGEVTDTADSDGEVLES : 98
sp_Q57JI1_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVQE : 98
sp_Q5PLB2_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVQE : 98
sp_Q8ZLT2_ : RIYNANRAGHTGALDPLATGMLPICLGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVQE : 98
sp_B8CKH5_ : RFFNANKAGHTGALDPLATGMLPICLGE-ATKFSQHLLDADKRYLVTAKLGQRTDTSDSDGEVVQT : 97
sp_Q31W45_ : RIYNANRAGHTGALDPLATGMLPICFGE-ATKFSQYLLDSDKRYRVIARLGQRTDTSDADGQIVEE : 98
sp_Q82K56_ : GIARTRRVGHAGTLDPMATGVLVLGVEK-ATKLLGHLALTEKEYLGTIRLGQNTLTDDAEGEIISS : 95
sp_B1VYN2_ : GIARTRRVGHAGTLDPMATGVLVLGVQR-ATKLLGHLALTEKEYLGTIRLGQDTVTDDAEGEITSS : 95
sp_Q8DU15_ : KILKEKKIGHGGTLDPDVIGVLPIAVGK-ATRVLEYMTEAGKVYEGQITLGFSTTTEDASGELLQW : 89
sp_Q04KA7_ : KILGTKKIGHGGTLDPDVVGVLPIAVGK-ATRMVEFMQDEGKIYEGEIILGYSTTTEDASGEVVAE : 88
sp_Q5M4G2_ : KILHEKKIGHGGTLDPDVTGVLPIAVGK-ATRVIEYMTEAGKVYEGEITIGFSTTTEDASGEVVQT : 89

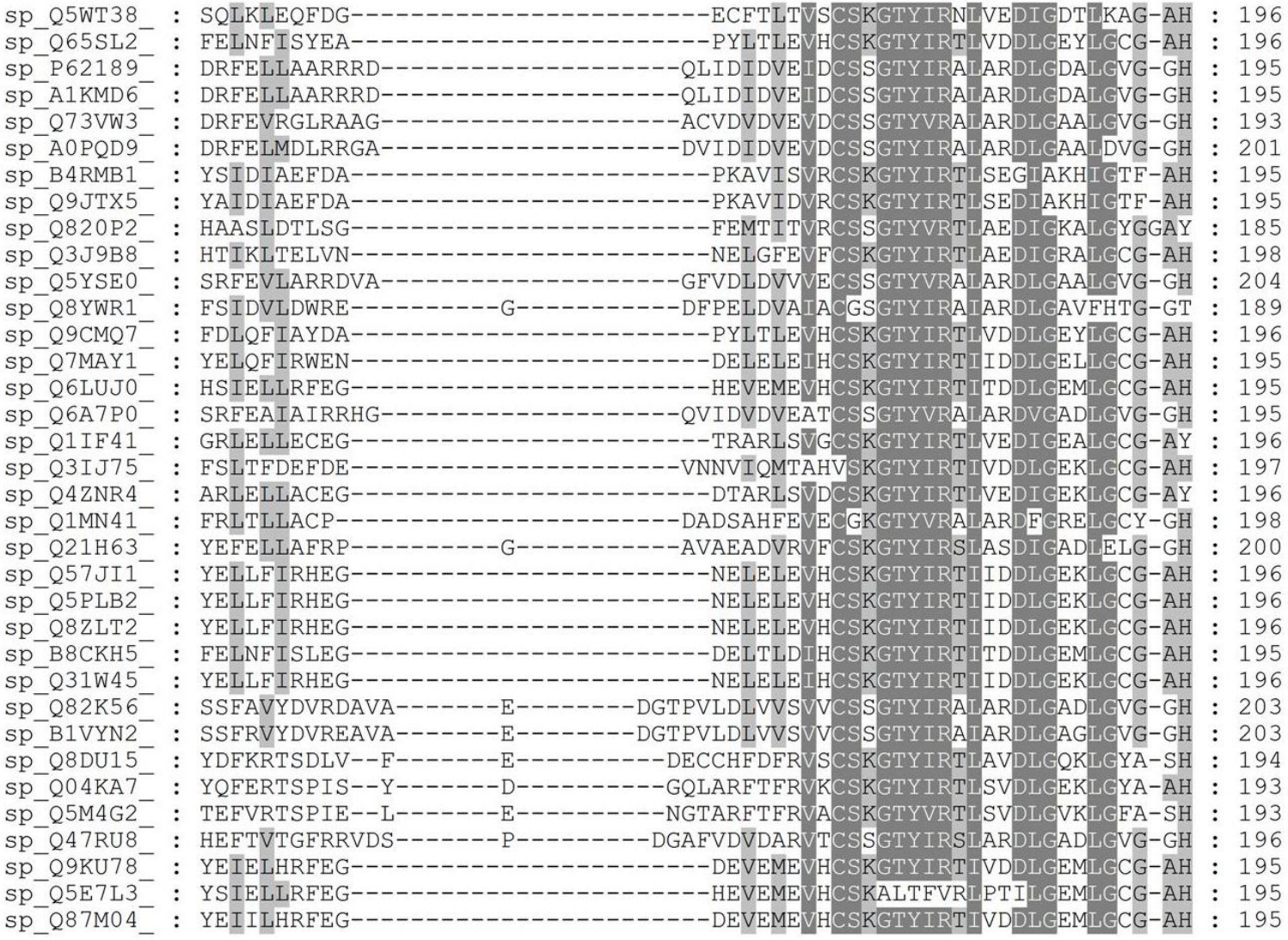

sp_Q47RU8_ : RLAGTRRVGHAGTLDPMATGVLVVGVGK-ATRLLGYLALTEKVYEATIRLGQSTTTDDAEGELLER : 89
sp_Q9KU78_ : RLYGAEKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDSDKRYRVIAKLGERTDTSDSDGQVVQT : 97
sp_Q5E7L3_ : RIYFAEKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDSDKRYRVVAKLGERTNTSDSDGEVVQT : 97
sp_Q87M04_ : RIYFAEKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDSDKRYRVIAKLGERTNTSDSDGEVVET : 97
sp_A7Z4T7_ : KLLKTKKVGHTGTLDPEVSGVLPICVGR-ATKIVEYVTDKSKTYDAEITLGFSTSTEDQTGETVSV : 89
sp_Q9KA80_ : RLLKTKKVGHTGTLDPDVYGVLPVCIGH-ATKVAQYMSDYPKAYEGEVTVGFSTTTEDRSGDTVET : 90
sp_Q0SM48_ : KYFNTNHVGHAGILDKFASGILIALVGK-YTKLANYFMSLDKEYVAEFRFGLETDTLDPNGRIVNK : 89
sp_Q5HU02_ : KKYKNKKAGYSGTLDPFAKGVLIVAFGQ-YTKLFRFLKKTPKTYKATLWLGVYSLSLDDQNIKEI- : 87
sp_P58063_ : RAFNAQKGGHAGTLDPLATGILPIALGE-ATKTVPFLMDADKAYRFTIAWGRDTTTLDREGETTGT : 97
sp_B0BB81_ : RLIGVKKIGHAGTLDPFATGVMVMLIGRKFTRLSDIMLFEDKEYAAVAHLGTTTDTYDCDGKIVGR : 96
sp_A5N845_ : FMLKNEKVGHTGTLDPMASGVLPICVGR-ATKFADYMVESKKIYLAELRLGITTETYDREGSVVNT : 88
sp_Q0TPS0_ : RVAKMKKVGHTGTLDPAASGVLPVCLGK-ATKIIDYIMENKKVYRVNLKLGMVTDTYDLEGEVLRE : 88
sp_Q895J5_ : KISRIKKVGHTGTLDPMATGVLPICLGG-STKIVDFIMNEHKEYRAKLKLGLITDTYDREGKVLKE : 88
sp_Q3Z7U5_ : HIYSQKKVGHGGMLDPSATGVIPVFLGS-ATRLIEYLSSVRKTYLAEIELGTETDSYDSEGEITSR : 88
sp_Q3YSC6_ : RILNVKKAGHAGTLDPLASGVLPIAIGE-ATKVMPYAVDVIKSYLFTVQWGEQRTTDDAEGEIVDK : 88
sp_Q7U330_ : KHFGMSKAGYLGTLDPFAKGVLVVGFGS-YTRLFPHLQKVPKAYRATLWLGAKSASLDIEHIESI- : 89
sp_Q03QT7_ : HILQTKKVGHSGTLDPSVDGVLPICIGS-ATKVVPYLMASGKVYRGSVTLGLATTTEDLDGDVVER : 88
sp_Q038M8_ : KILGIKKIGHSGTLDPNVDGVLPLAIGA-GTKAVPQLMASGKVYTGEITLGFATTTEDLDGEVVDK : 88
sp_Q92C26_ : KILHTKKVGHTGTLDPEVEGVLPICIGR-ATKLAEYVTDEGKVYVAEITLGKSTTTEDATGETVAT : 88
sp_B0JJJ5_ : KILKTKRVGHGGTLDPMATGVLPIAVGA-ATRLLAYLP-ENKAYRAKIQLGLSTDTDDITGKAIAT : 87
sp_Q98Q19_ : WQNNIKKIGHSGTLDPEATGLLLLASDE-DTKLLDYVDKKFKSYRATMILGLQSQSFDSQGKIINS : 86
sp_Q04ED7_ : KIFQQKQVGHTGTLDPSVTGLLVIVLGK-ATKLIDYLQENQKQYRGTLILGLKTDTQDMDGQVTEM : 89
sp_A5D2S3_ : RIFGVKKAGHTGTLDPGAAGVLVVCLGV-ATRLARFLLGEDKEYRVEITFGMSTSTGDSYGEITDQ : 88
sp_Q1RIG2_ : KVLGKVKIGHSGTLDVEAEGVLPLAIGE-ATKLVQMLIDAKKTYIFTVKFGKQTDSGDYAGKVIAI : 89
sp_A7X1Q5_ : KILKTKKIGHTGTLDPEVAGVLPVCIGN-ATRVSDYVMDMGKAYEATVSIGRSTTTEDQTGDTLET : 89
sp_Q8NWZ0_ : KILKTKKIGHTGTLDPEVAGVLPVCIGN-ATRVSDYVMDMGKAYEATVSIGRSTTTEDQTGDTLET : 89
sp_P65856_ : KILHTKKIGHGGTLDPDVVGVLPIAVGK-ATRVIEYMTESGKIYEGEITLGYATSTEDSSGEVISR : 89
sp_Q8CWM3_ : RRLRLKRIGHGGTLDPMATGVLPIALGA-ATRLLPYLS-DRKAYIGTVRFGMSTTTDDITGEICQE : 87
sp_O83859_ : RLLGVKKVGHTGTLDRFADGLLLLLVGG-FTKLAPVMTRLEKSYEARIQFGVQTDTLDPEGAVVRC : 94
sp_P45142_ : RLFQANKAGHTGALDPLATGMLPICLGE-ATKFSQFLLDADKRYLVTAKLGERTDTSDAEGQVVET : 98
Gh G LDp a G66 g a 4 K Y g d g
140 * 160 * 180 * 2
sp_P62190_ : --VPA---KH--LTIEAIDAAMERLRGEIRQVPSSVSAIKVGGRRAYRLARQGRS-VQLEARPIRI : 150
sp_P60340_ : --RPV----T--FSAEQLAAALDTFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155

