Identification of pseudouridine methyltransferase in *Escherichia coli*

RYA ERO,¹ LAURI PEIL,² AIVAR LIIV,¹ and JAANUS REMME¹

¹Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia ²Institute of Technology, University of Tartu, Tartu, Estonia

ABSTRACT

In ribosomal RNA, modified nucleosides are found in functionally important regions, but their function is obscure. Stem–loop 69 of *Escherichia coli* 23S rRNA contains three modified nucleosides: pseudouridines at positions 1911 and 1917, and N3 methylpseudouridine ($m^3\Psi$) at position 1915. The gene for pseudouridine methyltransferase was previously not known. We identified *E. coli* protein YbeA as the methyltransferase methylating Ψ 1915 in 23S rRNA. The *E. coli* ybeA gene deletion strain lacks the N3 methylation at position 1915 of 23S rRNA as revealed by primer extension and nucleoside analysis by HPLC. Methylation at position 1915 is restored in the *ybeA* deletion strain when recombinant YbeA protein is expressed from a plasmid. In addition, we show that purified YbeA protein is able to methylate pseudouridine in vitro using 70S ribosomes but not 50S subunits from the *ybeA* deletion strain as substrate. Pseudouridine is the preferred substrate as revealed by the inability of YbeA to methylate uridine at position 1915. This shows that YbeA is acting at the final stage during ribosome assembly, probably during translation initiation. Hereby, we propose to rename the YbeA protein to RlmH according to uniform nomenclature of RNA methyltransferases. RlmH belongs to the SPOUT superfamily of methyltransferases. RlmH was found to be well conserved in bacteria, and the gene is present in plant and in several archaeal genomes. RlmH is the first pseudouridine specific methyltransferase identified so far and is likely to be the only one existing in bacteria, as $m^3\Psi$ 1915 is the only methylated pseudouridine in bacteria described to date.

Keywords: post-transcriptional modifications; ribosome; methyltransferase; ybeA; RlmH; pseudouridine; 23S rRNA

INTRODUCTION

The physiological importance of post-transcriptional modifications in rRNA is largely enigmatic. Although a ribosome lacking post-transcriptional modifications is able to synthesize peptides in vitro (Krzyzosiak et al. 1987; Cunningham et al. 1991; Green and Noller 1999; Khaitovich et al. 1999), the conservation and clustering of modified nucleosides in functionally important regions of the ribosome suggest that they might be important for efficient translation, rRNA folding, ribosome assembly, or stability of ribosomes in vivo (Noller and Woese 1981; Brimacombe et al. 1993; Ofengand and Fournier 1998; Decatur and Fournier 2002; Xu et al. 2008). Indeed, several rRNA modifications have been shown to be important for 30S and 50S assembly and ribosome functioning (Igarashi et al. 1981; Green and Noller 1996; Gustafsson and Persson 1998; Caldas et al. 2000), and a number of additional modifications provide advantages under particular conditions, such as conferring resistance against ribosometargeting antibiotics (Cundliffe 1989; Weisblum 1995; Mann et al. 2001; Toh et al. 2008). However, the possible functional roles of the vast majority of modified nucleosides in rRNA remain unknown. A useful tool for studying functional roles of modified nucleosides is to construct the bacterial strains carrying mutant genes for modification enzymes.

Thirty-six modified nucleosides are found at precisely determined locations of *Escherichia coli* K12 strain ribosomes, 11 in 16S rRNA and 25 in 23S rRNA. Pseudouridine is found at 11 positions, and various ribose and base methylations are found at 24 positions across ribosomal rRNA (Ofengand and Del Campo 2004; Andersen and Douthwaite 2006; 3D Ribosomal Modification Maps database, http://people.biochem.umass.edu/fournierlab/3dmod map/main.php). Uridine at position 1915 of 23S rRNA is both isomerized to pseudouridine and methylated ($m^{3}\Psi$).

Reprint requests to: Jaanus Remme, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010, Tartu, Estonia; e-mail: jremme@ebc.ee; fax: 372 7420286.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.1186608.

In addition to pseudouridines and various methylated residues, one dihydrouridine (hU2449) and one 2-thiocytidine (s^2 C2501) are found in 23S rRNA (Andersen et al. 2004; for review, see Ofengand and Del Campo 2004). Most of the genes encoding enzymes that modify rRNA have been identified. Identification of remaining genes encoding modification enzymes is a prerequisite for the use of genetic and biochemical tools for functional studies on the modified nucleosides.

Stem-loop 69 (H69) of E. coli 23S rRNA forms a distinct structure at the interface side of 50S subunit. H69 was the first RNA structural element that was identified as the RNA component of an intersubunit bridge (Mitchell et al. 1992), later named B2a (Gabashvili et al. 2000; Yusupov et al. 2001). In addition, H69 has been shown to participate in several ribosomal functions: H69 contacts A-site tRNA and translation factors; it is functioning during ribosome assembly and translation termination (Agrawal et al. 2004; Ali et al. 2006; Hirabayashi et al. 2006). The loop region of H69 contains several post-transcriptional modifications in all known large subunit RNAs (Ofengand et al. 2001). Pseudouridine (Ψ) is found at positions 1911, 1915, and 1917, all of which are synthesized by pseudouridine synthase RluD (Huang et al. 1998; Raychaudhuri et al. 1998). Pseudouridines of H69 were shown to be important during translation termination (Ejby et al. 2007). In addition, the pseudouridine residue at position 1915 of E. coli 23S rRNA is further methylated to form $m^{3}\Psi$ (Fig. 1; Kowalak et al. 1996). The methyltransferase responsible for this modification was previously unknown, and the functional role of $m^{3}\Psi$ modification has not been explored.