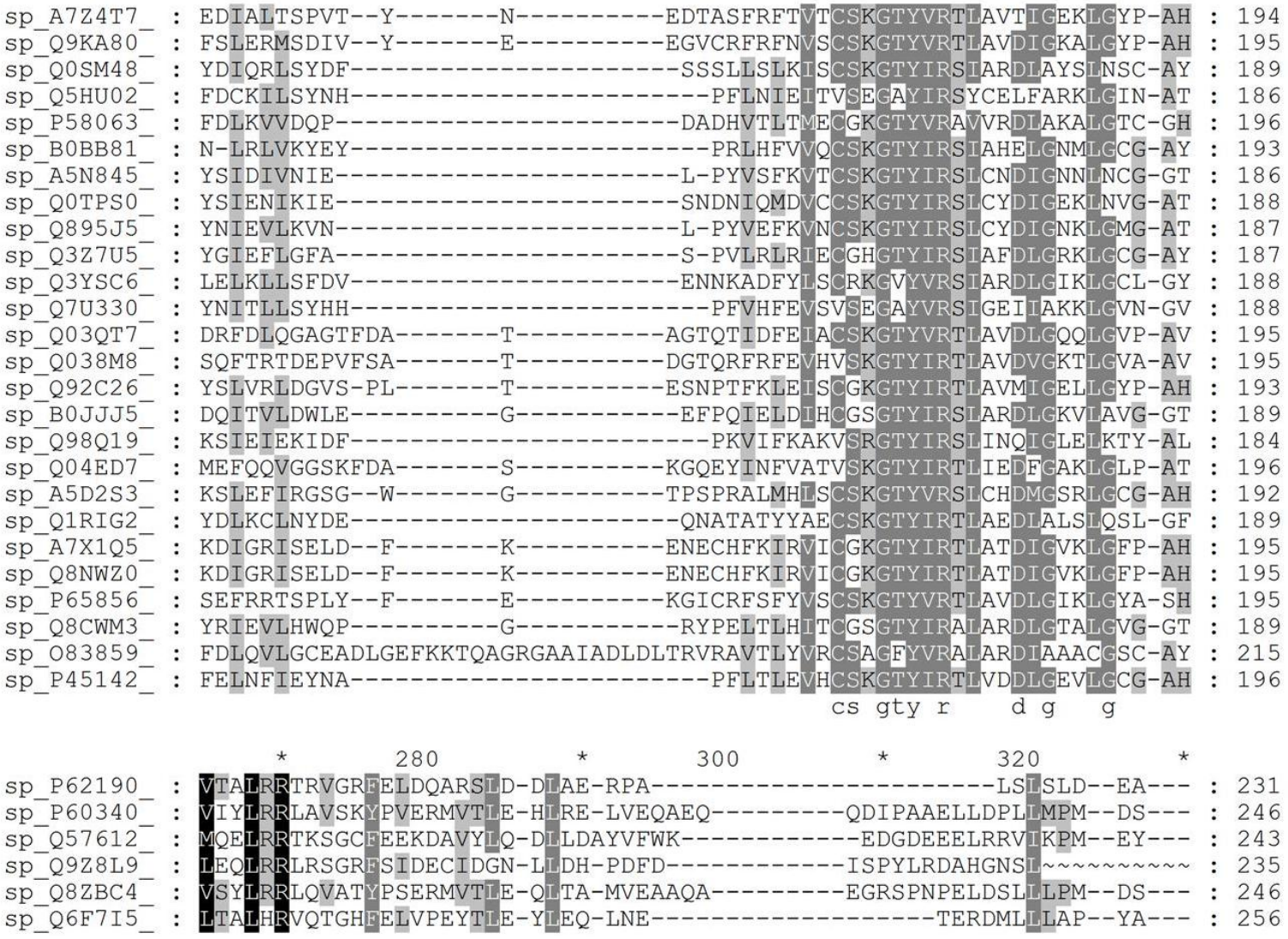

sp_Q57612_ : ----------ASEEDILRVFKEFTGRIYQRPPLKAAVKR----------RLRIRKI : 154
sp_Q9Z8L9_ : --SKK----I--PSLEEVLSAAEYFQGEIQQLPPMFSAKKVQGKKLYEYARKGLS-IERHHSTVQV : 153
sp_Q8ZBC4_ : --REV----N--LTQAQIDTALESFRGESQQIPSMYSALKHQGKPLYEYARQGIE-VEREARSITV : 155
sp_Q6F7I5_ : --RSV---PP--LTRDMLELVLDGFRGDIQQVPPMYSALKRDGRPLYELARQGIE-IEREARPVTI : 175
sp_A3N007_ : --RAV----N--VGEAEIITALEQFRGDILQVPTMFSALKHNGKPLYEYARQGIT-VEREARPITI : 155
sp_A4SJR7_ : --RPV----N--VALGTLIESLDQFRGPIMQVPSMYSALKHNGRPLYEYAREGIE-IEREARPITV : 156
sp_O66922_ : --YEG----E--LNCDKLKEVLNEFRGEILQTPPPFSAKKIKGRRAYELARKGKK-VELKPVKITV : 145
sp_A1R520_ : --RTA---AH--ITDDAVAVGVAALTGPIQQVPSSVSAIKVNGERSYARVRSGEE-VKLAARPVTI : 147
sp_A1K7B7_ : --HPV---S---VTREALEEVLSRFRGEIDQLPPMYSALKRDGKPLYEYARAGIE-IEREVRRVTI : 154
sp_Q6G5F5_ : --SSK----R--PTREEILALLPQYTGVILQTPPQFSAIKIAGNRAYDLAREGEV-VEIPPRQVEI : 156
sp_Q1LSL0_ : --RPV----T--INQAMIEHIMEQFYGDIYQIPPMFSSIKYQGRALYKYARKGIN-IPRSARLVHI : 156
sp_Q8YEB5_ : --SDK----R--PSREEVEALLPDYTGVISQVPPQFSAIKIDGERAYDLAREGET-VEIPAREVEI : 156
sp_P59876_ : --SDK----R--PSREEVEALLPDYTGVISQVPPQFSAIKIDGERAYDLAREGET-VEIPAREVEI : 156
sp_P57456_ : --RPI----L--INSFKIKSALKELTGLIEQIPPMYSAIKHNGVPLYKYARQGLN-IKRSIRKVLI : 153
sp_Q8K9H3_ : --RPI----L--FTPIQLSLALKALTGSVNQIPSMYSAVKYKGIPLYKYARQGIN-VKRNIRCVFI : 153
sp_Q89AF6_ : --RPI----T--FTNLQLQKVLKSFHGKIKQIPSMYSAIKYHGHALYKYARQGIV-ISRKVRDAII : 152
sp_Q62KL1_ : --RDV---S---CDEAAVRAALARFVGEIVQVPPMYSALKRDGKPLYEYARAGQT-VEREGRTVTI : 159
sp_Q482T7_ : --KTV----D--VSSEQLAKALDSFRGTTKQVPSMYSALKHQGQPLYKYAREGIE-VPREARDITV : 154
sp_C3PH13_ : --EPA---TA--VEDSAIAREIAALTGDIMQRPAAVSAIKVDGKRAHERVRAGEE-VELPARPVTV : 152
sp_P60343_ : --ASPDQLAG--ITETKISDAVKQFRGSIMQRPAAVSAIKIDGKRAHQRVREGEK-VEIPARPVTI : 155
sp_Q8FPB3_ : --ADA---ST--LDDEQITTAVTSLTGEIMQKPASVSAIKIDGKRAHERVRDGEV-VDIPARPVTV : 152
sp_A4QEY6_ : --TDA---SG--LDHRAILAEIANLTGDIMQKPTKVSAIKIDGKRAHERVRDGEE-VDIPARPVTV : 152
sp_Q4JV56_ : --ASPQDLQA--LTEQQVREAFAAQRGDIMQRPTSVSSIKIDGKRAHELVREGHD-VVLPERPVTI : 160
sp_A9KBM3_ : --RPI---PK--LTKRALEKTLSAFRGVIDQTPSMYSALKHKGQPLYKLARQGIE-VERKTRQVTI : 159
sp_Q4AAX7_ : --RPI---PK--LTKRALEKTLFAFRGVIDQTPSMYSALKHKGQPLYKLARQGIE-VERKTRQVTI : 159
sp_Q0TCU3_ : --RPV----T--FSAEQLAAALDTFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155
sp_P60341_ : --RPV----T--FSAEQLAAALDTFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155
sp_Q7MBB8_ : --AGA---AD--LDLGRVQVHLQAFRGTISQVPPRYSAIHQEGERLYDLARRGVAIAEIAPRTVEI : 146
sp_Q4QK43_ : --REV----H--VETPQILTALEQFRGDILQVPTMFSALKHNGKPLYEYARQGIT-VEREARPITI : 155
sp_A5UBU2_ : --REV----H--VETPQILTALEQFRGDILQVPTMFSALKHNGKPLYEYARQGIT-VEREARPITI : 155
sp_A5UF28_ : --REV----H--VETPQILTALEQFRGDILQVPTMFSALKHNGKPLYEYARQGIT-VEREARPITI : 155
sp_Q0I3P3_ : --REV----N--IDLAKILTALEQFRGEIMQVPTMFSALKYQGKALYEYARAGIT-IEREARPISI : 155
sp_Q2SML1_ : --SPV---PG--LSVSDLESVLGKFRGNLKQIPSMYSALKHKGTPLYKLARQGIE-VEREARDIII : 155
sp_C4K3E8_ : --REI----T--FTQPQLEEALEKFRGPILQIPSMYSALKHQGRPLYEYARQGVT-LDREARRITV : 156
sp_Q5QTZ0_ : --KPV----A--VDTAKVADAIKQFIGEQDQSPSIYSALKHEGRPLYYYARQGIE-VPKKTRTITV : 156

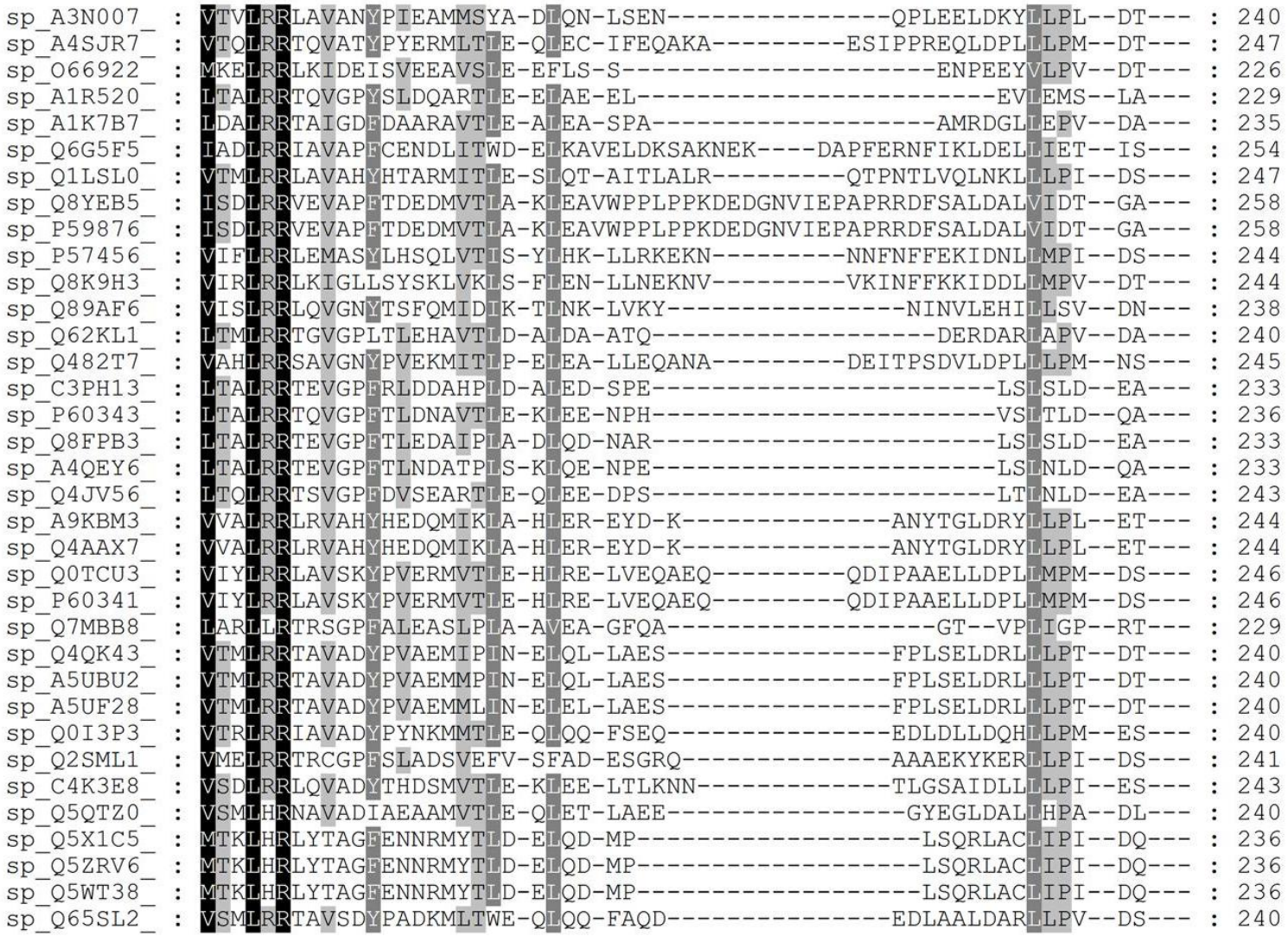

sp_Q5X1C5_ : I-ENY----T--VSHEEMIATLEKYKGKIKQIPSMFSALKHKGTPLYRLAREGIE-IERKARDIVI : 155
sp_Q5ZRV6_ : I-ENY----T--VSHEEMIATLEKYKGKIKQIPSMFSALKHKGTPLYRLAREGIE-IERKARDIVI : 155
sp_Q5WT38_ : I-ENY----T--VSHEELIATLEKYKGKTKQIPSMFSALKHKGTPLYRLAREGIE-IERKARDIVI : 155
sp_Q65SL2_ : --RSV----N--VTEQKILDSLPHFRGDIMQVPTMFSALKHKGKPLYEYARAGIV-VEREARPISI : 155
sp_P62189_ : --VPA---KH--LTIEAIDAAMERLRGEIRQVPSSVSAIKVGGRRAYRLARQGRS-VQLEARPIRI : 150
sp_A1KMD6_ : --VPA---KH--LTIEAIDAAMERLRGEIRQVPSSVSAIKVGGRRAYRLARQGRS-VQLEARPIRI : 150
sp_Q73VW3_ : --VDA---RH--LTSQAIEAAVGGLRGDIHQVPSTVSAIKVAGKRAYKLVREGQA-VELPARPVRI : 148
sp_A0PQD9_ : --VSA---EH--VAPEAIAAAILDLTGDIRQVPSAVSAIKVDGRRAYQLAREGQT-VELAARPVRI : 156
sp_B4RMB1_ : --ARA---D---ISLAEFQTACQALTGNIRQVPPMFSALKHEGKPLYEYARKGIV-IERKPRDITV : 154
sp_Q9JTX5_ : --ARA---D---ISLAEFQTACQALTGNIRQVPPMFSALKHEGKPLYEYARKGIV-IERKARDITV : 154
sp_Q820P2_ : VGSDV---NP--PDFGQVTGILQTFLGRSSQIPPMFSALKQHGKPLYRYAREGIT-VERKAREIVI : 143
sp_Q3J9B8_ : --HPV---NE--LDRDEVAKFLSGFSGPQEQVPPMYSAIKHQGQRLYKLARQGIE-VERKSRQVTI : 157
sp_Q5YSE0_ : --VPA---GH--LGDAEVAAGVAALTGDIQQVPATVSAIKIGGERAYARHRAGEQ-VELAARPVTV : 159
sp_Q8YWR1_ : --QPC---RG--LSLSEVKTALPQFIGKIEQIPPIYSAIQVEGKRLYDLARKGEI-IEVPARTVEV : 145
sp_Q9CMQ7_ : --RDV----Q--VDVQDILAALPHFRGNLMQVPTMFSALKHQGKPLYEYARAGIT-VEREARPITI : 155
sp_Q7MAY1_ : --RAI----Q--LSQVQLDAALDKFRGDTMQIPSMYSALKYQGKPLYEYARQGIE-VEREARPITV : 154
sp_Q6LUJ0_ : --REV----K--VDRGQLERCIAKFRGTTDQIPSMFSALKYQGRPLYEYAREGIE-IPRESRKITV : 154
sp_Q6A7P0_ : --TDA---SA--IDDRAIHAAMVRQTGEIQQVPAAVSAIKVNGRRAYAKVRAGED-VVLRPRAVTV : 150
sp_Q1IF41_ : --REV----T--VGRADIEAVIPRFRGDILQVPPMYSALKRDGQPLYKLARAGEV-VEREARSVTI : 155
sp_Q3IJ75_ : --KDV----N--VTREQLIKEIAAFVGESDQYPSMYSALKYQGRPLYKYAREGIE-VPRKCRKINV : 154
sp_Q4ZNR4_ : --RPV----T--VGRDDIEAALPHFRGQISQIPPMYSALKRDGQPLYKLARAGEV-VEREPRSVTI : 155
sp_Q1MN41_ : --SDK----R--PSEQQIRDILPGYIGTISQVPPQFSAIKIAGERAYDLAREGET-IEIPSREVDI : 156
sp_Q21H63_ : --IDA---SK--LTKADVLKAIKAYIGDIDQVPPMYSALKRNGQPLYKLARQGIE-VEREPRPVTV : 156
sp_Q57JI1_ : --RPV----T--FSAEQLASALETFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155
sp_Q5PLB2_ : --RPV----T--FSAEQLASALETFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155
sp_Q8ZLT2_ : --RPV----T--FSAEQLASALETFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITV : 155
sp_B8CKH5_ : --REI----D--FTQEQLDTALEYFRGKTMQVPSMYSALKYQGQPLYKYAREGIE-VPREARPINV : 154
sp_Q31W45_ : --RPV----T--FSAEQLAAALDTFRGDIEQIPSMYSALKYQGKKLYEYARQGIE-VPREARPITI : 155
sp_Q82K56_ : --TDA---SR--VTRDAIDAGVAKLSGAIMQVPSKVSAIKIDGVRSYKRAREGED-FEIPARPVTV : 153
sp_B1VYN2_ : --TDA---SG--VTRASIDAGVAALTGRIMQVPSKVSAIKIDGKRSYARVRGGEE-FEIPARPVTI : 153
sp_Q8DU15_ : --TPV---DE-TLSVELIDQAMTCFMGQITQVPPMYSAVKVNGKKLYEYARAGQE-VERPQRQVRI : 148
sp_Q04KA7_ : --TPV---LS-SLDEKLVDEAIASLTGPITQIPPMYSAVKVNGRKLYEYARAGQE-VERPERQVII : 147
sp_Q5M4G2_ : --TPI---TE--LDGATVDQAMASFEGEITQIPPMYSAVKINGKKLYEYARAGEE-VERPQRQVKI : 147
sp_Q47RU8_ : --RPA---DH--IDEAAVHAGTRALTGVIHQVPPQVSAVKVRGQRAYRRARAGET-VELKAREVTV : 147
sp_Q9KU78_ : --RPV----H--VDYDTLLACIAKFRGETDQVPSMFSALKYQGRPLYEYARQGIE-VPREARKITV : 154

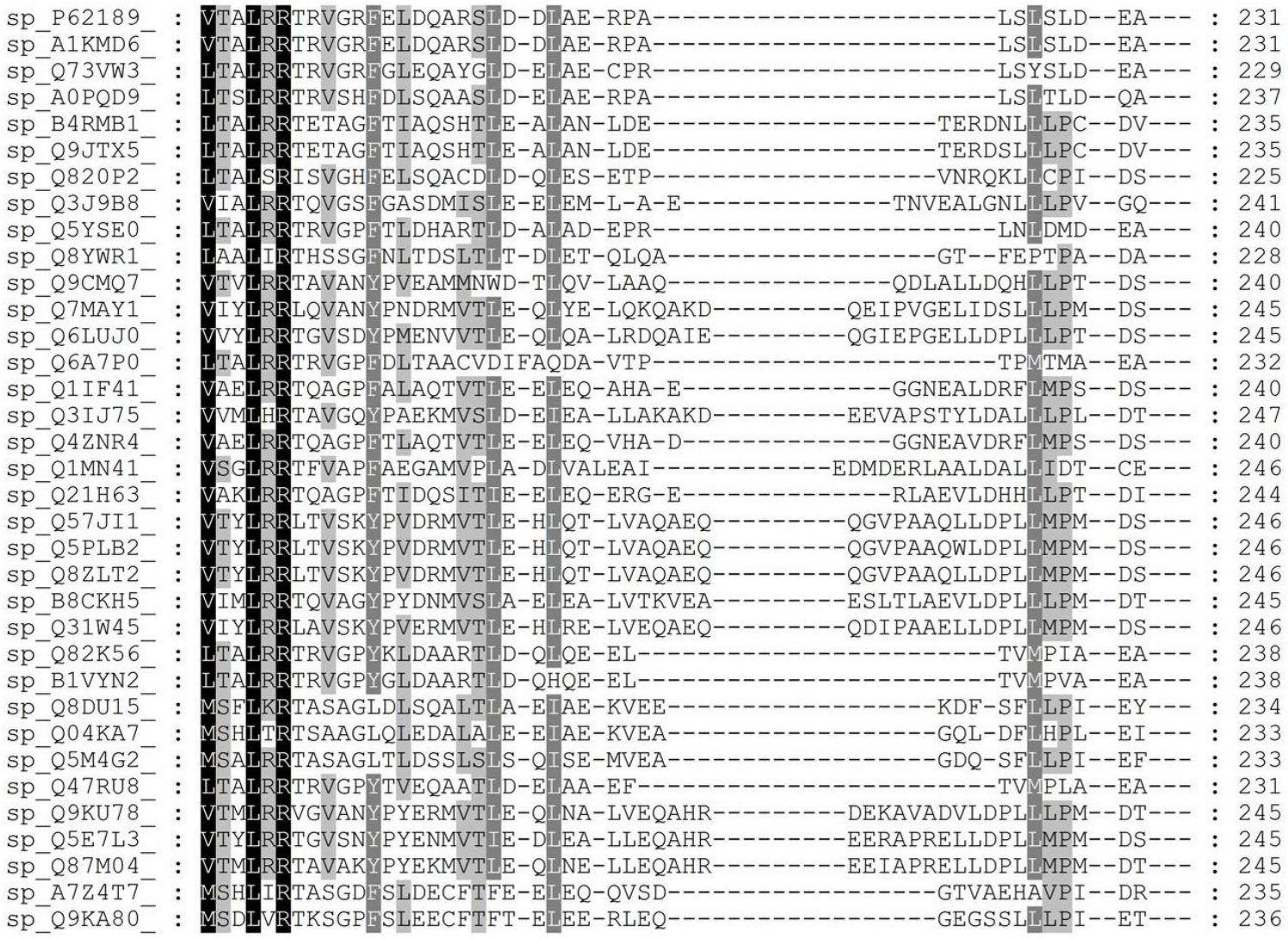

sp_Q5E7L3_ : --REV----K--VDRGQLERCIAKFRGTTDQIPSMFSALKYQGRPLYEYAREGIE-VPRESRKITV : 154
sp_Q87M04_ : --RPV----D--VTLEKLEACIEKFRGESDQVPSMFSALKYQGKPLYEYARKGIE-VPRESRKITV : 154
sp_A7Z4T7_ : --KPV---KE-PLKEADIKAVLDELKGPQEQVPPMYSAVKVNGKKLYEYARAGIE-VERPKRNITI : 148
sp_Q9KA80_ : --KTI---QQ-PFVEAVVDQVLATFVGEIKQIPPMYSAVKVRGKRLYEYARAGIT-VERPERTVTI : 149
sp_Q0SM48_ : --TDY----I--PNLEDIDLKLKDFVGEIYQSPPRFSSVHINGSRAYKLALNGKF-FEIKKRRVNV : 146
sp_Q5HU02_ : -----QNIKE--FDLANLKQIIDQMQGIISYTPPQFSAKRINGTRAYELAKKGIE-ANLKPCQMEV : 145
sp_P58063_ : --SDV----R--PTREQVEAALPAFIGEVDQIPPNFSAIKVDGERAYDLARDGVE-FELPTRKVSI : 154
sp_B0BB81_ : --SKK----V--PTMDEVLTCTSYFQGEIQQVPPMFSAKKVQGKKLYEYARQGLS-IERRFATVTV : 153
sp_A5N845_ : --RGV----Y--LKKKDIIEEILSFQGEIEQVPPMYSALKVNGRRLYELARKGIE-IERKKRKITI : 145
sp_Q0TPS0_ : --EDA---SH--ITKDEILNCINSFLGTIDQVPPMYSALKQNGVRLYELARQGIE-VHREARKITI : 146
sp_Q895J5_ : --EDA---SK--ILEEEVVNCINSFKGEIIQIPPMYSAIKIKGERLYNLARKGIE-VEREGRKINI : 146
sp_Q3Z7U5_ : --KSC---EH--ITADMVRNALPDFLGEITQIPPMYSAVKHRGVRLYNLARQGIE-VERNPRKAAI : 146
sp_Q3YSC6_ : --SDM----I--PCVENIKKIIPDFIGLLKQVPPSFSAVHVNGVRAFELARSGQD-VNLSSRFVDV : 145
sp_Q7U330_ : -----EIIPE--YNQSDIEKILFSLKGTFDYTPPAFSAKHINGQRAYKLAREGKV-FTLQQIQMSI : 147
sp_Q03QT7_ : --QPL---ER-PFTADQVAAAAQALTGTIQQTPPMYSAVKVNGRKLYEYARAGET-VERPTRTITV : 147
sp_Q038M8_ : --TPL---TQ-PFTADQLDAALTAWTGNITQIPPMFSAVKVNGRRLYEYARAGET-VKRPERQATV : 147
sp_Q92C26_ : --KEL---AE--ISAEELQAALTKLTGKITQIPPMFSAVKVNGKKLYEYARAGIE-VERPSRQVDI : 146
sp_B0JJJ5_ : --CPW---PD--LTLEAVKPHLAEFIGNIAQIPPMYSAIHKDGRRLYELARKGEI-IAVEPRQVKI : 145
sp_Q98Q19_ : --SNL----K--VDNLTIEKTIKNFVGPFVQIPPIFSAKKINGKRAYEYARQGSE-ISMKAQEVFV : 143
sp_Q04ED7_ : --QFL---KE-AIADFKKKAAFDSFLGSSDQLPPMYSAVKVGGKHLYELARKGET-IERKSRKIKV : 148
sp_A5D2S3_ : --RDA---SF--LKEHDIIRVLPEFTGEVRQVPPMTSAIKWRGKKLYELAREGLV-VERQERAVYI : 146
sp_Q1RIG2_ : --TDY----I--PSKENAYAICSKFIGTITQIPPAFSALKVNGVRAYKLARDGKE-VELKPRNITI : 146
sp_A7X1Q5_ : --KGV---HSADFNKDDIDRLLENFKGVIEQIPPMYSSVKVNGKKLYEYARNNET-VERPKRKVNI : 149
sp_Q8NWZ0_ : --KGV---HSADFNKDDIDRLLESFKGIIEQIPPMYSSVKVNGKKLYEYARNNET-VERPKRKVNI : 149
sp_P65856_ : --TPL---TQSDLSEDVVDHAMKSFTGPITQVPPMYSAVKVNGKKLYEYARSGEE-VERPKRQITI : 149
sp_Q8CWM3_ : --QGA---SH--LTLAAIQEQLPQFIGEIEQLPPAYSAIQVAGQRLYARARAGEV-VSVPPRIVTV : 145
sp_O83859_ : --SLF-------PTFARVRAALPHFTGSIDQVPPEYSALKFGGVRASDRVRRGEA-VCMKARRVFV : 150
sp_P45142_ : --REV----N--LETQQILTALEQFRGDILQVPTMFSALKHNGKPLYEYARQGIT-VEREARPITI : 155
G q P sa g a g r 6
00 * 220 * 240 * 260
sp_P62190_ : DRFELLAARRRD----------QLIDIDVEIDCSSGTYIRALARDLGDALGVG-GH : 195
sp_P60340_ : YELLFIRHEG----------NELELEIHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_Q57612_ : HELELLDKDG----------KDVLFRVKCQSGTYIRKLCEDIGEALGTS-AH : 195
sp_Q9Z8L9_ : H-LQITKYEY----------PLLHFVVSCSKGTYIRSIAHELGTMLGCG-AY : 193

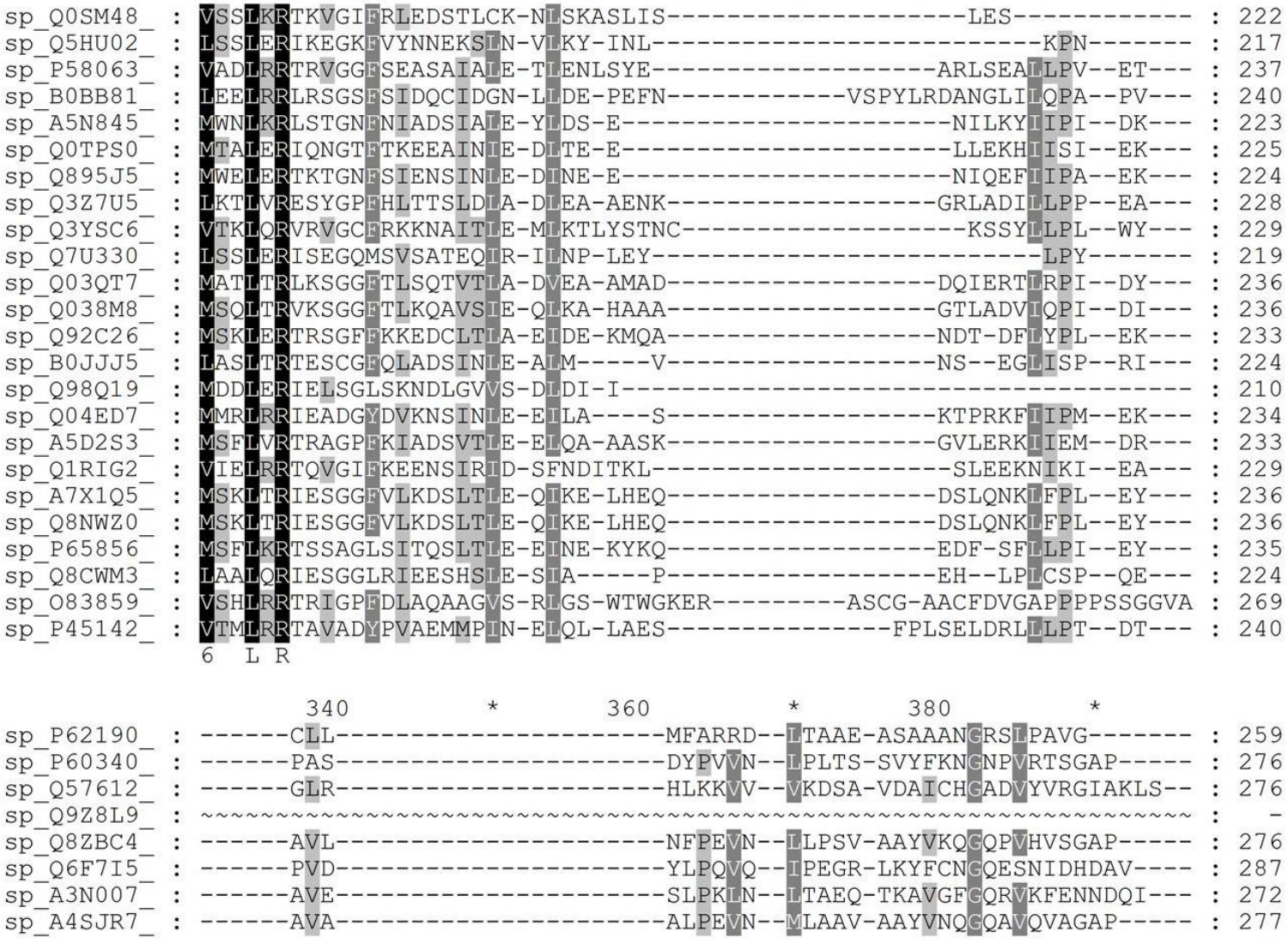

sp_Q8ZBC4_ : YELLFIRWEG----------NDLELEIHCSKGTYIRTIIDDLGELLGCG-AH : 196
sp_Q6F7I5_ : HALKLVDFDT----------ESLTLDVTCSKGTYIRVLGEDIGEALGCG-GH : 216
sp_A3N007_ : FELRFIEYAA----------PYLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
sp_A4SJR7_ : FELKLISFEG----------DEVKLEVHCSKGTYIRSLVDDLGEVLGCG-AH : 197
sp_O66922_ : YSLELLSCNP----------REKEAEFLAEISSGGYVRSLAYDIGKKLGIG-GY : 188
sp_A1R520_ : HRFDVHSITRID----------GGRVVDVDVTVECSSGTYIRALARDLGNALGIG-GH : 194
sp_A1K7B7_ : HDLELIAFSG----------EHFSMRVRCSKGTYIRTLAMDIGAALGCG-AY : 195
sp_Q6G5F5_ : ETFKLIEMI----------TKDHSIFEITCGKGTYVRSLARDMGRDLGCY-AH : 198
sp_Q1LSL0_ : YNLQILDWDN----------THIELQIHCSKGTYIRTIIDDIGELLGCG-AH : 197
sp_Q8YEB5_ : DRLEIVGFP----------DADRTEFEVECSKGTYVRSLARDMGRDLGCY-GH : 198
sp_P59876_ : DRLEIVGFP----------DADRTEFEVECSKGTYVRSLARDMGRDLGCY-GH : 198
sp_P57456_ : HDISSIHQEK----------NLIEFKIFCSKGTYVRTLVEDLGEKLGCG-AH : 194
sp_Q8K9H3_ : YKIDLVDQKD----------NWIELNVHCSKGTYIRTLIEDLGEKLFCG-AH : 194
sp_Q89AF6_ : YELKVLGYSY----------KHKYLELDIICSKGTYIRTLIEDVGEKLNCG-AH : 195
sp_Q62KL1_ : RALALVSCAL----------PDVTFRVTCSKGTYVRTLAEDIGEALGCG-AH : 200
sp_Q482T7_ : FNLELLRFEH----------DEVELNIHVSKGTYIRTIVDDLGELLGCG-AH : 195
sp_C3PH13_ : SRFDILATRREG----------EFIDLDVSVACSSGTYIRSLARDLGEALGVG-GH : 197
sp_P60343_ : SRYDILEIRRDA----------AFIDIDVEVDCSSGTYIRSLARDLGEELGVG-GH : 200
sp_Q8FPB3_ : SVFDVLEQRREG----------GFVDLDVRVHCSSGTYIRSLARDLGAALGVG-GH : 197
sp_A4QEY6_ : SVFDVLDYRVDS----------EFYDLDVRVHCSSGTYIRALARDLGNALQVG-GH : 197
sp_Q4JV56_ : FSLEVLDVVVDD----------ATSCIDARISVHCSSGTYIRAIARDVGEALGVG-GH : 207
sp_A9KBM3_ : YELTLLDWDN----------ESIELYVHCSKGTYIRTLLDDVGEALGCG-AH : 200
sp_Q4AAX7_ : YELTLLDWDN----------ESIELYVHCSKGTYIRTLLDDVGEALGCG-AH : 200
sp_Q0TCU3_ : YELLFIRHEG----------NELELEIHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_P60341_ : YELLFIRHEG----------NELELEIHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_Q7MBB8_ : QALTVLDWYP----------G----------HYPELQIEIACSTGTYIRSIARDLGTALGTG-AT : 190
sp_Q4QK43_ : FELNFIEYHA----------PFLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
sp_A5UBU2_ : FELNFIEYHA----------PFLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
sp_A5UF28_ : FELNFIEYHA----------PFLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
sp_Q0I3P3_ : FELKFIDYQI----------PYLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
sp_Q2SML1_ : HELELLDYTP----------PFLRLRVLCSKGTYIRNLVEDIGQVLGCG-AH : 196
sp_C4K3E8_ : FDLQCLRWEG----------HELELEIHCSKGTYIRTIVDDLGEVLGCG-AY : 197
sp_Q5QTZ0_ : HSIELLNIQD----------NKVTLQVSCSKGTYIRTLVDDLGQLLGCG-AH : 197
sp_Q5X1C5_ : SQLKLEQFDG----------ECFTLTVSCSKGTYIRNLVEDIGDTLKAG-AH : 196
sp_Q5ZRV6_ : SQLNLEQFDG----------ECFTLTVSCSKGTYIRNLVEDIGDTLKAG-AH : 196