According to current knowledge, $m^3\Psi1915$ is the only methylated pseudouridine in bacterial RNAs (Rozenski et al. 1999; RNA Modification Database, http://library.med.utah. edu/RNAmods/), hence making the corresponding $m^3\Psi$ methyltransferase a likely candidate for the only pseudouridine-specific methyltransferase in bacteria. In eukaryotes, $m^1\Psi$, Ψ m, and $m^1acp^3\Psi$ are found (Gray and Keddy 1974; Maden and Salim 1974; Saponara and Enger 1974; Maden et al. 1975; Brand et al. 1978). So far, no pseudouridinespecific methyltransferases have been described.



FIGURE 1. Secondary structure of *E. coli* 23S rRNA stem–loop 69 and the structural formula of $m^3\Psi$. (*Left*) Stem–loop 69 contains three post-transcriptional modifications: two pseudouridines (Ψ) and one 3-methylpseudouridine ($m^3\Psi$) located at position 1915 according to standard *E. coli* 23S rRNA numeration. (*Right*, bold) The methyl group of the $m^3\Psi$ -modified base.

rRNA methyltransferases are enzymes that catalyze the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to an acceptor residue in rRNA. Five classes of methyltransferases with structurally distinct folds have evolved convergently to bind the cofactor SAM and to perform the methyltransferase reaction (for review, see Ofengand and Del Campo 2004). There is no significant sequence conservation across all five methyltransferase classes. All known rRNA methyltransferases belong to class I or class IV. The genes corresponding to 16 out of the 24 rRNA methyltransferases predicted in E. coli have been identified (Andersen and Douthwaite 2006; Sergiev et al. 2007, 2008; Toh et al. 2008), and the majority of them belong to class I, characterized by the presence of a common, conserved Rossmann fold SAM binding domain (Schubert et al. 2003; for review, see Ofengand and Del Campo 2004). Much less conservation is noticed at the sequence level, where only a few conserved motifs are present, most of them being a part of the SAM binding region (Fauman et al. 1999). E. coli Gm2251 methyltransferase RlmB and m³U1498 methyltransferase RsmE are class IV methyltransferases and belong to the superfamily of proteins characterized by an intriguing α/β knot structure (Anantharaman et al. 2002; Forouhar et al. 2003; Schubert et al. 2003; Basturea et al. 2006; Basturea and Deutscher 2007). Recently, Tkaczuk et al. (2007) proposed to include the whole group of proteins with the α/β knot domain to the SPOUT superfamily of methyltransferases, regardless of the level of their functional characterization and the degree of sequence similarity to the SpoU and TrmD methyltransferases that historically gave the name to the SPOUT (SpoU-TrmD) family.

In this study, we report the identification of the methyltransferase gene (*ybeA*) responsible for the m³ Ψ 1915 methylation in *E. coli* 23S rRNA. The gene was identified by analysis of several putative RNA methyltransferase deletion strains and was confirmed by gene complementation, overexpression, and in vitro characterization of the purified YbeA protein. The enzyme has high specificity for Ψ 1915 and does not methylate U1915. YbeA belongs to a SPOUT superfamily of methyltransferases and is the first pseudouridine-specific methyltransferase to be identified. In agreement with the accepted convention (Ofengand and Del Campo 2004; Andersen and Douthwaite 2006), we propose to rename the YbeA protein as RlmH (rRNA large subunit methyltransferase gene H).

RESULTS

Identification of the gene associated with the methylation of Ψ 1915 of *E. coli* 23S rRNA

The loop region of H69 in 23S rRNA contains three modified uridines. In addition to the methylated pseudouridine $(m^{3}\Psi)$ at position 1915, two additional pseudouridines are located at positions 1911 and 1917 (Fig. 1). Pseudouridines at positions 1911, 1915, and 1917 are synthesized by the RluD protein, while the methyltransferase responsible for $m^{3}\Psi$ 1915 synthesis is not identified yet (Kowalak et al. 1996; Raychaudhuri et al. 1998).

In order to identify the gene involved in methylation of Ψ 1915, for analysis we selected 11 genes predicted to encode RNA methyltransferases. Eleven putative RNA methyltransferase deletion strains (Fig. 2) were selected from the pool of *E. coli* clones ("KEIO" collection) containing single gene knockouts of all non-essential genes in the K-12 strain BW25113 background (Baba et al. 2006).

The methylation status of Ψ 1915 was initially screened by primer extension analysis of total rRNA isolated from 11 Keio collection knockout strains and the wild-type MG1655 strain (Fig. 2). Total ribosomal RNA was isolated from exponentially growing cells, and the methylation at position 1915 of 23S rRNA was monitored. Methylation at the N3 position of the pseudouridine (m³ Ψ) perturbs Watson–Crick base-pairing and results in a strong reverse transcriptase stop.

In the case of *E. coli* MG1655 wild-type strain (Fig. 2, lane 12) and 10 out of 11 putative RNA methyltransferase knockout strains analyzed, the strong stop signal corresponding to the position 1915 of 23S rRNA was observed (Fig. 2, lanes 1–8,10,11), indicating the presence of methylation. In contrast, no stop signal was detected in rRNA extracted from strain JW0631 (Fig. 2, lane 9), referring to the absence of methylation at position Ψ 1915 of 23S rRNA. Strain JW0631 has the kanamycin-resistance cassette inserted into the putative RNA methyltransferase gene *ybeA* and is hence renamed Δ ybeA.