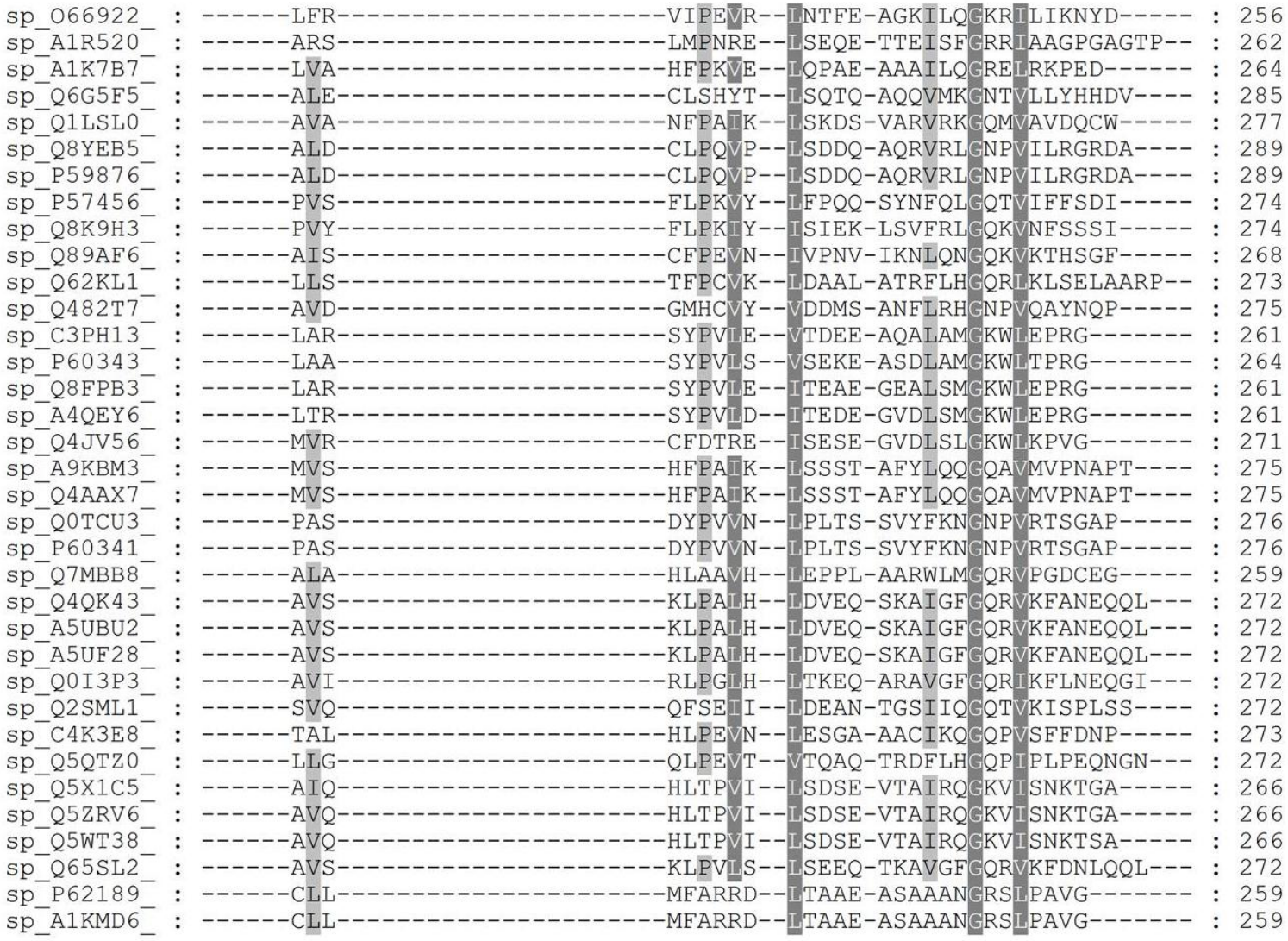

sp_Q5WT38_ : SQLKLEQFDG----------ECFTLTVSCSKGTYIRNLVEDIGDTLKAG-AH : 196
sp_Q65SL2_ : FELNFISYEA----------PYLTLEVHCSKGTYIRTLVDDLGEYLGCG-AH : 196
sp_P62189_ : DRFELLAARRRD----------QLIDIDVEIDCSSGTYIRALARDLGDALGVG-GH : 195
sp_A1KMD6_ : DRFELLAARRRD----------QLIDIDVEIDCSSGTYIRALARDLGDALGVG-GH : 195
sp_Q73VW3_ : DRFEVRGLRAAG----------ACVDVDVEVDCSSGTYVRALARDLGAALGVG-GH : 193
sp_A0PQD9_ : DRFELMDLRRGA----------DVIDIDVEVDCSSGTYIRALARDLGAALDVG-GH : 201
sp_B4RMB1_ : YSIDIAEFDA----------PKAVISVRCSKGTYIRTLSEGIAKHIGTF-AH : 195
sp_Q9JTX5_ : YAIDIAEFDA----------PKAVIDVRCSKGTYIRTLSEDIAKHIGTF-AH : 195
sp_Q820P2_ : HAASLDTLSG----------FEMTITVRCSSGTYVRTLAEDIGKALGYGGAY : 185
sp_Q3J9B8_ : HTIKLTELVN----------NELGFEVFCSKGTYIRTLAEDIGRALGCG-AH : 198
sp_Q5YSE0_ : SRFEVLARRDVA----------GFVDLDVVVECSSGTYVRALARDLGAALGVG-GH : 204
sp_Q8YWR1_ : FSIDVLDWRE----------G----------DFPELDVAIACGSGTYIRAIARDLGAVFHTG-GT : 189
sp_Q9CMQ7_ : FDLQFIAYDA----------PYLTLEVHCSKGTYIRTLVDDLGEYLGCG-AH : 196
sp_Q7MAY1_ : YELQFIRWEN----------DELELEIHCSKGTYIRTIIDDLGELLGCG-AH : 195
sp_Q6LUJ0_ : HSIELLRFEG----------HEVEMEVHCSKGTYIRTITDDLGEMLGCG-AH : 195
sp_Q6A7P0_ : SRFEAIAIRRHG----------QVIDVDVEATCSSGTYVRALARDVGADLGVG-GH : 195
sp_Q1IF41_ : GRLELLECEG----------TRARLSVGCSKGTYIRTLVEDIGEALGCG-AY : 196
sp_Q3IJ75_ : FSLTFDEFDE----------VNNVIQMTAHVSKGTYIRTIVDDLGEKLGCG-AH : 197
sp_Q4ZNR4_ : ARLELLACEG----------DTARLSVDCSKGTYIRTLVEDIGEKLGCG-AY : 196
sp_Q1MN41_ : FRLTLLACP----------DADSAHFEVECGKGTYVRALARDFGRELGCY-GH : 198
sp_Q21H63_ : YEFELLAFRP----------G----------AVAEADVRVFCSKGTYIRSLASDIGADLELG-GH : 200
sp_Q57JI1_ : YELLFIRHEG----------NELELEVHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_Q5PLB2_ : YELLFIRHEG----------NELELEVHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_Q8ZLT2_ : YELLFIRHEG----------NELELEVHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_B8CKH5_ : FELNFISLEG----------DELTLDIHCSKGTYIRTITDDLGEMLGCG-AH : 195
sp_Q31W45_ : YELLFIRHEG----------NELELEIHCSKGTYIRTIIDDLGEKLGCG-AH : 196
sp_Q82K56_ : SSFAVYDVRDAVA-------E--------DGTPVLDLVVSVVCSSGTYIRALARDLGADLGVG-GH : 203
sp_B1VYN2_ : SSFRVYDVREAVA-------E--------DGTPVLDLVVSVVCSSGTYIRAIARDLGAGLGVG-GH : 203
sp_Q8DU15_ : YDFKRTSDLV--F-------E----------DECCHFDFRVSCSKGTYIRTLAVDLGQKLGYA-SH : 194
sp_Q04KA7_ : YQFERTSPIS--Y-------D----------GQLARFTFRVKCSKGTYIRTLSVDLGEKLGYA-AH : 193
sp_Q5M4G2_ : TEFVRTSPIE--L-------E----------NGTARFTFRVACSKGTYVRTLSVDLGVKLGFA-SH : 193
sp_Q47RU8_ : HEFTVTGFRRVDS-------P---------DGAFVDVDARVTCSSGTYIRSLARDLGADLGVG-GH : 196
sp_Q9KU78_ : YEIELHRFEG----------DEVEMEVHCSKGTYIRTIVDDLGEMLGCG-AH : 195
sp_Q5E7L3_ : YSIELLRFEG----------HEVEMEVHCSKALTFVRLPTILGEMLGCG-AH : 195
sp_Q87M04_ : YEIILHRFEG----------DEVEMEVHCSKGTYIRTIVDDLGEMLGCG-AH : 195

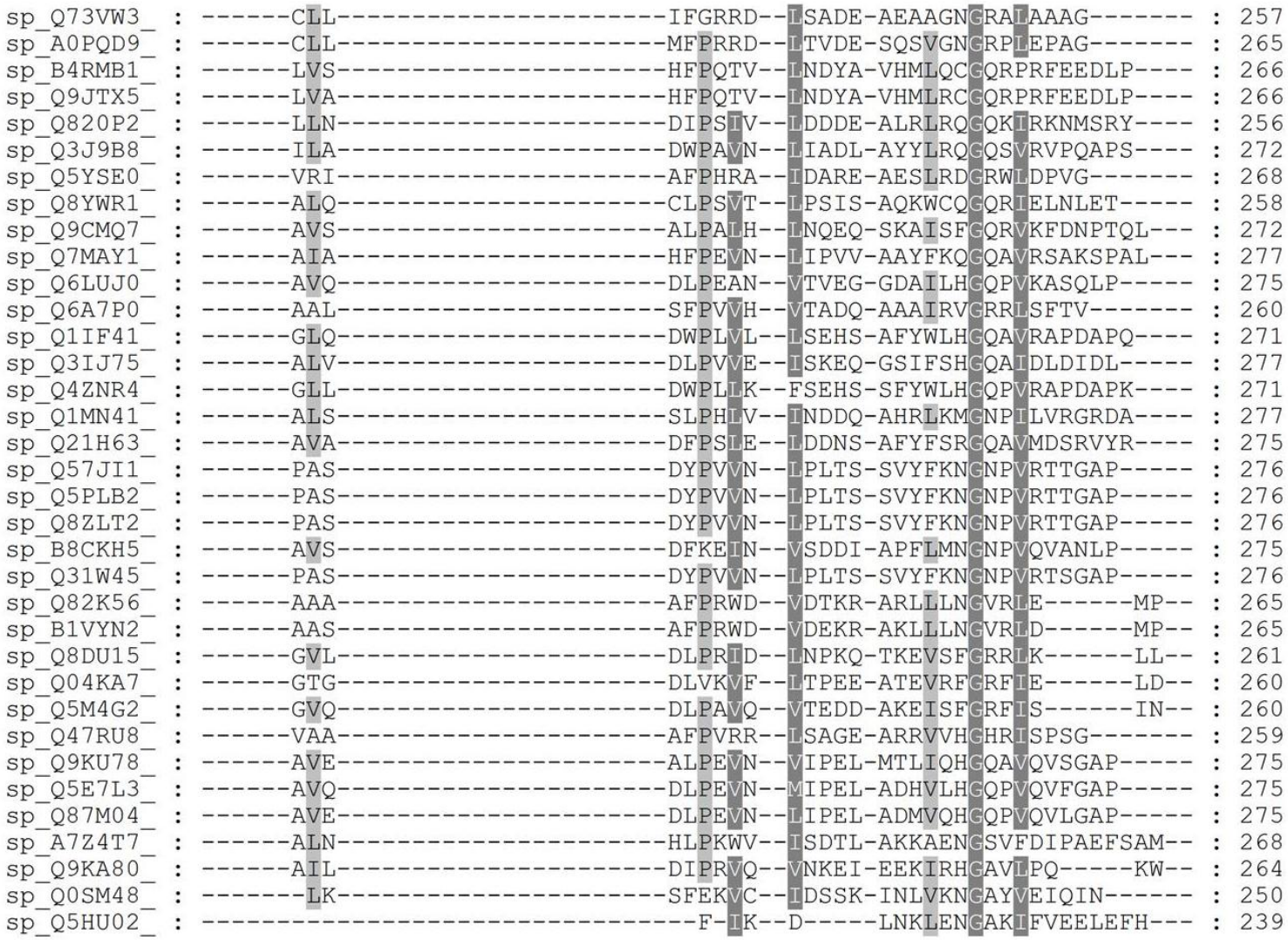

sp_A7Z4T7_ : EDIALTSPVT--Y-------N----------EDTASFRFTVTCSKGTYVRTLAVTIGEKLGYP-AH : 194
sp_Q9KA80_ : FSLERMSDIV--Y-------E----------EGVCRFRFNVSCSKGTYVRTLAVDIGKALGYP-AH : 195
sp_Q0SM48_ : YDIQRLSYDF----------------------SSSLLSLKISCSKGTYIRSIARDLAYSLNSC-AY : 189
sp_Q5HU02_ : FDCKILSYNH------------------------PFLNIEITVSEGAYIRSYCELFARKLGIN-AT : 186
sp_P58063_ : FDLKVVDQP----------------------DADHVTLTMECGKGTYVRAVVRDLAKALGTC-GH : 196
sp_B0BB81_ : N-LRLVKYEY------------------------PRLHFVVQCSKGTYIRSIAHELGNMLGCG-AY : 193
sp_A5N845_ : YSIDIVNIE----------------------L-PYVSFKVTCSKGTYIRSLCNDIGNNLNCG-GT : 186
sp_Q0TPS0_ : YSIENIKIE----------------------SNDNIQMDVCCSKGTYIRSLCYDIGEKLNVG-AT : 188
sp_Q895J5_ : YNIEVLKVN----------------------L-PYVEFKVNCSKGTYIRSLCYDIGNKLGMG-AT : 187
sp_Q3Z7U5_ : YGIEFLGFA----------------------S-PVLRLRIECGHGTYIRSIAFDLGRKLGCG-AY : 187
sp_Q3YSC6_ : LELKLLSFDV---------------------ENNKADFYLSCRKGVYVRSIARDLGIKLGCL-GY : 188
sp_Q7U330_ : YNITLLSYHH------------------------PFVHFEVSVSEGAYVRSIGEIIAKKLGVN-GV : 188
sp_Q03QT7_ : DRFDLQGAGTFDA-------T----------AGTQTIDFEIACSKGTYVRTLAVDLGQQLGVP-AV : 195
sp_Q038M8_ : SQFTRTDEPVFSA-------T----------DGTQRFRFEVHVSKGTYIRTLAVDVGKTLGVA-AV : 195
sp_Q92C26_ : YSLVRLDGVS-PL-------T----------ESNPTFKLEISCGKGTYIRTLAVMIGELLGYP-AH : 193
sp_B0JJJ5_ : DQITVLDWLE----------G----------EFPQIELDIHCGSGTYIRSLARDLGKVLAVG-GT : 189
sp_Q98Q19_ : KSIEIEKIDF----------------------PKVIFKAKVSRGTYIRSLINQIGLELKTY-AL : 184
sp_Q04ED7_ : MEFQQVGGSKFDA-------S----------KGQEYINFVATVSKGTYIRTLIEDFGAKLGLP-AT : 196
sp_A5D2S3_ : KSLEFIRGSG--W-------G----------TPSPRALMHLSCSKGTYVRSLCHDMGSRLGCG-AH : 192
sp_Q1RIG2_ : YDLKCLNYDE---------------------QNATATYYAECSKGTYIRTLAEDLALSLQSL-GF : 189
sp_A7X1Q5_ : KDIGRISELD--F-------K----------ENECHFKIRVICGKGTYIRTLATDIGVKLGFP-AH : 195
sp_Q8NWZ0_ : KDIGRISELD--F-------K----------ENECHFKIRVICGKGTYIRTLATDIGVKLGFP-AH : 195
sp_P65856_ : SEFRRTSPLY--F-------E----------KGICRFSFYVSCSKGTYVRTLAVDLGIKLGYA-SH : 195
sp_Q8CWM3_ : YRIEVLHWQP----------G----------RYPELTLHITCGSGTYIRALARDLGTALGVG-GT : 189
sp_O83859_ : FDLQVLGCEADLGEFKKTQAGRGAAIADLDLTRVRAVTLYVRCSAGFYVRALARDIAAACGSC-AY : 215
sp_P45142_ : FELNFIEYNA------------------------PFLTLEVHCSKGTYIRTLVDDLGEVLGCG-AH : 196
cs gty r d g g
* 280 * 300 * 320 *
sp_P62190_ : VTALRRTRVGRFELDQARSLD-DLAE-RPA----------LSLSLD--EA--- : 231
sp_P60340_ : VIYLRRLAVSKYPVERMVTLE-HLRE-LVEQAEQ---------QDIPAAELLDPLLMPM--DS--- : 246
sp_Q57612_ : MQELRRTKSGCFEEKDAVYLQ-DLLDAYVFWK----------EDGDEEELRRVIKPM--EY--- : 243
sp_Q9Z8L9_ : LEQLRRLRSGRFSIDECIDGN-LLDH-PDFD----------ISPYLRDAHGNSL~~~~~~~~~~ : 235
sp_Q8ZBC4_ : VSYLRRLQVATYPSERMVTLE-QLTA-MVEAAQA---------EGRSPNPELDSLLLPM--DS--- : 246
sp_Q6F7I5_ : LTALHRVQTGHFELVPEYTLE-YLEQ-LNE----------TERDMLLLAP--YA--- : 256