FIGURE 2. Analysis of 1915 region of *E. coli* 23S rRNA by primer extension. rRNA was isolated from various putative RNA methyltransferase gene knockout strains from (lanes 1–11) the Keio collection, (lanes 12,17) *E. coli* wild-type MG1655 strain, (lane 18) the MG Δ ybeA strain, and (lane 19) the Δ ybeA/pQE30-ybeA strain. MG Δ ybeA is the *ybeA* gene knockout in the MG1655 genetic background, and the Δ ybeA/pQE30-ybeA strain is the Δ ybeA strain transformed with a plasmid (pQE30-ybeA) expressing YbeA protein. A strong stop signal indicates the presence of methylation at position 1915. (Lanes 13–16) The sequence of the corresponding region of the *E. coli* 23S rRNA gene.

A ybeA gene knockout strain in the E. coli wild-type MG1655 background was constructed to eliminate the possibility that unknown second-site mutations in the JW0631 strain are responsible for the loss of the methylation at position 1915 of 23S rRNA. The MG1655ybeA:: kan strain was generated by transducing E. coli strain MG1655 with P1 phage lysate derived from the phageinfected Δ ybeA strain (JW0631); the resulting knockout strain was named MG Δ ybeA. Deletion of the *ybeA* gene did not have a detectable effect on the bacterial growth rate in rich medium at 37°C (data not shown). Primer extension analysis of rRNA demonstrated that the MG Δ ybeA strain, like the original JW0631 strain, lacked the methylation at position 1915 of 23S rRNA (Fig. 2, lane 18), hence making the *ybeA* gene a potential candidate for encoding $m^3\Psi 1915$ methyltransferase.

ybeA gene encodes $m^3 \Psi$ methyltransferase

To test the possibility that the *ybeA* gene encodes a protein responsible for the formation of m³ Ψ 1915 of 23S rRNA, gene complementation analysis was conducted. The *ybeA* gene was PCR-amplified using genomic DNA prepared from the *E. coli* MG1655 strain as a template, and cloned into a pQE-30 expression vector, introducing an N-terminal His-tag. The recombinant YbeA protein was constitutively expressed in the Δ ybeA strain, and the methylation pattern of Ψ 1915 was monitored by primer extension analysis. Figure 2 shows a strong reverse transcriptase stop signal at position 1915 of 23S rRNA when the Δ ybeA strain (JW0631) is transformed with the plasmid (pQE30/ybeA) expressing N-terminally His-tagged YbeA protein (Fig. 2,

lane 19). This finding suggests that the *ybeA* gene is required for the introduction of the methyl group to Ψ 1915.

To verify the loss of $m^3 \Psi 1915$ methvlation in the ybeA deletion strain, RP-HPLC analysis of nucleosides was conducted. 70S ribosomes were isolated from the wild-type MG1655 strain, ybeA deletion strain (MG Δ ybeA), and ybeA deletion strain complemented with a plasmid encoding YbeA protein (MG∆ybeA/pQE30-ybeA). Phenol extracted RNA was used for oligonucleotide-directed RNase H excision of the 23S rRNA fragments corresponding to nucleotides 1778-1921. Nucleoside composition of RNA fragments was analyzed by RP-HPLC using a modified multilinear gradient (Gehrke and Kuo 1989) on a Supelcosil LC-18-S HPLC column at 30°C. The retention time of $m^{3}\Psi$ under the conditions used is 11.73 min (Gehrke and Kuo 1989).

HPLC chromatograms of the wild-type MG1655 strain revealed a nucleoside with the retention time 11.73 min (Fig. 3A), contrary to the nucleoside composition of the *ybeA* knockout strain (MG Δ ybeA), where no signal was



FIGURE 3. HPLC analysis of nucleoside composition of 23S rRNA fragment corresponding to positions 1778–1921. The 23S rRNA fragment was isolated from 70S ribosomes of *E. coli* strains (*A*) MG1655, (*B*) MG Δ ybeA, (*C*) MG Δ ybeA/pQE30-YbeA, and (*D*) MG Δ ybeA strain ribosomes treated with YbeA in vitro. Peaks corresponding to three standard nucleosides (C, U, and G) and two modified nucleosides, pseudouridine (Ψ) and 3-methylpseudouridine ($m^{3}\Psi$), are indicated.

detected in the corresponding region (Fig. 3B). When the *ybeA* deletion strain was complemented with a plasmid expressing recombinant YbeA protein, a strong signal (retention time 11.73 min) corresponding to the m³ Ψ appeared on HPLC chromatograms (Fig. 3C). These data indicate that there is a causative relationship between the presence of the *ybeA* gene and posttranscriptional methylation of Ψ 1915 in *E. coli* ribosomes, making YbeA protein the main candidate for m³ Ψ methyltransferase.

Table 1 represents a relative nucleoside composition of each RNA fragment compared to the wild-type reference probe. Interestingly, the absence of YbeA protein leads also to the reduced levels of H69 pseudouridines and m^2G1835 in the 70S ribosomes.

YbeA protein has Ψ 1915-specific methyltransferase activity in vitro

In order to test the functional properties of the ybeA gene product, in vitro methylation assays were performed. Recombinant YbeA protein with N-terminal His-tag was purified by metal affinity chromatography. YbeA-dependent incorporation of [14C]-methyl groups into ribosomes was monitored by TCA precipitation and scintillation counting. The purified YbeA protein was able to catalyze transfer of [¹⁴C]-methyl groups from [¹⁴C]-SAM to the 70S ribosomes isolated from the MG Δ ybeA strain but not to the ribosomes purified from the wild-type MG1655 strain (Fig. 4). About 70%-75% of the ribosomes isolated from the MG Δ ybeA strain were methylated giving an average of 0.7 methyl groups per ribosome (Fig. 4). Only 0.05 methyl groups were incorporated per 23S rRNA when 50S subunits were used as substrate (Fig. 4), indicating that YbeA is specific for fully assembled 70S ribosomes. Specificity of the methylation reaction was verified by HPLC analysis. An aliquot of ribosomes from in vitro methylation assay was used for RNA fragment preparation and nucleoside composition determination. Ribosomal rRNA fragment isolated from *ybeA* deletion strain MG Δ ybeA did not show any signal in the region corresponding to the $m^{3}\Psi$ (retention time 11.73 min) (Fig. 3B). Incubation of ribosomes with purified YbeA protein leads to appearance of strong signal at 11.73 min on the chromatogram (Fig. 3D). Approximately 0.65 methyl groups were incorporated per 70S ribosome according to the chromatographic analysis, which is in good agreement with radioactivity incorporation data (Figs. 3, 4; Table 1).

To test whether the uridine or pseudouridine at position 1915 is a substrate for YbeA, 70S ribosomes from the Δ ybeA/ Δ rluD double knockout strain were isolated, and the efficiency of YbeA-directed methyl group incorporation in vitro was tested. The purified YbeA protein was unable to incorporate methyl groups into 70S ribosomes isolated from the Δ ybeA/ Δ rluD double-knockout strain lacking

TABLE 1. Quantification of nucleosides in 23S rRNA fragment 1778–1921											
	С	U	G	А	Ψ	${ m m}^{3}\Psi$	m ² G				
MG1655 MG∆ <i>ybeA</i> MG∆ <i>ybeA</i> /pQE30-YbeA MG∆ <i>ybeA</i> +YbeA ^a	$\begin{array}{rrrrr} 100 \ \pm \ 0.9 \\ 98 \ \pm \ 0.7 \\ 100 \ \pm \ 1.0 \\ 98 \ \pm \ 0.6 \end{array}$	$\begin{array}{rrrr} 100 \ \pm \ 2.7 \\ 103 \ \pm \ 0.9 \\ 105 \ \pm \ 1.0 \\ 103 \ \pm \ 0.4 \end{array}$	$\begin{array}{rrrr} 100 \ \pm \ 2.5 \\ 103 \ \pm \ 1.7 \\ 105 \ \pm \ 1.0 \\ 102 \ \pm \ 1.2 \end{array}$	$\begin{array}{rrrr} 100 \ \pm \ 0.5 \\ 97 \ \pm \ 1.3 \\ 94 \ \pm \ 1.0 \\ 98 \ \pm \ 1.2 \end{array}$	$\begin{array}{rrrr} 100 \ \pm \ 2.7 \\ 79 \ \pm \ 7.0 \\ 64 \ \pm \ 1.0 \\ 85 \ \pm \ 2.1 \end{array}$	$\begin{array}{rrrr} 100 \ \pm \ 3.0 \\ 0 \ \pm \ 0.2 \\ 50 \ \pm \ 1.0 \\ 63 \ \pm \ 1.6 \end{array}$	$100 \pm 3.7 \\ 83 \pm 5.6 \\ 64 \pm 1.0 \\ 83 \pm 2.8$				

70S ribosomes were isolated, nucleosides of 23S rRNA fragment 1778–1921 were analyzed, and molar amounts of nucleosides were calculated as described in Materials and Methods. The relative molar amounts of nucleosides of three independent ribosome preparations are presented as relation to the 100% value of wild-type ribosomes.

^a70S ribosomes from MG $\Delta y beA$ strain treated with purified YbeA protein as described in Materials and Methods.

both *ybeA* and *rluD* genes (Fig. 4). RluD protein was shown to synthesize 1911, 1915, and 1917 pseudouridines in vitro using purified 70S ribosomes as substrates, albeit at reduced efficiency as compared to the free 50S subunits (Vaidyanathan et al. 2007). When Δ ybeA/ Δ rluD 70S ribosomes were pre-incubated with purified RluD protein, they became substrates for YbeA-directed methylation (Fig. 4). These results show that the pseudouridine rather than uridine at position 1915 is the substrate for YbeA.

Taking together, we have shown that the *ybeA* gene encodes a methyltransferase specific for methylation of Ψ 1915 of 23S rRNA. Therefore, we propose the *ybeA* gene to be renamed *rlmH* (<u>r</u>ibosomal <u>l</u>arge subunit <u>m</u>ethyltransferase H). The same result was found by our accompanying paper in this issue (Purta et al. 2008) using different methods.

ybeA is a widely conserved gene

YbeA belongs to the COG1576 cluster of SPOUT superfamily methyltransferases (Tkaczuk et al. 2007). Clusters of Orthologous Groups, also known as COGs, is a systematic grouping of gene families from completely sequenced genomes (Tatusov et al. 1997, 2003). Multiple sequence alignment of proteins from the COG1576 cluster of the SPOUT methyltransferase superfamily is shown in Figure 5. From each class of organisms with an annotated putative COG1576 member, one representative sequence was taken (31 sequences in total) and aligned with MUSCLE (Edgar 2004). Sequences for alignment were selected from COG (Tatusov et al. 2003) and RefSeq (Pruitt et al. 2007) databases as of March 2008.