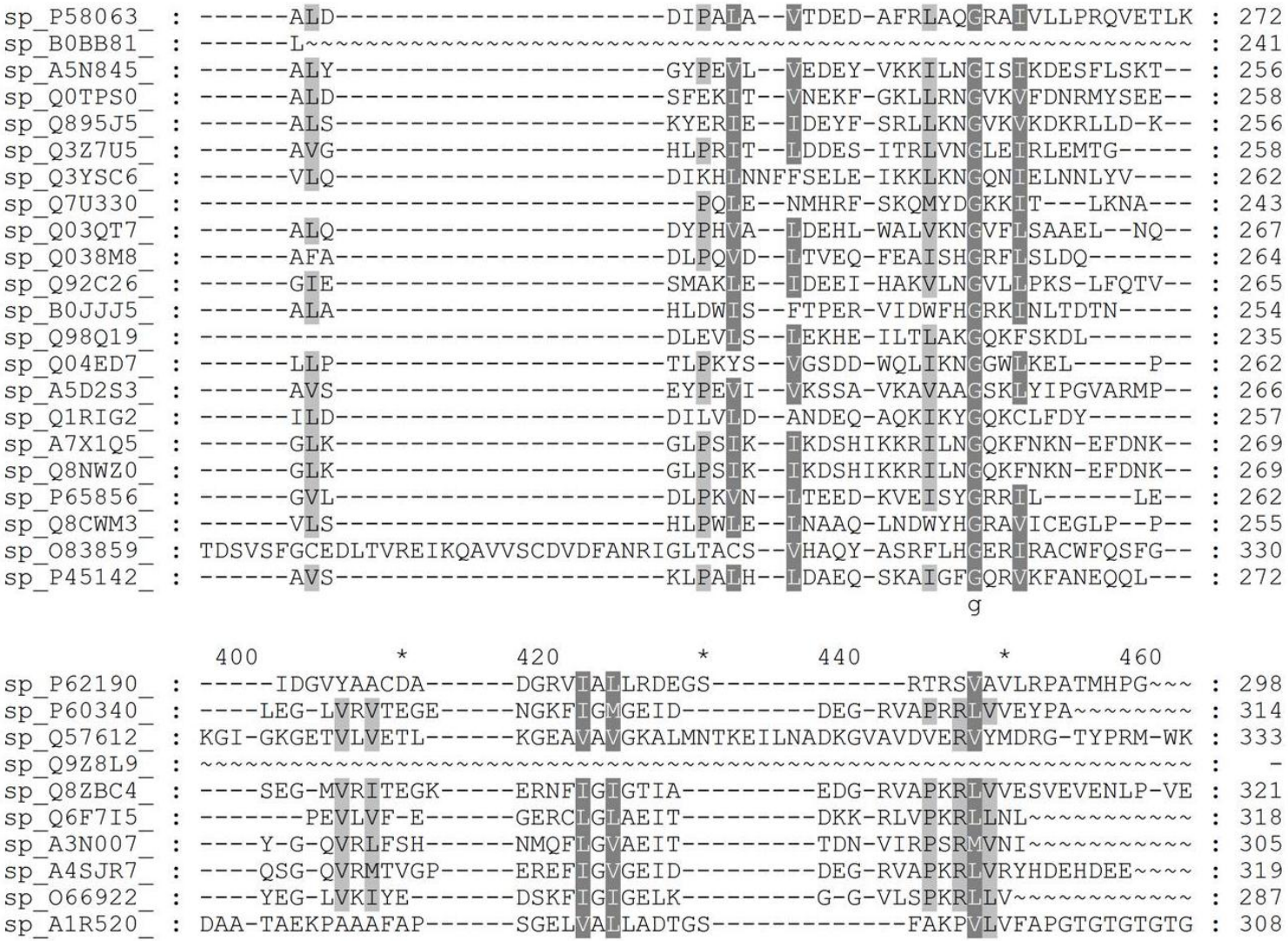

sp_A3N007_ : VTVLRRLAVANYPIEAMMSYA-DLQN-LSEN--------------QPLEELDKYLLPL--DT--- : 240
sp_A4SJR7_ : VTQLRRTQVATYPYERMLTLE-QLEC-IFEQAKA---------ESIPPREQLDPLLLPM--DT--- : 247
sp_O66922_ : MKELRRLKIDEISVEEAVSLE-EFLS-S--------------------ENPEEYVLPV--DT--- : 226
sp_A1R520_ : LTALRRTQVGPYSLDQARTLE-ELAE-EL----------------------EVLEMS--LA--- : 229
sp_A1K7B7_ : LDALRRTAIGDFDAARAVTLE-ALEA-SPA------------------AMRDGLLEPV--DA--- : 235
sp_Q6G5F5_ : IADLRRIAVAPFCENDLITWD-ELKAVELDKSAKNEK----DAPFERNFIKLDELLIET--IS--- : 254
sp_Q1LSL0_ : VTMLRRLAVAHYHTARMITLE-SLQT-AITLALR---------QTPNTLVQLNKLLLPI--DS--- : 247
sp_Q8YEB5_ : ISDLRRVEVAPFTDEDMVTLA-KLEAVWPPLPPKDEDGNVIEPAPRRDFSALDALVIDT--GA--- : 258
sp_P59876_ : ISDLRRVEVAPFTDEDMVTLA-KLEAVWPPLPPKDEDGNVIEPAPRRDFSALDALVIDT--GA--- : 258
sp_P57456_ : VIFLRRLEMASYLHSQLVTIS-YLHK-LLRKEKN---------NNFNFFEKIDNLLMPI--DS--- : 244
sp_Q8K9H3_ : VIRLRRLKIGLLSYSKLVKLS-FLEN-LLNEKNV---------VKINFFKKIDDLLMPV--DT--- : 244
sp_Q89AF6_ : VISLRRLQVGNYTSFQMIDIK-TLNK-LVKY----------------NINVLEHILLSV--DN--- : 238
sp_Q62KL1_ : LTMLRRTGVGPLTLEHAVTLD-ALDA-ATQ------------------DERDARLAPV--DA--- : 240
sp_Q482T7_ : VAHLRRSAVGNYPVEKMITLP-ELEA-LLEQANA---------DEITPSDVLDPLLLPM--NS--- : 245
sp_C3PH13_ : LTALRRTEVGPFRLDDAHPLD-ALED-SPE----------------------LSLSLD--EA--- : 233
sp_P60343_ : LTALRRTQVGPFTLDNAVTLE-KLEE-NPH----------------------VSLTLD--QA--- : 236
sp_Q8FPB3_ : LTALRRTEVGPFTLEDAIPLA-DLQD-NAR----------------------LSLSLD--EA--- : 233
sp_A4QEY6_ : LTALRRTEVGPFTLNDATPLS-KLQE-NPE----------------------LSLNLD--QA--- : 233
sp_Q4JV56_ : LTQLRRTSVGPFDVSEARTLE-QLEE-DPS----------------------LTLNLD--EA--- : 243
sp_A9KBM3_ : VVALRRLRVAHYHEDQMIKLA-HLER-EYD-K-------------ANYTGLDRYLLPL--ET--- : 244
sp_Q4AAX7_ : VVALRRLRVAHYHEDQMIKLA-HLER-EYD-K-------------ANYTGLDRYLLPL--ET--- : 244
sp_Q0TCU3_ : VIYLRRLAVSKYPVERMVTLE-HLRE-LVEQAEQ---------QDIPAAELLDPLLMPM--DS--- : 246
sp_P60341_ : VIYLRRLAVSKYPVERMVTLE-HLRE-LVEQAEQ---------QDIPAAELLDPLLMPM--DS--- : 246
sp_Q7MBB8_ : LARLLRTRSGPFALEASLPLA-AVEA-GFQA-----------------GT--VPLIGP--RT--- : 229
sp_Q4QK43_ : VTMLRRTAVADYPVAEMIPIN-ELQL-LAES---------------FPLSELDRLLLPT--DT--- : 240
sp_A5UBU2_ : VTMLRRTAVADYPVAEMMPIN-ELQL-LAES---------------FPLSELDRLLLPT--DT--- : 240
sp_A5UF28_ : VTMLRRTAVADYPVAEMMLIN-ELEL-LAES---------------FPLSELDRLLLPT--DT--- : 240
sp_Q0I3P3_ : VTRLRRIAVADYPYNKMMTLE-QLQQ-FSEQ---------------EDLDLLDQHLLPM--ES--- : 240
sp_Q2SML1_ : VMELRRTRCGPFSLADSVEFV-SFAD-ESGRQ-------------AAAEKYKERLLPI--DS--- : 241
sp_C4K3E8_ : VSDLRRLQVADYTHDSMVTLE-KLEE-LTLKNN------------TLGSAIDLLLLPI--ES--- : 243
sp_Q5QTZ0_ : VSMLHRNAVADIAEAAMVTLE-QLET-LAEE---------------GYEGLDALLHPA--DL--- : 240
sp_Q5X1C5_ : MTKLHRLYTAGFENNRMYTLD-ELQD-MP------------------LSQRLACLIPI--DQ--- : 236
sp_Q5ZRV6_ : MTKLHRLYTAGFENNRMYTLD-ELQD-MP------------------LSQRLACLIPI--DQ--- : 236
sp_Q5WT38_ : MTKLHRLYTAGFENNRMYTLD-ELQD-MP------------------LSQRLACLIPI--DQ--- : 236
sp_Q65SL2_ : VSMLRRTAVSDYPADKMLTWE-QLQQ-FAQD---------------EDLAALDARLLPV--DS--- : 240

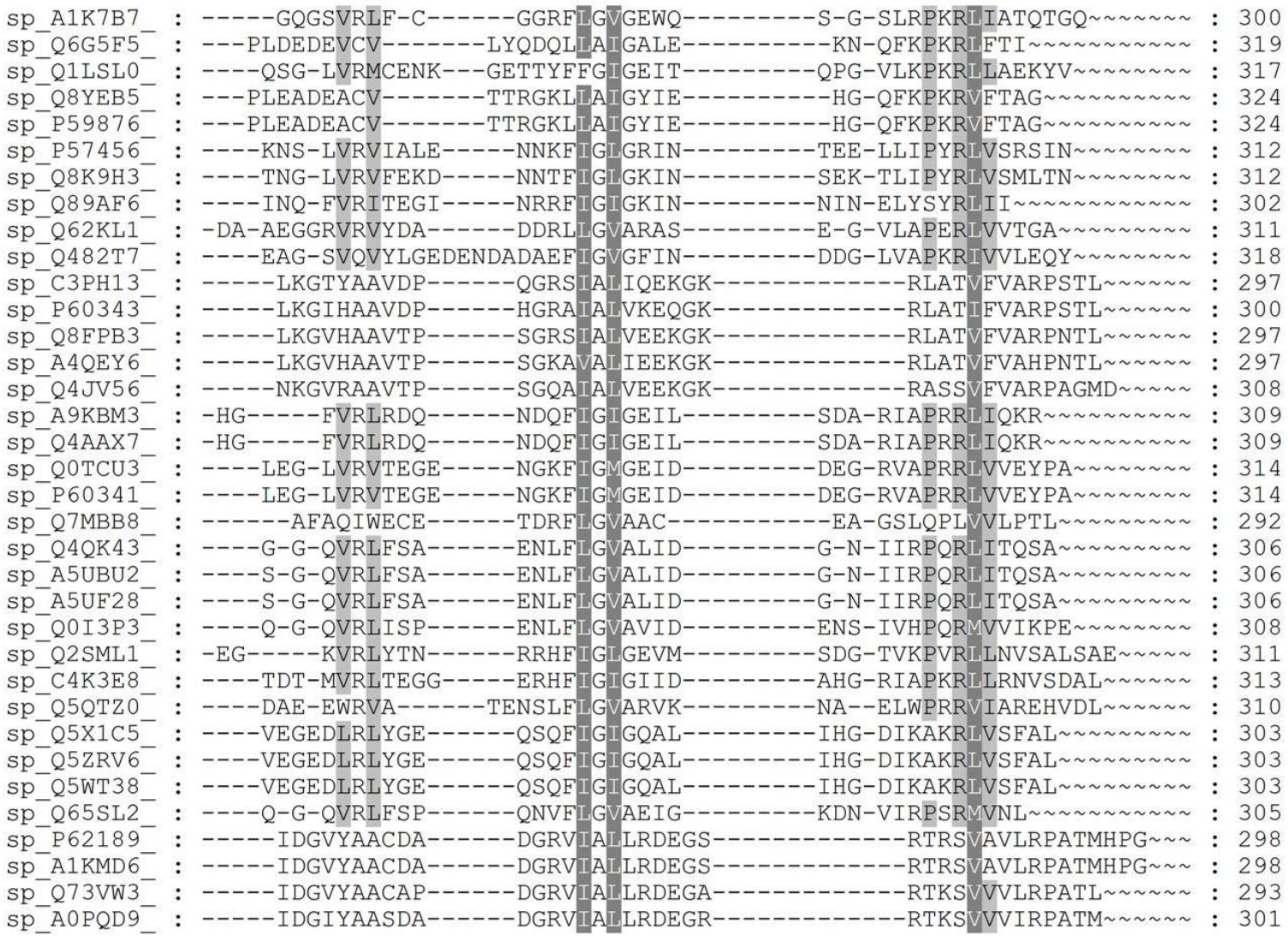

sp_P62189_ : VTALRRTRVGRFELDQARSLD-DLAE-RPA----------LSLSLD--EA--- : 231
sp_A1KMD6_ : VTALRRTRVGRFELDQARSLD-DLAE-RPA----------LSLSLD--EA--- : 231
sp_Q73VW3_ : LTALRRTRVGRFGLEQAYGLD-ELAE-CPR----------LSYSLD--EA--- : 229
sp_A0PQD9_ : LTSLRRTRVSHFDLSQAASLD-ELAE-RPA----------LSLTLD--QA--- : 237
sp_B4RMB1_ : LTALRRTETAGFTIAQSHTLE-ALAN-LDE----------TERDNLLLPC--DV--- : 235
sp_Q9JTX5_ : LTALRRTETAGFTIAQSHTLE-ALAN-LDE----------TERDSLLLPC--DV--- : 235
sp_Q820P2_ : LTALSRISVGHFELSQACDLD-QLES-ETP----------VNRQKLLCPI--DS--- : 225
sp_Q3J9B8_ : VIALRRTQVGSFGASDMISLE-ELEM-L-A-E----------TNVEALGNLLLPV--GQ--- : 241
sp_Q5YSE0_ : LTALRRTRVGPFTLDHARTLD-ALAD-EPR----------LNLDMD--EA--- : 240
sp_Q8YWR1_ : LAALIRTHSSGFNLTDSLTLT-DLET-QLQA----------GT--FEPTPA--DA--- : 228
sp_Q9CMQ7_ : VTVLRRTAVANYPVEAMMNWD-TLQV-LAAQ----------QDLALLDQHLLPT--DS--- : 240
sp_Q7MAY1_ : VIYLRRLQVANYPNDRMVTLE-QLYE-LQKQAKD---------QEIPVGELIDSLLLPM--DS--- : 245
sp_Q6LUJ0_ : VVYLRRTGVSDYPMENVVTLE-QLQA-LRDQAIE---------QGIEPGELLDPLLLPT--DS--- : 245
sp_Q6A7P0_ : LTALRRTRVGPFDLTAACVDIFAQDA-VTP----------TPMTMA--EA--- : 232
sp_Q1IF41_ : VAELRRTQAGPFALAQTVTLE-ELEQ-AHA-E----------GGNEALDRFLMPS--DS--- : 240
sp_Q3IJ75_ : VVMLHRTAVGQYPAEKMVSLD-EIEA-LLAKAKD---------EEVAPSTYLDALLLPL--DT--- : 247
sp_Q4ZNR4_ : VAELRRTQAGPFTLAQTVTLE-ELEQ-VHA-D----------GGNEAVDRFLMPS--DS--- : 240
sp_Q1MN41_ : VSGLRRTFVAPFAEGAMVPLA-DLVALEAI----------EDMDERLAALDALLIDT--CE--- : 246
sp_Q21H63_ : VAKLRRTQAGPFTIDQSITIE-ELEQ-ERG-E----------RLAEVLDHHLLPT--DI--- : 244
sp_Q57JI1_ : VTYLRRLTVSKYPVDRMVTLE-HLQT-LVAQAEQ---------QGVPAAQLLDPLLMPM--DS--- : 246
sp_Q5PLB2_ : VTYLRRLTVSKYPVDRMVTLE-HLQT-LVAQAEQ---------QGVPAAQWLDPLLMPM--DS--- : 246
sp_Q8ZLT2_ : VTYLRRLTVSKYPVDRMVTLE-HLQT-LVAQAEQ---------QGVPAAQLLDPLLMPM--DS--- : 246
sp_B8CKH5_ : VIMLRRTQVAGYPYDNMVSLA-ELEA-LVTKVEA---------ESLTLAEVLDPLLLPM--DT--- : 245
sp_Q31W45_ : VIYLRRLAVSKYPVERMVTLE-HLRE-LVEQAEQ---------QDIPAAELLDPLLMPM--DS--- : 246
sp_Q82K56_ : LTALRRTRVGPYKLDAARTLD-QLQE-EL----------TVMPIA--EA--- : 238
sp_B1VYN2_ : LTALRRTRVGPYGLDAARTLD-QHQE-EL----------TVMPVA--EA--- : 238
sp_Q8DU15_ : MSFLKRTASAGLDLSQALTLA-EIAE-KVEE----------KDF-SFLLPI--EY--- : 234
sp_Q04KA7_ : MSHLTRTSAAGLQLEDALALE-EIAE-KVEA----------GQL-DFLHPL--EI--- : 233
sp_Q5M4G2_ : MSALRRTASAGLTLDSSLSLS-QISE-MVEA----------GDQ-SFLLPI--EF--- : 233
sp_Q47RU8_ : LTALRRTRVGPYTVEQAATLD-ELAA-EF----------TVMPLA--EA--- : 231
sp_Q9KU78_ : VTMLRRVGVANYPYERMVTLE-QLNA-LVEQAHR---------DEKAVADVLDPLLLPM--DT--- : 245
sp_Q5E7L3_ : VTYLRRTGVSNYPYENMVTIE-DLEA-LLEQAHR---------EERAPRELLDPLLMPM--DS--- : 245
sp_Q87M04_ : VTMLRRTAVAKYPYEKMVTLE-QLNE-LLEQAHR---------EEIAPRELLDPLLMPM--DT--- : 245
sp_A7Z4T7_ : MSHLIRTASGDFSLDECFTFE-ELEQ-QVSD----------GTVAEHAVPI--DR--- : 235
sp_Q9KA80_ : MSDLVRTKSGPFSLEECFTFT-ELEE-RLEQ----------GEGSSLLLPI--ET--- : 236

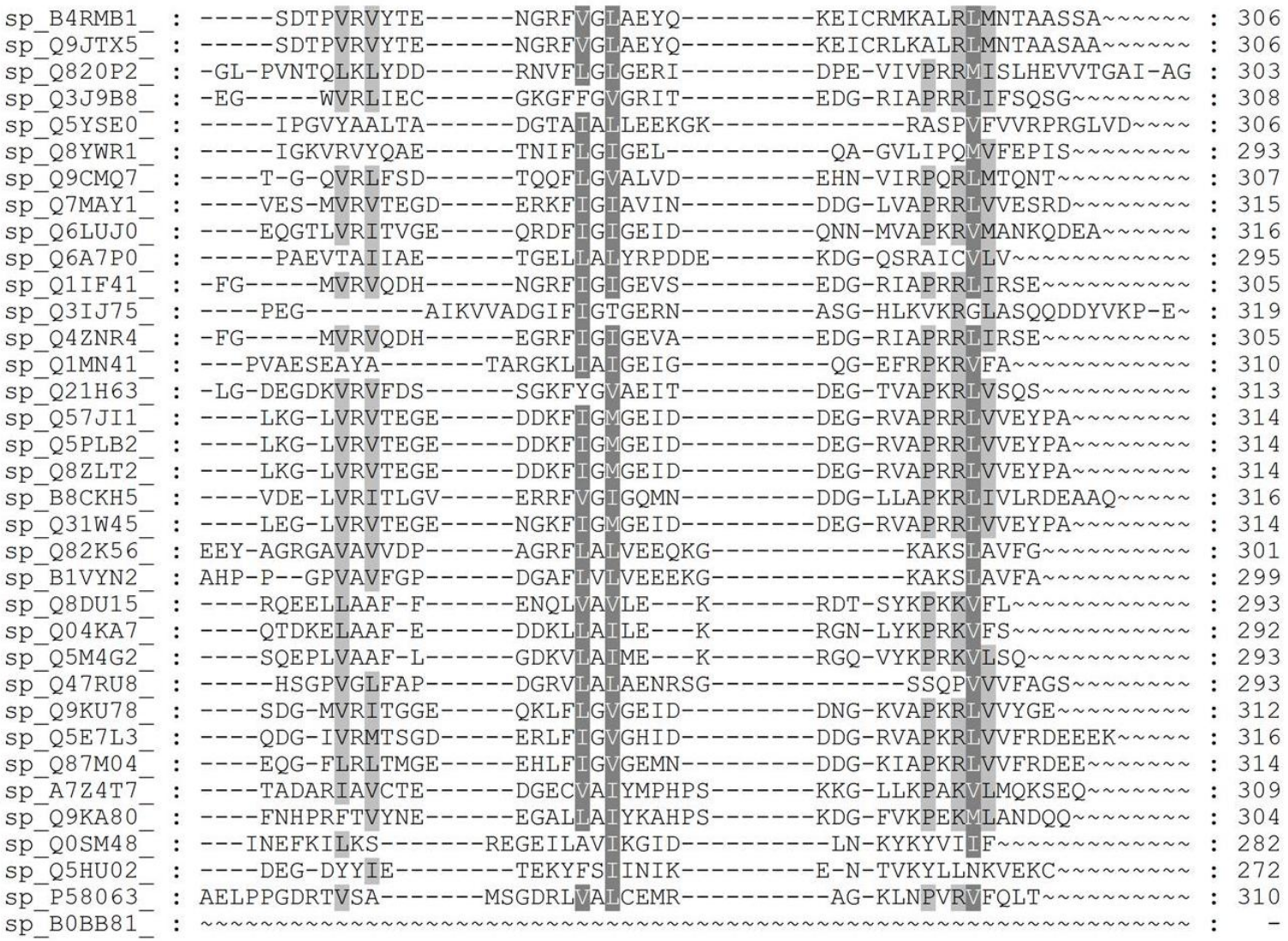



sp_O66922_ : ------LFR----------VIPEVR--LNTFE-AGKILQGKRILIKNYD----- : 256
sp_A1R520_ : ------ARS----------LMPNRE--LSEQE-TTEISFGRRIAAGPGAGTP-- : 262
sp_A1K7B7_ : ------LVA----------HFPKVE--LQPAE-AAAILQGRELRKPED------ : 264
sp_Q6G5F5_ : ------ALE----------CLSHYT--LSQTQ-AQQVMKGNTVLLYHHDV---- : 285
sp_Q1LSL0_ : ------AVA----------NFPAIK--LSKDS-VARVRKGQMVAVDQCW----- : 277
sp_Q8YEB5_ : ------ALD----------CLPQVP--LSDDQ-AQRVRLGNPVILRGRDA---- : 289
sp_P59876_ : ------ALD----------CLPQVP--LSDDQ-AQRVRLGNPVILRGRDA---- : 289
sp_P57456_ : ------PVS----------FLPKVY--LFPQQ-SYNFQLGQTVIFFSDI----- : 274
sp_Q8K9H3_ : ------PVY----------FLPKIY--ISIEK-LSVFRLGQKVNFSSSI----- : 274
sp_Q89AF6_ : ------AIS----------CFPEVN--IVPNV-IKNLQNGQKVKTHSGF----- : 268
sp_Q62KL1_ : ------LLS----------TFPCVK--LDAAL-ATRFLHGQRLKLSELAARP-- : 273
sp_Q482T7_ : ------AVD----------GMHCVY--VDDMS-ANFLRHGNPVQAYNQP----- : 275
sp_C3PH13_ : ------LAR----------SYPVLE--VTDEE-AQALAMGKWLEPRG------- : 261
sp_P60343_ : ------LAA----------SYPVLS--VSEKE-ASDLAMGKWLTPRG------- : 264
sp_Q8FPB3_ : ------LAR----------SYPVLE--ITEAE-GEALSMGKWLEPRG------- : 261
sp_A4QEY6_ : ------LTR----------SYPVLD--ITEDE-GVDLSMGKWLEPRG------- : 261
sp_Q4JV56_ : ------MVR----------CFDTRE--ISESE-GVDLSLGKWLKPVG------- : 271
sp_A9KBM3_ : ------MVS----------HFPAIK--LSSST-AFYLQQGQAVMVPNAPT---- : 275
sp_Q4AAX7_ : ------MVS----------HFPAIK--LSSST-AFYLQQGQAVMVPNAPT---- : 275
sp_Q0TCU3_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTSGAP----- : 276
sp_P60341_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTSGAP----- : 276
sp_Q7MBB8_ : ------ALA----------HLAAVH--LEPPL-AARWLMGQRVPGDCEG----- : 259
sp_Q4QK43_ : ------AVS----------KLPALH--LDVEQ-SKAIGFGQRVKFANEQQL--- : 272
sp_A5UBU2_ : ------AVS----------KLPALH--LDVEQ-SKAIGFGQRVKFANEQQL--- : 272
sp_A5UF28_ : ------AVS----------KLPALH--LDVEQ-SKAIGFGQRVKFANEQQL--- : 272
sp_Q0I3P3_ : ------AVI----------RLPGLH--LTKEQ-ARAVGFGQRIKFLNEQGI--- : 272
sp_Q2SML1_ : ------SVQ----------QFSEII--LDEAN-TGSIIQGQTVKISPLSS---- : 272
sp_C4K3E8_ : ------TAL----------HLPEVN--LESGA-AACIKQGQPVSFFDNP----- : 273
sp_Q5QTZ0_ : ------LLG----------QLPEVT--VTQAQ-TRDFLHGQPIPLPEQNGN--- : 272
sp_Q5X1C5_ : ------AIQ----------HLTPVI--LSDSE-VTAIRQGKVISNKTGA----- : 266
sp_Q5ZRV6_ : ------AVQ----------HLTPVI--LSDSE-VTAIRQGKVISNKTGA----- : 266
sp_Q5WT38_ : ------AVQ----------HLTPVI--LSDSE-VTAIRQGKVISNKTSA----- : 266
sp_Q65SL2_ : ------AVS----------KLPVLS--LSEEQ-TKAVGFGQRVKFDNLQQL--- : 272
sp_P62189_ : ------CLL----------MFARRD--LTAAE-ASAAANGRSLPAVG------- : 259
sp_A1KMD6_ : ------CLL----------MFARRD--LTAAE-ASAAANGRSLPAVG------- : 259


sp_Q73VW3_ : ------CLL----------IFGRRD--LSADE-AEAAGNGRALAAAG------- : 257
sp_A0PQD9_ : ------CLL----------MFPRRD--LTVDE-SQSVGNGRPLEPAG------- : 265
sp_B4RMB1_ : ------LVS----------HFPQTV--LNDYA-VHMLQCGQRPRFEEDLP---- : 266
sp_Q9JTX5_ : ------LVA----------HFPQTV--LNDYA-VHMLRCGQRPRFEEDLP---- : 266
sp_Q820P2_ : ------LLN----------DIPSIV--LDDDE-ALRLRQGQKIRKNMSRY---- : 256
sp_Q3J9B8_ : ------ILA----------DWPAVN--LIADL-AYYLRQGQSVRVPQAPS---- : 272
sp_Q5YSE0_ : ------VRI----------AFPHRA--IDARE-AESLRDGRWLDPVG------- : 268
sp_Q8YWR1_ : ------ALQ----------CLPSVT--LPSIS-AQKWCQGQRIELNLET----- : 258
sp_Q9CMQ7_ : ------AVS----------ALPALH--LNQEQ-SKAISFGQRVKFDNPTQL--- : 272
sp_Q7MAY1_ : ------AIA----------HFPEVN--LIPVV-AAYFKQGQAVRSAKSPAL--- : 277
sp_Q6LUJ0_ : ------AVQ----------DLPEAN--VTVEG-GDAILHGQPVKASQLP----- : 275
sp_Q6A7P0_ : ------AAL----------SFPVVH--VTADQ-AAAIRVGRRLSFTV------- : 260
sp_Q1IF41_ : ------GLQ----------DWPLVL--LSEHS-AFYWLHGQAVRAPDAPQ---- : 271
sp_Q3IJ75_ : ------ALV----------DLPVVE--ISKEQ-GSIFSHGQAIDLDIDL----- : 277
sp_Q4ZNR4_ : ------GLL----------DWPLLK--FSEHS-SFYWLHGQPVRAPDAPK---- : 271
sp_Q1MN41_ : ------ALS----------SLPHLV--INDDQ-AHRLKMGNPILVRGRDA---- : 277
sp_Q21H63_ : ------AVA----------DFPSLE--LDDNS-AFYFSRGQAVMDSRVYR---- : 275
sp_Q57JI1_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTTGAP----- : 276
sp_Q5PLB2_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTTGAP----- : 276
sp_Q8ZLT2_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTTGAP----- : 276
sp_B8CKH5_ : ------AVS----------DFKEIN--VSDDI-APFLMNGNPVQVANLP----- : 275
sp_Q31W45_ : ------PAS----------DYPVVN--LPLTS-SVYFKNGNPVRTSGAP----- : 276
sp_Q82K56_ : ------AAA----------AFPRWD--VDTKR-ARLLLNGVRLE------MP-- : 265
sp_B1VYN2_ : ------AAS----------AFPRWD--VDEKR-AKLLLNGVRLD------MP-- : 265
sp_Q8DU15_ : ------GVL----------DLPRID--LNPKQ-TKEVSFGRRLK------LL-- : 261
sp_Q04KA7_ : ------GTG----------DLVKVF--LTPEE-ATEVRFGRFIE------LD-- : 260
sp_Q5M4G2_ : ------GVQ----------DLPAVQ--VTEDD-AKEISFGRFIS------IN-- : 260
sp_Q47RU8_ : ------VAA----------AFPVRR--LSAGE-ARRVVHGHRISPSG------- : 259
sp_Q9KU78_ : ------AVE----------ALPEVN--VIPEL-MTLIQHGQAVQVSGAP----- : 275
sp_Q5E7L3_ : ------AVQ----------DLPEVN--MIPEL-ADHVLHGQPVQVFGAP----- : 275
sp_Q87M04_ : ------AVE----------DLPEVN--LIPEL-ADMVQHGQPVQVLGAP----- : 275
sp_A7Z4T7_ : ------ALN----------HLPKWV--ISDTL-AKKAENGSVFDIPAEFSAM-- : 268
sp_Q9KA80_ : ------AIL----------DIPRVQ--VNKEI-EEKIRHGAVLPQ-----KW-- : 264
sp_Q0SM48_ : -------LK----------SFEKVC--IDSSK-INLVKNGAYVEIQIN------ : 250
sp_Q5HU02_ : ---------------------------F-IK--D-----LNKLENGAKIFVEELEFH--- : 239