Putative members of COG1576 were found in all three domains of life (Fig. 5). In eukaryotes, they are present in genomes of green plants (*Viridiplantae*). In *Arabidopsis thaliana*, the gene was annotated as a chloroplast protein. In archaea, the corresponding gene was found only in the phylum *Euryarchaeota*, and in bacteria in phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Fusobacteria*, *Lentisphaerae*, *Proteobacteria*, *Synergistetes*, *Thermotogae*, and *Verrucomicrobia* (Fig. 5; Supplemental Fig. 1; Supplemental Table 1). The taxonomy represented here is based on The NCBI Entrez Taxonomy Homepage (http://www.ncbi.nlm.nih. gov/sites/entrez?db=taxonomy) as of March 2008. Interestingly, while annotated the YbeA homolog is present in *Mycoplasma pulmonis* and *Ureaplasma urealyticum*, it is not present in *Mycoplasma pneumoniae* and *Mycoplasma genitalium*, and it is also missing from Rickettsiales among alpha-proteobacteria. The fact that parasites with small genomes have lost the *ybeA* gene is interesting and probably indicates the secondary nature of such loss. Whether it has some biological importance remains to be seen.

The sequences of YbeA homologs are well conserved, especially the C-terminal part that is thought to contain the catalytic center. Six amino acids were universally conserved among the aligned sequences presented in Figure 5—G103, G107, H129, E138, R142, and Y152 (*E. coli* numbering). Of those six, G103 and G107 are in the predicted SAM binding motif (Anantharaman et al. 2002), and R142 is of particular interest as a potential catalytic amino acid (Fig. 5). Structurally equivalent arginine is also conserved in TrmD (R154 in *E. coli*) and is shown to be part of its catalytic center (Elkins et al. 2003). The role of highly conserved arginine in the catalytic center of another SPOUT



FIGURE 4. Methyltransferase activity of YbeA in vitro. 70S ribosomes were isolated from *E. coli* wild-type MG1655, MG Δ ybeA, and Δ ybeA/ Δ rluD strains. Twenty-four picomoles of 70S ribosomes or 50S subunits was incubated with 100 μ M [¹⁴C]-SAM, 3 μ g of YbeA, and 1 μ g of RluD (if indicated). Incorporation of [¹⁴C]-methyl groups into TCA insoluble material was determined. The data presented are the average of three experiments.