sp_P58063_ : ------ALD----------DIPALA--VTDED-AFRLAQGRAIVLLPRQVETLK : 272
sp_B0BB81_ : ------L~~~~~~ : 241
sp_A5N845_ : ------ALY----------GYPEVL--VEDEY-VKKILNGISIKDESFLSKT-- : 256
sp_Q0TPS0_ : ------ALD----------SFEKIT--VNEKF-GKLLRNGVKVFDNRMYSEE-- : 258
sp_Q895J5_ : ------ALS----------KYERIE--IDEYF-SRLLKNGVKVKDKRLLD-K-- : 256
sp_Q3Z7U5_ : ------AVG----------HLPRIT--LDDES-ITRLVNGLEIRLEMTG----- : 258
sp_Q3YSC6_ : ------VLQ----------DIKHLNNFFSELE-IKKLKNGQNIELNNLYV---- : 262
sp_Q7U330_ : ----------------------------PQLE--NMHRF-SKQMYDGKKIT---LKNA--- : 243
sp_Q03QT7_ : ------ALQ----------DYPHVA--LDEHL-WALVKNGVFLSAAEL--NQ-- : 267
sp_Q038M8_ : ------AFA----------DLPQVD--LTVEQ-FEAISHGRFLSLDQ------- : 264
sp_Q92C26_ : ------GIE----------SMAKLE--IDEEI-HAKVLNGVLLPKS-LFQTV-- : 265
sp_B0JJJ5_ : ------ALA----------HLDWIS--FTPER-VIDWFHGRKINLTDTN----- : 254
sp_Q98Q19_ : ----------------------------DLEVLS--LEKHE-ILTLAKGQKFSKDL------- : 235
sp_Q04ED7_ : ------LLP----------TLPKYS--VGSDD-WQLIKNGGWLKEL-----P-- : 262
sp_A5D2S3_ : ------AVS----------EYPEVI--VKSSA-VKAVAAGSKLYIPGVARMP-- : 266
sp_Q1RIG2_ : ------ILD----------DILVLD--ANDEQ-AQKIKYGQKCLFDY------- : 257
sp_A7X1Q5_ : ------GLK----------GLPSIK--IKDSHIKKRILNGQKFNKN-EFDNK-- : 269
sp_Q8NWZ0_ : ------GLK----------GLPSIK--IKDSHIKKRILNGQKFNKN-EFDNK-- : 269
sp_P65856_ : ------GVL----------DLPKVN--LTEED-KVEISYGRRIL------LE-- : 262
sp_Q8CWM3_ : ------VLS----------HLPWLE--LNAAQ-LNDWYHGRAVICEGLP--P-- : 255
sp_O83859_ : TDSVSFGCEDLTVREIKQAVVSCDVDFANRIGLTACS--VHAQY-ASRFLHGERIRACWFQSFG-- : 330
sp_P45142_ : ------AVS----------KLPALH--LDAEQ-SKAIGFGQRVKFANEQQL--- : 272
g
400 * 420 * 440 * 460
sp_P62190_ : -----IDGVYAACDA------DGRVIALLRDEGS-------------RTRSVAVLRPATMHPG~~~ : 298
sp_P60340_ : ----LEG-LVRVTEGE-----NGKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_Q57612_ : KGI-GKGETVLVETL------KGEAVAVGKALMNTKEILNADKGVAVDVERVYMDRG-TYPRM-WK : 333
sp_Q9Z8L9_ : ~~~~~~ : -
sp_Q8ZBC4_ : ----SEG-MVRITEGK-----ERNFIGIGTIA---------EDG-RVAPKRLVVESVEVENLP-VE : 321
sp_Q6F7I5_ : -------PEVLVF-E------GERCLGLAEIT---------DKK-RLVPKRLLNL~~~~~~~~~~~ : 318
sp_A3N007_ : ----Y-G-QVRLFSH------NMQFLGVAEIT---------TDN-VIRPSRMVNI~~~~~~~~~~~ : 305
sp_A4SJR7_ : ----QSG-QVRMTVGP-----EREFIGVGEID---------DEG-RVAPKRLVRYHDEHDEE~~~~ : 319
sp_O66922_ : ----YEG-LVKIYE-------DSKFIGIGELK---------G-G-VLSPKRLLV~~~~~~~~~~~~ : 287
sp_A1R520_ : DAA-TAEKPAAAFAP------SGELVALLADTGS-------------FAKPVLVFAPGTGTGTGTG : 308


sp_A1K7B7_ : -----GQGSVRLF-C------GGRFLGVGEWQ---------S-G-SLRPKRLIATQTGQ~~~~~~~ : 300
sp_Q6G5F5_ : ---PLDEDEVCV-------LYQDQLLAIGALE----------KN-QFKPKRLFTI~~~~~~~~~~~ : 319
sp_Q1LSL0_ : ----QSG-LVRMCENK---GETTYFFGIGEIT---------QPG-VLKPKRLLAEKYV~~~~~~~~ : 317
sp_Q8YEB5_ : ---PLEADEACV-------TTRGKLLAIGYIE----------HG-QFKPKRVFTAG~~~~~~~~~~ : 324
sp_P59876_ : ---PLEADEACV-------TTRGKLLAIGYIE----------HG-QFKPKRVFTAG~~~~~~~~~~ : 324
sp_P57456_ : ----KNS-LVRVIALE-----NNKFIGLGRIN---------TEE-LLIPYRLVSRSIN~~~~~~~~ : 312
sp_Q8K9H3_ : ----TNG-LVRVFEKD-----NNTFIGLGKIN---------SEK-TLIPYRLVSMLTN~~~~~~~~ : 312
sp_Q89AF6_ : ----INQ-FVRITEGI-----NRRFIGIGKIN---------NIN-ELYSYRLII~~~~~~~~~~~~ : 302
sp_Q62KL1_ : -DA-AEGGRVRVYDA------DDRLLGVARAS---------E-G-VLAPERLVVTGA~~~~~~~~~ : 311
sp_Q482T7_ : ----EAG-SVQVYLGEDENDADAEFIGVGFIN---------DDG-LVAPKRIVVLEQY~~~~~~~~ : 318
sp_C3PH13_ : -----LKGTYAAVDP------QGRSIALIQEKGK-------------RLATVFVARPSTL~~~~~~ : 297
sp_P60343_ : -----LKGIHAAVDP------HGRAIALVKEQGK-------------RLATIFVARPSTL~~~~~~ : 300
sp_Q8FPB3_ : -----LKGVHAAVTP------SGRSIALVEEKGK-------------RLATVFVARPNTL~~~~~~ : 297
sp_A4QEY6_ : -----LKGVHAAVTP------SGKAVALIEEKGK-------------RLATVFVAHPNTL~~~~~~ : 297
sp_Q4JV56_ : -----NKGVRAAVTP------SGQAIALVEEKGK-------------RASSVFVARPAGMD~~~~~ : 308
sp_A9KBM3_ : -HG-----FVRLRDQ------NDQFIGIGEIL---------SDA-RIAPRRLIQKR~~~~~~~~~~ : 309
sp_Q4AAX7_ : -HG-----FVRLRDQ------NDQFIGIGEIL---------SDA-RIAPRRLIQKR~~~~~~~~~~ : 309
sp_Q0TCU3_ : ----LEG-LVRVTEGE-----NGKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_P60341_ : ----LEG-LVRVTEGE-----NGKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_Q7MBB8_ : ------AFAQIWECE------TDRFLGVAAC-----------EA-GSLQPLVVLPTL~~~~~~~~~ : 292
sp_Q4QK43_ : ----G-G-QVRLFSA------ENLFLGVALID---------G-N-IIRPQRLITQSA~~~~~~~~~ : 306
sp_A5UBU2_ : ----S-G-QVRLFSA------ENLFLGVALID---------G-N-IIRPQRLITQSA~~~~~~~~~ : 306
sp_A5UF28_ : ----S-G-QVRLFSA------ENLFLGVALID---------G-N-IIRPQRLITQSA~~~~~~~~~ : 306
sp_Q0I3P3_ : ----Q-G-QVRLISP------ENLFLGVAVID---------ENS-IVHPQRMVVIKPE~~~~~~~~ : 308
sp_Q2SML1_ : -EG-----KVRLYTN------RRHFIGLGEVM---------SDG-TVKPVRLLNVSALSAE~~~~~ : 311
sp_C4K3E8_ : ----TDT-MVRLTEGG-----ERHFIGIGIID---------AHG-RIAPKRLLRNVSDAL~~~~~~ : 313
sp_Q5QTZ0_ : ----DAE-EWRVA------TENSLFLGVARVK---------NA--ELWPRRVIAREHVDL~~~~~~ : 310
sp_Q5X1C5_ : ----VEGEDLRLYGE------QSQFIGIGQAL---------IHG-DIKAKRLVSFAL~~~~~~~~~ : 303
sp_Q5ZRV6_ : ----VEGEDLRLYGE------QSQFIGIGQAL---------IHG-DIKAKRLVSFAL~~~~~~~~~ : 303
sp_Q5WT38_ : ----VEGEDLRLYGE------QSQFIGIGQAL---------IHG-DIKAKRLVSFAL~~~~~~~~~ : 303
sp_Q65SL2_ : ----Q-G-QVRLFSP------QNVFLGVAEIG---------KDN-VIRPSRMVNL~~~~~~~~~~~ : 305
sp_P62189_ : -----IDGVYAACDA------DGRVIALLRDEGS-------------RTRSVAVLRPATMHPG~~~ : 298
sp_A1KMD6_ : -----IDGVYAACDA------DGRVIALLRDEGS-------------RTRSVAVLRPATMHPG~~~ : 298
sp_Q73VW3_ : -----IDGVYAACAP------DGRVIALLRDEGA-------------RTKSVVVLRPATL~~~~~~ : 293
sp_A0PQD9_ : -----IDGIYAASDA------DGRVIALLRDEGR-------------RTKSVVVIRPATM~~~~~~ : 301


sp_B4RMB1_ : -----SDTPVRVYTE------NGRFVGLAEYQ---------KEICRMKALRLMNTAASSA~~~~~~ : 306
sp_Q9JTX5_ : -----SDTPVRVYTE------NGRFVGLAEYQ---------KEICRLKALRLMNTAASAA~~~~~~ : 306
sp_Q820P2_ : -GL-PVNTQLKLYDD------RNVFLGLGERI---------DPE-VIVPRRMISLHEVVTGAI-AG : 303
sp_Q3J9B8_ : -EG-----WVRLIEC------GKGFFGVGRIT---------EDG-RIAPRRLIFSQSG~~~~~~~~ : 308
sp_Q5YSE0_ : -----IPGVYAALTA------DGTAIALLEEKGK-------------RASPVFVVRPRGLVD~~~~ : 306
sp_Q8YWR1_ : -----IGKVRVYQAE------TNIFLGIGEL-----------QA-GVLIPQMVFEPIS~~~~~~~~ : 293
sp_Q9CMQ7_ : ----T-G-QVRLFSD------TQQFLGVALVD---------EHN-VIRPQRLMTQNT~~~~~~~~ : 307
sp_Q7MAY1_ : ----VES-MVRVTEGD-----ERKFIGIAVIN---------DDG-LVAPRRLVVESRD~~~~~~~~ : 315
sp_Q6LUJ0_ : ----EQGTLVRITVGE-----QRDFIGIGEID---------QNN-MVAPKRVMANKQDEA~~~~~~ : 316
sp_Q6A7P0_ : -----PAEVTAIIAE------TGELLALYRPDDE-------KDG-QSRAICVLV~~~~~~~~~~~~ : 295
sp_Q1IF41_ : -FG-----MVRVQDH------NGRFIGIGEVS---------EDG-RIAPRRLIRSE~~~~~~~~~~ : 305
sp_Q3IJ75_ : ----PEG--------AIKVVADGIFIGTGERN---------ASG-HLKVKRGLASQQDDYVKP-E~ : 319
sp_Q4ZNR4_ : -FG-----MVRVQDH------EGRFIGIGEVA---------EDG-RIAPRRLIRSE~~~~~~~~~~ : 305
sp_Q1MN41_ : ---PVAESEAYA-------TARGKLIAIGEIG----------QG-EFRPKRVFA~~~~~~~~~~~~ : 310
sp_Q21H63_ : -LG-DEGDKVRVFDS------SGKFYGVAEIT---------DEG-TVAPKRLVSQS~~~~~~~~~~ : 313
sp_Q57JI1_ : ----LKG-LVRVTEGE-----DDKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_Q5PLB2_ : ----LKG-LVRVTEGE-----DDKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_Q8ZLT2_ : ----LKG-LVRVTEGE-----DDKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_B8CKH5_ : ----VDE-LVRITLGV-----ERRFVGIGQMN---------DDG-LLAPKRLIVLRDEAAQ~~~~~ : 316
sp_Q31W45_ : ----LEG-LVRVTEGE-----NGKFIGMGEID---------DEG-RVAPRRLVVEYPA~~~~~~~~ : 314
sp_Q82K56_ : EEY-AGRGAVAVVDP------AGRFLALVEEQKG-------------KAKSLAVFG~~~~~~~~~~ : 301
sp_B1VYN2_ : AHP-P--GPVAVFGP------DGAFLVLVEEEKG-------------KAKSLAVFA~~~~~~~~~~ : 299
sp_Q8DU15_ : ----RQEELLAAF-F------ENQLVAVLE---K-------RDT-SYKPKKVFL~~~~~~~~~~~~ : 293
sp_Q04KA7_ : ----QTDKELAAF-E------DDKLLAILE---K-------RGN-LYKPRKVFS~~~~~~~~~~~~ : 292
sp_Q5M4G2_ : ----SQEPLVAAF-L------GDKVLAIME---K-------RGQ-VYKPRKVLSQ~~~~~~~~~~~ : 293
sp_Q47RU8_ : -----HSGPVGLFAP------DGRVLALAENRSG-------------SSQPVVVFAGS~~~~~~~~ : 293
sp_Q9KU78_ : ----SDG-MVRITGGE-----QKLFLGVGEID---------DNG-KVAPKRLVVYGE~~~~~~~~~ : 312
sp_Q5E7L3_ : ----QDG-IVRMTSGD-----ERLFIGVGHID---------DDG-RVAPKRLVVFRDEEEK~~~~~ : 316
sp_Q87M04_ : ----EQG-FLRLTMGE-----EHLFIGVGEMN---------DDG-KIAPKRLVVFRDEE~~~~~~~ : 314
sp_A7Z4T7_ : ----TADARIAVCTE------DGECVAIYMPHPS-------KKG-LLKPAKVLMQKSEQ~~~~~~~ : 309
sp_Q9KA80_ : ----FNHPRFTVYNE------EGALLAIYKAHPS-------KDG-FVKPEKMLANDQQ~~~~~~~~ : 304
sp_Q0SM48_ : ---INEFKILKS-------REGEILAVIKGID----------LN-KYKYVIIF~~~~~~~~~~~~ : 282
sp_Q5HU02_ : ----DEG-DYYIE--------TEKYFSIINIK---------E-N-TVKYLLNKVEKC~~~~~~~~~ : 272
sp_P58063_ : AELPPGDRTVSA-------MSGDRLVALCEMR----------AG-KLNPVRVFQLT~~~~~~~~~~ : 310
sp_B0BB81_ : ~~~~~~ : -

```
sp_A5N845_ : ----IKDKLYRVYIE------GNKFIGIGM-N---------KDF-QFKMVKLFV~~~~~~~~~~~~ : 289
sp_Q0TPS0_ : --V-EFNKLYRVYED------NGVFLGLGK-R---------DEK-GFKLEKLLIEE~~~~~~~~~~ : 294
sp_Q895J5_ : --I-KANDILRVYQE-------DKFIGLGQ-K---------TKE-GFKIVKLLV~~~~~~~~~~~~ : 289
sp_Q3Z7U5_ : -----QPEAMAVYSA------ENRFAAVIRPE---------TDG-SWHPAKVFLSPCPKKNAD~~~ : 300
sp_Q3YSC6_ : ---IRNCDICYV-------STGNVPVAICSIV----------NS-VVRPVRIFNVRGLVF~~~~~~ : 301
sp_Q7U330_ : ----QKG-KYIVC--------FEDFFSIIEIF---------SNG-GIQYILNRIEYVDTFKKT--- : 283
sp_Q03QT7_ : ----ATEPELALT-Y------QGETKCLYRWS-D-------EKK-QYRPLKMFAVN~~~~~~~~~~ : 303
sp_Q038M8_ : -----QTPRVRLH-F------AGVLKAIYRRE----------DD-QYRPDLMFLANEKNV~~~~~~ : 301
sp_Q92C26_ : ----ENEPRVALI-F------QEKLTAIYKPHPE-------KQD-LFKPEKVIELQQA~~~~~~~~ : 304
sp_B0JJJ5_ : --V-IIGSLVAVESL------EAQFLGIGEIVVA-------EDE-YYLQPKIVIQQ~~~~~~~~~~ : 293
sp_Q98Q19_ : ----ADGKYAFIYKN------TKKILGICKIESK-----------IIAPIKIFNKKIEKSLKKDEK : 280
sp_Q04ED7_ : ----ISVSPVKIF-Y------NNSFQAIYE---K-------QDG-FYKPKKMLIHEDH~~~~~~~~ : 298
sp_A5D2S3_ : LDL-DCGKLVRLTGP-------DGLLAIAE-A---------GRE-PFDKEKLFFKPVCVLARQ-AG : 312
sp_Q1RIG2_ : ---DKNVDFMWV-------RYKGVLLTIGSLN----------KN-CFHSLRVFNLTQ~~~~~~~~~ : 293
sp_A7X1Q5_ : ----IKDQIVFIDDD------SEKVLAIYMVHPT-------KES-EIKPKKVFN~~~~~~~~~~~~ : 305
sp_Q8NWZ0_ : ----IKDQIVFIDGD------SEKVLAIYMVHPT-------KES-EIKPKKVFN~~~~~~~~~~~~ : 305
sp_P65856_ : ----NEADTLAAF-Y------ENRVIAILE---K-------RGN-EFKPHKVLL~~~~~~~~~~~~ : 294
sp_Q8CWM3_ : ADS-CVGVTFAT-----------ACVGVGI----------SGG-DRLHPKVVLKE~~~~~~~~~~ : 287
sp_O83859_ : -TR-RPGERALVF-S------EGRCLGLIRKA---------ANG--FSYDAVFCTE~~~~~~~~~~ : 366
sp_P45142_ : ----S-G-QVRLFSA------ENLFLGVLNRR---------E-Y-Y-SPTTINYTIRITSLPF-L~ : 312


sp_P62190_ : ~~~~~ :   -
sp_P60340_ : ~~~~~ :   -
sp_Q57612_ : RKK~~ : 336
sp_Q9Z8L9_ : ~~~~~ :   -
sp_Q8ZBC4_ : NKK~~ : 324
sp_Q6F7I5_ : ~~~~~ :   -
sp_A3N007_ : ~~~~~ :   -
sp_A4SJR7_ : ~~~~~ :   -
sp_O66922_ : ~~~~~ :   -
sp_A1R520_ : QAK~~ : 311
sp_A1K7B7_ : ~~~~~ :   -
sp_Q6G5F5_ : ~~~~~ :   -
```

```
sp_Q1LSL0_ : ~~~~~ :   -
sp_Q8YEB5_ : ~~~~~ :   -
sp_P59876_ : ~~~~~ :   -
sp_P57456_ : ~~~~~ :   -
sp_Q8K9H3_ : ~~~~~ :   -
sp_Q89AF6_ : ~~~~~ :   -
sp_Q62KL1_ : ~~~~~ :   -
sp_Q482T7_ : ~~~~~ :   -
sp_C3PH13_ : ~~~~~ :   -
sp_P60343_ : ~~~~~ :   -
sp_Q8FPB3_ : ~~~~~ :   -
sp_A4QEY6_ : ~~~~~ :   -
sp_Q4JV56_ : ~~~~~ :   -
sp_A9KBM3_ : ~~~~~ :   -
sp_Q4AAX7_ : ~~~~~ :   -
sp_Q0TCU3_ : ~~~~~ :   -
sp_P60341_ : ~~~~~ :   -
sp_Q7MBB8_ : ~~~~~ :   -
sp_Q4QK43_ : ~~~~~ :   -
sp_A5UBU2_ : ~~~~~ :   -
sp_A5UF28_ : ~~~~~ :   -
sp_Q0I3P3_ : ~~~~~ :   -
sp_Q2SML1_ : ~~~~~ :   -
sp_C4K3E8_ : ~~~~~ :   -
sp_Q5QTZ0_ : ~~~~~ :   -
sp_Q5X1C5_ : ~~~~~ :   -
sp_Q5ZRV6_ : ~~~~~ :   -
sp_Q5WT38_ : ~~~~~ :   -
sp_Q65SL2_ : ~~~~~ :   -
sp_P62189_ : ~~~~~ :   -
sp_A1KMD6_ : ~~~~~ :   -
sp_Q73VW3_ : ~~~~~ :   -
sp_A0PQD9_ : ~~~~~ :   -
sp_B4RMB1_ : ~~~~~ :   -
sp_Q9JTX5_ : ~~~~~ :   -
```

```
sp_Q820P2_ : SVDLQ : 308
sp_Q3J9B8_ : ~~~~~ :   -
sp_Q5YSE0_ : ~~~~~ :   -
sp_Q8YWR1_ : ~~~~~ :   -
sp_Q9CMQ7_ : ~~~~~ :   -
sp_Q7MAY1_ : ~~~~~ :   -
sp_Q6LUJ0_ : ~~~~~ :   -
sp_Q6A7P0_ : ~~~~~ :   -
sp_Q1IF41_ : ~~~~~ :   -
sp_Q3IJ75_ : ~~~~~ :   -
sp_Q4ZNR4_ : ~~~~~ :   -
sp_Q1MN41_ : ~~~~~ :   -
sp_Q21H63_ : ~~~~~ :   -
sp_Q57JI1_ : ~~~~~ :   -
sp_Q5PLB2_ : ~~~~~ :   -
sp_Q8ZLT2_ : ~~~~~ :   -
sp_B8CKH5_ : ~~~~~ :   -
sp_Q31W45_ : ~~~~~ :   -
sp_Q82K56_ : ~~~~~ :   -
sp_B1VYN2_ : ~~~~~ :   -
sp_Q8DU15_ : ~~~~~ :   -
sp_Q04KA7_ : ~~~~~ :   -
sp_Q5M4G2_ : ~~~~~ :   -
sp_Q47RU8_ : ~~~~~ :   -
sp_Q9KU78_ : ~~~~~ :   -
sp_Q5E7L3_ : ~~~~~ :   -
sp_Q87M04_ : ~~~~~ :   -
sp_A7Z4T7_ : ~~~~~ :   -
sp_Q9KA80_ : ~~~~~ :   -
sp_Q0SM48_ : ~~~~~ :   -
sp_Q5HU02_ : ~~~~~ :   -
sp_P58063_ : ~~~~~ :   -
sp_B0BB81_ : ~~~~~ :   -
sp_A5N845_ : ~~~~~ :   -
sp_Q0TPS0_ : ~~~~~ :   -
```

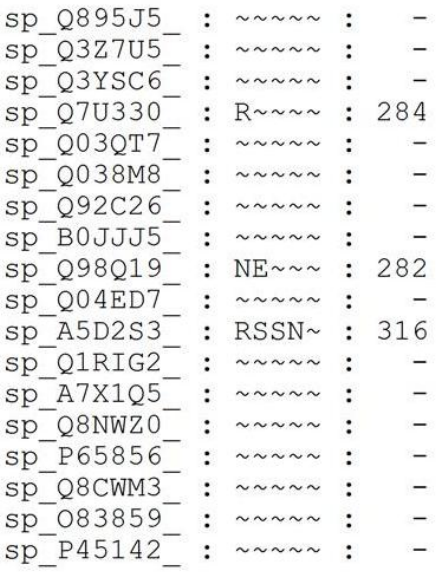


Figure A-1: Alignment of TruB bacterial sequeneces. 100 bacterial TruB sequences aligned using GenDoc. Black shows 100% conserved, dark grey shows mostly conserved, light grey marginally conserved and white is not conserved. Species number and name are in order as they appear in the alignment below.

P62190|TRUB_MYCTU	*Mycobacterium tuberculosis*
P60340|TRUB_ECOLI	*Escherichia coli (strain K12)*
Q57612|TRUB_METJA	*Methanocaldococcus jannaschii*
Q9Z8L9|TRUB_CHLPN	*Chlamydia pneumoniae*
Q8ZBC4|TRUB_YERPE	*Yersinia pestis*
Q6F7I5|TRUB_ACIAD	*Acinetobacter sp. (strain ADP1)*
A3N007|TRUB_ACTP2	*Actinobacillus pleuropneumoniae serotype 5b (strain L20)*

A4SJR7|TRUB_AERS4	*Aeromonas salmonicida (strain A449)*
O66922|TRUB_AQUAE	*Aquifex aeolicus*
A1R520|TRUB_ARTAT	*Arthrobacter aurescens*
A1K7B7|TRUB_AZOSB	*Azoarcus sp. (strain BH72)*
Q6G5F5|TRUB_BARHE	*Bartonella henselae*
Q1LSL0|TRUB_BAUCH	*Baumannia cicadellinicola subsp. Homalodisca coagulata*
Q8YEB5|TRUB_BRUME	*Brucella melitensis biotype 1 (strain 16M / ATCC 23456 / NCTC 10094)*
P59876|TRUB_BRUSU	*Brucella suis biovar 1 (strain 1330)*
P57456|TRUB_BUCAI	*Buchnera aphidicola subsp. Acyrthosiphon pisum (strain APS)*
Q8K9H3|TRUB_BUCAP	*Buchnera aphidicola subsp. Schizaphis graminum*
Q89AF6|TRUB_BUCBP	*Buchnera aphidicola subsp. Baizongia pistaciae (strain Bp)*
Q62KL1|TRUB_BURMA	*Burkholderia mallei*
Q482T7|TRUB_COLP3	*Colwellia psychrerythraea (strain 34H / ATCC BAA-681)*
C3PH13|TRUB_CORA7	*Corynebacterium aurimucosum (strain ATCC 700975 / DSM 44827 / CN-1)*
P60343|TRUB_CORDI	*Corynebacterium diphtheriae*
Q8FPB3|TRUB_COREF	*Corynebacterium efficiens*
A4QEY6|TRUB_CORGB	*Corynebacterium glutamicum (strain R)*
Q4JV56|TRUB_CORJK	*Corynebacterium jeikeium (strain K411)*
A9KBM3|TRUB_COXBN	*Coxiella burnetii (strain Dugway 5J108-111)*
Q4AAX7|TRUB_COXBU	*Coxiella burnetii*
Q0TCU3|TRUB_ECOL5	*Escherichia coli O6:K15:H31 (strain 536 / UPEC)*
P60341|TRUB_ECOL6	*Escherichia coli O6*
Q7MBB8|TRUB_GLOVI	*Gloeobacter violaceus*
Q4QK43|TRUB_HAEI8	*Haemophilus influenzae (strain 86-028NP)*
A5UBU2|TRUB_HAEIE	*Haemophilus influenzae (strain PittEE)*
A5UF28|TRUB_HAEIG	*Haemophilus influenzae (strain PittGG)*
Q0I3P3|TRUB_HAES1	*Haemophilus somnus (strain 129Pt)*

Q2SML1|TRUB_HAHCH	*Hahella chejuensis (strain KCTC 2396)*
C4K3E8|TRUB_HAMD5	*Hamiltonella defensa subsp. Acyrthosiphon pisum (strain 5AT)*
Q5QTZ0|TRUB_IDILO	*Idiomarina loihiensis (strain ATCC BAA-735 / DSM 15497 / L2-TR)*
Q5X1C5|TRUB_LEGPA	*Legionella pneumophila (strain Paris)*
Q5ZRV6|TRUB_LEGPH	*Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513)*
Q5WT38|TRUB_LEGPL	*Legionella pneumophila (strain Lens)*
Q65SL2|TRUB_MANSM	*Mannheimia succiniciproducens (strain MBEL55E)*
P62189|TRUB_MYCBO	*Mycobacterium bovis*
A1KMD6|TRUB_MYCBP	*Mycobacterium bovis (strain BCG / Pasteur 1173P2)*
Q73VW3|TRUB_MYCPA	*Mycobacterium paratuberculosis*
A0PQD9|TRUB_MYCUA	*Mycobacterium ulcerans (strain Agy99)*
B4RMB1|TRUB_NEIG2	*Neisseria gonorrhoeae (strain NCCP11945)*
Q9JTX5|TRUB_NEIMA	*Neisseria meningitidis serogroup A*
Q820P2|TRUB_NITEU	*Nitrosomonas europaea*
Q3J9B8|TRUB_NITOC	*Nitrosococcus oceani (strain ATCC 19707 / NCIMB 11848)*
Q5YSE0|TRUB_NOCFA	*Nocardia farcinica*
Q8YWR1|TRUB_NOSS1	*Nostoc sp. (strain PCC 7120 / UTEX 2576)*
Q9CMQ7|TRUB_PASMU	*Pasteurella multocida (strain Pm70)*
Q7MAY1|TRUB_PHOLL	*hotorhabdus luminescens subsp. laumondii (strain TT01)*
Q6LUJ0|TRUB_PHOPR	*Photobacterium profundum*
Q6A7P0|TRUB_PROAC	*Propionibacterium acnes (strain KPA171202 / DSM 16379)*
Q1IF41|TRUB_PSEE4	*Pseudomonas entomophila (strain L48)*
Q3IJ75|TRUB_PSEHT	*Pseudoalteromonas haloplanktis (strain TAC 125)*
Q4ZNR4|TRUB_PSEU2	*Pseudomonas syringae pv. syringae (strain B728a)*
Q1MN41|TRUB_RHIL3	*Rhizobium leguminosarum bv. viciae (strain 3841)*
Q21H63|TRUB_SACD2	*Saccharophagus degradans (strain 2-40 / ATCC 43961 / DSM 17024)*
Q57JI1|TRUB_SALCH	*Salmonella choleraesuis*

Q5PLB2|TRUB_SALPA	*Salmonella paratyphi*
Q8ZLT2|TRUB_SALTY	*Salmonella typhimurium*
B8CKH5|TRUB_SHEPW	*Shewanella piezotolerans (strain WP3 / JCM 13877)*
Q31W45|TRUB_SHIBS	*Shigella boydii serotype 4 (strain Sb227)*
Q82K56|TRUB_STRAW	*Streptomyces avermitilis*
B1VYN2|TRUB_STRGG	*Streptomyces griseus subsp. griseus (strain JCM 4626 / NBRC 13350)*
Q8DU15|TRUB_STRMU	*Streptococcus mutans*
Q04KA7|TRUB_STRP2	*Streptococcus pneumoniae serotype 2 (strain D39 / NCTC 7466)*
Q5M4G2|TRUB_STRT2	*Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311)*
Q47RU8|TRUB_THEFY	*Thermobifida fusca (strain YX)*
Q9KU78|TRUB_VIBCH	*Vibrio cholerae*
Q5E7L3|TRUB_VIBF1	*Vibrio fischeri (strain ATCC 700601 / ES114)*
Q87M04|TRUB_VIBPA	*Vibrio parahaemolyticus*
A7Z4T7|TRUB_BACA2	*Bacillus amyloliquefaciens (strain FZB42)*
Q9KA80|TRUB_BACHD	*Bacillus halodurans*
Q0SM48|TRUB_BORAP	*Borrelia afzelii (strain PKo)*
Q5HU02|TRUB_CAMJR	*Campylobacter jejuni (strain RM1221)*
P58063|TRUB_CAUCR	*Caulobacter crescentus*
B0BB81|TRUB_CHLTB	*Chlamydia trachomatis serovar L2b (strain UCH-1/proctitis)*
A5N845|TRUB_CLOK5	*Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680)*
Q0TPS0|TRUB_CLOP1	*Clostridium perfringens (strain ATCC 13124 / NCTC 8237 / Type A)*
Q895J5|TRUB_CLOTE	*Clostridium tetani*
Q3Z7U5|TRUB_DEHE1	*Dehalococcoides ethenogenes (strain 195)*
Q3YSC6|TRUB_EHRCJ	*Ehrlichia canis (strain Jake)*
Q7U330|TRUB_HELHP	*Helicobacter hepaticus*
Q03QT7|TRUB_LACBA	*Lactobacillus brevis (strain ATCC 367 / JCM 1170)*
Q038M8|TRUB_LACC3	*Lactobacillus casei (strain ATCC 334)*

Q92C26|TRUB_LISIN	*Listeria innocua*
B0JJJ5|TRUB_MICAN	*Microcystis aeruginosa (strain NIES-843)*
Q98Q19|TRUB_MYCPU	*Mycoplasma pulmonis (strain UAB CTIP)*
Q04ED7|TRUB_OENOB	*Oenococcus oeni (strain BAA-331 / PSU-1)*
A5D2S3|TRUB_PELTS	*Pelotomaculum thermopropionicum (strain DSM 13744 / JCM 10971 / SI)*
Q1RIG2|TRUB_RICBR	*Rickettsia bellii (strain RML369-C)*
A7X1Q5|TRUB_STAA1	*Staphylococcus aureus (strain Mu3 / ATCC 700698)*
Q8NWZ0|TRUB_STAAW	*Staphylococcus aureus (strain MW2)*
P65856|TRUB_STRA3	*Streptococcus agalactiae serotype III*
Q8CWM3|TRUB_THEEB	*Thermosynechococcus elongatus (strain BP-1)*
O83859|TRUB_TREPA	*Treponema pallidum (strain Nichols)*
P45142|TRUB_HAEIN	*Haemophilus influenzae (strain ATCC 51907 / DSM 11121 / KW20 / Rd)*