	1ns5		α1 ,-	β1		α2		β2		α3	-
			-	51 61	71	8 1		101	111	121	
-	Esccol.gil16128619	1	MKLOLVAV	TKMPDWVOTGFT	EYLRRFPK	DMPFELI	EIPAGKRGKN	NADIKRIL	DKEGEOMLAA	AGKNRIVTLI	D 73
	Roslit.gi 163732522	1	MRIHIIAV	RLRAGPEKDLID	DYLTRFARSGR	PLGLGPARIV	EVEDRKNGG-	M	SNEAVLLRRA	L-PDGALIAVLI	D 73
	Burdol.gi 84362417	1	MKLYIVAV	HKMPGWIASGFD	EYAKRMPP	ELRIELF	EIKPELRSG	SRSAESVM	AAERQKIDAA	L-PKGARIVALI	D 74
	Camjej.gi 15791514	5	LQVNIFCI	KSDEFKTCSE	KYSKLISK	YATLKEI	NVFNKKIAL	AQNLNAIEAKK	SYEEAFMP	YKKGYCIALI	D 76
	Bdebac.gi 42525168	1	MKFILYNL	TAKEPWADEVSE	LYKKKISF	FIPFDIQ	SLKAKKSAR	EDADFKR	NEESELILKN	II-NSDDYVVLFI	D 73
	Marfer.gi 114776729	1	MKLRLLVV	RGSRELAEFESR	-FDQRLRP	FADFQVV	ELPEGRAKQ-	PVQRK	QEEAKQILT-	HAGKGFILFI	D 68
	Polirg.gi 88803363	1	MKIKLLVI	KTDHKSLLQLIE	EYKNRLKH	YIKFEIE	TIPDIKNVK	NLSEIQQK	EKEGTLILSK	L-QNTDQLVLLI	D 74
	Baccap.gi 154496020	2	LGIHVICV	KLKEKFYTDASA	EYAKRLGG	YCKFQLI	ELPEERLPD	NPSQAQIDAAL	RKEADAILQR	L-PKGAAVTAM	C 78
	Pedobacter_sp.gi 149279619	1	MKITLLVV	KTEDKYLIEGIE	KYLKRLKH	YIGFNLI	VIPDVKNTKI	NLTTDQQK	SKEADLILKÇ	L-NNTDTIVLLI	D 74
	Synechococcus_sp.gi 116074639	4	ARCRILAV	KVRKSWVQEGIA	LYLKRLPG	LTI3	ELRDSTP		EKEAEAIRQA	L-RPDEVPVIL	M 64
	Promar.gi 159903250	5	SRYRILAI	KTRKAWIQNGLN	LYIKRLPG	LTTT	ELKDSD		KKEAQSIRSS	SI-KINELLILL	T 65
	Bilden.gi 11/1/41054	1	PRITLITY	KVKEKILKDAIA	EISKRLGR	VEDUEU	EVADENTPE	HAGEGLERQIN	AREGERIARI		A //
	Rubxy1.g1 100004490	1	MDVDTAVT	KIL-GWAALGCE	UVEKEI DD	FCKDEVI	ETERUNDCC.	LGKGEAL	DEFTEDI TNE		D 70
р	Fuenuc gi 119703798	1	MNTNTTCT	KIKDKYINDGIA	EFSKRMTS	FVSLNT	FLEENKED	VILLET V	EKESLETIKC	TSKSNSYNTLL	D 73
	Deirad gill5805646	1	MRIHITTV	EPKLAYARSGWD	EYEKBLER	YHKVOVS	RV	SGKTC	OAESEATLKA	AGKSPLILL	D 63
	Akkmuc.gil166834251	1	MOFLILAA	KPSLGYAKEGVE	LYLNRLRP	FGKTELH	LVRDGSS		ODVSKRLLAA	SEGCLRIAM	D 63
	Lenara.gi 149195686	2	IRVDFLVI	KTKEKWIQTGID	KYLKRLKP	YANLTLF	ELPDQAVD-		KRKDSKILQA	V-SSRDLLILLI	D 66
	Acibac.gi 94967055	1	MKLRVVWI	KTKESAIQTLTG	EYLKRLSR	YVATEGI	EIGSEEA	LLKLK	DRPG	RTAPVLVLM	D 62
	Solusi.gi 116624738	1	MKIYLYFI	KPKDPHANAIAE	DFLARAGR	YSPCEME	REIRP		ERIDLW	TKHPTARKIFL	D 58
	Synwol.gi 114567973	1	MKYRIISV	RIRESFYLEGVR	EYLKRLGP	YTSIEL	DGLEEKIGP	RAGEKEIQAII	QKEAEKIRRW	IL-DKDEILVVL	D 77
	Bacsub.gi 16081075	1	MNINIVTI	KLKEKYLKQGIE	EYTKRLSA	YAKIDI	ELPDEKAPE	NLSDQDMKIIK	DKEGDRILSK	I-SPDAHVIAL	A 77
	Cloace.gi 15896772	1	MNITLITV	KLKEKYLKDAVN	EYAKRLQK	YCKLNI	ELQDEKTPE	KASLKEEKLIF	EKEGEKILSS	SI-KDNSYVVSM	D 77
-	Mycpul.gi 15828725	8	MKINIISV	TLSKEF-QVIFC	DYIKKINF	YSNVNL	IKIKEFKSNN	KDLII	KNETMAILER	(I-PKNSKVFLC:	S 77
4	Metmar.gi 159906004	1	MNITIISV	KIKEKYLSGAII	EYSKRISR	YSKLDI	EVADEKTPE	NPSDVEKSKLI	EKEAERILKY	L-KKDSFVITL	E 77
-	Methun.gi 88603833	2	VRIRIRAI	KIKESFIRDAIA	DYAKRMCS	FCQVEC	LEYPEAPVPD	THPST-IEMAC	VSEGEKLCSG	SI-DSLDYVIIL	D 77
	Drysat.gil15220214	42	IDIDUTTV	KKRSRGAQLIVE	EIKEKLGI	ICDIED.	LINSNPRLI.	ODVRVQV	EAEDMAMML	T-CEDDWART	D 114
61	Phypat gil168060349	1	MDMKLVTV	KUBCAALENVAL	EXIKETOR	VCPFFF	OL PDNPKNS	SDVRTOL	SSECERUMPS	ST-GSDDWVVVL	D 111 D 73
	Chlrei gil159475549	28	ALVOLTTV	SKENSKGATLFAL	ELLEKVOR	YAPVOS	INTERNER	PDPAVOR	ETEGEKVIKA	AL-DSRDLVVVI.	D 100
	Ostluc.gi1145344864	40	VPLVVVVV	KGGGAKVDDAVE	EYARRCAR	YAPFEEH	TVKONPKNV	KDTELOP	VHEGERVMRA	I-TARDYVVLL	D 112
					Dave dd ach a d	03W 14-44					
					Predicted	SAM DINGI	ng motii				
	lns5		β	3	α4 ,-	β4	-α5,-	β5			-
	1ns5		β: 131	3 141	α4 151	β4 161	-α5 - 171	β5 181	191	201	-
-	lns5	74	β: 131 IP- S KPWD	3 141 IPQLA-AELERWE	α4 151 LDGR-DVSLL	β4 161 GPECLSPA	-α5 171 CKAAAEQSWS	β5 181 LSALTLP P EL	191 RVLVA	201 AWS TTNHP	- R 154
-	lns5 Esccol.gi 16128619 Roslit.gi 163732522	74 74	β 131 IP-CKPWD ER-CRVES	141 IPQLA-AELERWK PDFA-GRLGGWF	α4 151 KLDGR-DVSLL RDQGRGDLA V	β4 161 GGPECLSPAC GADGIDPAI	-α5 171 CKAAAEQSWS LRAEADFALS	β5 181 LSALTLPHPLN FGKMVWPHMLN	191 VRVLVA SL VRVMLA L	201 AWS TTNHPY 1 AAS LAGAPY H	- R 154 R 155
-	lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417	74 74 75	β 131 IP-CKPWD ER-RVES ER-CRDWT	141 PQLA-AELERWY PDFA-GRLGGWF TMQLA-HALPGWQ	α4 151 CLDGR-DVSLL RDQGRGDLA V QQDGR-DVA M	β4 161 GGPECLSPAC GGADCIDPAI GGADCLDPEI	-α5 - 171 CKAAAEQSWS LRAEADFALS LKARADLLLR	β5 181 LSALTLPPPL\ FGKMVWPPML\ ISSMTLPPGM\	191 7RVLVA SL 7RVMLA CL 7RVMLA CL	201 AWS TTNHPY H AAS LAGAPY H AWS TQNHPY H	- R 154 R 155 R 155
-	lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514	74 74 75 77	β: 131 IP-SKPWD ER-RVES ER-RDWT EK-KDLT	141 PQLA-AELERWK PDFA-GRLGGWF TMQLA-HALPGWC IEFA-KLIQD	α4 151 KLDGR-DVSLL RDQGRGDLA V QDGR-DVA M KNELS F	β4 161 GGPE LSPAC GAD IDPAI GAD LDPEI GAY LREEI	171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS	β5 181 LSALTLP PL\ FGKMVWPML\ ISSMTLP GM\ LSKLTLA OF\	191 YRVLVA SL YRVMLA CL YRVLLA CL YKTLLL CL	201 AWS TTNHPY AAS LAGAPY AWS TQNHPY AFC NNNHPY	- R 154 R 155 R 155 K 153
-	lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168	74 74 75 77 74	β: 131 IP- KPWD ER- RVES ER- RDWT EK- KDLT ER- SVLD	141 IPQLA-AELERWK PDFA-GRLGGWF IMQLA-HALPGWQ IEFA-KLIQD IQFS-KKVENII	151 LDGR-DVSLL RDGRGDLA V 20DGR-DVA M KNELS F LGSSKKRAI Ι	161 GPE LSPAC GAD IDPAI GAD LDPE GAY LREE GAF VNEE	171 CKAAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA	β5 181 LSALTLP PLV FGKMVWP MLV ISSMTLP GMV LSKLTLA QFV LSPMVMN LMP	191 TRVLVA SL TRVMLA CL TRVLLA CL TRVLLA CL TRTLLL OT NGAMSL C T	201 AWS TTNHP AAS LAGAPY AWS TONHPY AFC NNNHPY AFT IKKIPY	R 154 R 155 R 155 K 153 N 155
	lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729	74 74 75 77 74 69	β: 131 IP- KPWD ER- RVES ER- RDWT EK- KDLT ER- SVLD ER- SVLD	141 PQLA-AELERWA PDFA-GRLGGWA IMQLA-HALPGWQ IEFA-KLIQD IQFS-KKVENII VKWA-AHLER	151 LDGR-DVSLL NDGRGDLA V DODGR-DVA M KNELS F JGSSKKRAI I -QAGNAQLD V	161 GADCIDPA GADCIDPA GADCIDPA GACLDPE GATCVNEE GATCVNEE GADCVADE	-α5 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS	β5 181 LSALTLP PLV FGKMVWP MLV ISSMTLP GMV LSKLTLA QFV LSPMVMN LMP LSKLTLP QLV	191 RVLVANSL RVMLASCL VRVLLASCL VRVLLASCL VRVLLASCL VRVLSSI VRALUSSI VRALVLSSI VRALVLSSI	201 AWS TTNHP AAS LACAPY AWS TQNHPY AFT NNHPY AFT IKKIPY AFT ICGHPY	R 154 R 155 R 155 K 153 N 155 R 148
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Ponenee 1154406200</pre>	74 74 75 77 74 69 75	β: 131 IPKFWD ERRVES ERRVES ERRUST ERSVLD	141 PQLA-AELERWK PDFA-GRLGGWF IMQLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VEFS-EXLQKK	151 (LDGR-DVSLL RDGGRGDLA V 20DGR-DVA GSSKKAI GSSKKAI 1 -QAGNAQLD V MAGIKQLV V	161 GGPE LSPA(GGAD IDPA) GGAD LDPA GGAT LREEJ GGAT VNEET GGAT VNEET GGAD VADET	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS VYKKANEKMS	β5 181 ISALTLP PLV FGKMVWP MLV ISSMTLP GMV LSKLTLA OFV LSPMVMN LMM LSKLTLP QLV LSKMTFS QMV	191 YRVIVASIY YRVMLASIY YRVILASIY YKTLLLSIY YRAIVLSIY YRAIVLSIY YRAIVLSIY YRAIVLSIY	201 AWS TTNHPY AAS LAGAPY AWS TQNHPY AFC NNNHPY AFT IKKIPY AFT IQGHPY GFT LKKEPY	R 154 R 155 R 155 K 153 N 155 R 148 H 156
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pdebret.gr.gi 10276610</pre>	74 74 75 77 74 69 75 79	β 131 IP-CKPWD ER-CRVES ER-CRUET EK-CKDLT ER-SVLD ER-CALLE DK-CKPUS VE-CKSLS VE-CKSLS	141 FPQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGW(IEFA-KLIQD- IQFS-KKVENII VWA-AHLER VEFS-EYLQKKN SEELS-RRVTDWR ULES-RVTDWR	151 151 KLDGR-DVSLL KDQGRGDLA W QDGR-DVA M KNELS F GSSKKRAI I KNELS F 	A 161 CGPE LSPA(CGAD IDPAI CGAD LDPEI CGAT CVAET CGAT CVAET CGAD VADET CGPY CFSDDT CSYS LHPAT	-a5 - 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRKRADLKVA VRKRADLKVA VKKANEKMS VKKANEKMS VKKANEKMS	B5 181 ISALTLPHPLA FGKMVWPHLA ISSMTIPHGMT LSKLTLAPOFT LSKLTLPFOLA LSKMTFFHLA SKMTFFHLA SKMTFFHLA	191 RVLVABSL RVMLACLY RVLLABCLY KTLLLECY RAIVLECIY RAIVLECIY RLFIIECLY RVMLECLY	201 AWS TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKKIP AFT IQGHP GFT LKNEP GFT LKNEP AFK AEGSS	R 154 R 155 R 155 K 153 N 155 R 148 H 156 K 160
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Swnechococcus sp.gi 116074639</pre>	74 74 75 77 74 69 75 79 75	131 IP-GKPWD ER-GRVES ER-GRDWT ER-GRDWT ER-GRDWT ER-GALLE DK-KKYS VE-KSLS EK-GKKYT EQ-KKID	141 POLA-AELERWH PDFA-GRLGGWF TMOLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VEFS-EYLQKKN SEELS-RRVTDWA VLFS-DYLNKQD VUFBSSL(151 LDGR-DVSLL RDQGRGDLAFV QDGR-DVAAM KNELSF GSSKKRAIIII -QAGNAQLDV AGQGRSQLCV VAGQGRSQLCV VIGSVQHLAFV	GADE GADE GADE GADE GADE GADE GADE GADE	-α5 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VYKKANEKMS VYKKANE	β5 181 LSALTEP PLI ISSMTLP GM LSKLTLA QF LSKLTLA QF LSKLTLP QL LSKLTES QM MSKMTFP HL/ LSMTFS QM MSKMTFP HL/ LSMTFS DHFL	191 YRVIVAS KYMLAS VI YKTLLE QAMSIS IY YRAIVIS IN IN IN IN IN IN IN IN IN IN IN IN IN	201 AWS TTNHP AWS TONHP AFC NNNHP AFT IKKIP AFT ICHP AFT ICHP AFK AEGSS AFS LKGEP	R 154 R 155 R 155 K 153 N 155 R 148 H 156 K 160 H 156 R 143
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 15903250</pre>	74 74 75 77 75 75 75 75 65 66	β: 131 IP-SKPWD ER-SKVDT EK-KDLT EK-KKLT DK-KHYS VE-KSLS EK-KKYT EQ-KTID EE-FELT	141 PDFA-GRLGGWF TPQLA-AELERWF PDFA-GRLGGWF TEFA-KLIQD IQFS-KKVENII VKWA-AHLER VEFS-EXLQKK EELS-RVTDWA VLFS-DYLNKQM VTFASRLC LGFANRLF	44 151 KLDGR-DVSLL VQDGR-DVSLV VQDGR-DVABM 	β4 161 GGPE LSPAC GAD LDPEI GAD LDPEI GAT VNEET GAT VNEET GGAT VNEET GSY FSDDT SSY LHPAT GPY FDER GAD ITAE: SAN LDSE	-a5 - 171 CKAAAEQSWS LKARADILLR FNQSLDFRLS VRKRADLKVA VRREADLRLS IVERANGSLS LKASAAWQLS KAMANWSIS	B5 β5 β5 β5 β5 β5 β5 β5 β5 β5 β	191 /RVUARSI /RVULASI /KTLLL 0 /KTLLL 0 /KTLLL 0 /KLLI 0 /KLFI /RLFI /L /RLFI 0 / /RLFI 0 / / / / / / / / / / / / /	201 AWS TTNHP AAS LAGAP AWS TQNHP AFT IKKIP AFT IKKIP GFT LKNEP AFK AEGSS AFS LKGEP AQT RAGSP AKN CEGGS	R 154 R 155 R 155 K 153 N 155 R 148 H 156 K 160 H 156 R 143 R 144
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 8803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechcococus_sp.gi 116074639 Promar.gi 15903250 Bifden.gi 171741054</pre>	74 74 75 77 74 69 75 65 65 66 78	131 IP-GKFWD ER-GRVES ER-GRUWT EK-GKDLT ER-GALLE DK-GKLS VE-GKLS EK-GKYT EQ-GKTID EE-GSLT IE-SKOLT	141 POLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGWC IQFS-KKVENII VKWA-AHLER VEFS-EXLQKKN EELS-RRVTDWA VLFS-DYLNKQN VJFASRLC GGFANRLH	151 LDGR-DVSLL NQCRGDLAFV NQCRGDLAFV NQCRGDLAFV NQCGR-DVAFM SSSKRAI I QAGNAQLDGV MAGIKQLVV MIGSVQHLMFI GGYGSQRLAFV SLGSSRLLGV	GPE GPE GPE LSPA(GAD IDPA) GAD LDPA GAT LPE GAT VNEE GAT VNEE GPY FSDD SSY LHPA GPY FDER GAD ITAE SAN LDSE GSL LDPS	-a5 -171 CKAAAAEQSWS LKARADFALS VRKRADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS VYKKANEKMS IYERANGSLS LKASAAWQLS IKAMANWSIS	β5 181 LSALTLP PL1 LSATLP GM1 LSKTLP GM1 LSKTLP QL1 LSKTFS QM1 MSKMTFS QM1 MSKMTFS QM1 LSPMTFP EL1 LSPMTFP EL1 LSPMTFP EL1 LSPTFP EL1	191 (RVLLAS SLO (RVLLAS CLO (RVLLAS CLO (RVLLAS CLO (RVLLAS CLO (RVLLAS CLO (RLFI) (RLFI) (RLFI) (RLLLL CLE (RLLLL CLE (RLLLL CLE (RVLLC CLO (RVLL) (I) (I) (I) (I) (I) (I) (I) (I	201 AWS TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKKIP AFT ICGHP GFT LKNEP AFK AEGSS AFS LKGEP AQT RAGSP AKN CEGGS AKN CEGGS	R 154 R 155 R 155 K 153 N 155 R 148 H 156 K 160 H 156 R 143 R 144 K 159
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococccus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498</pre>	74 75 77 75 75 75 75 65 66 78 75	131 IP-GKPWD ER-GRVES ER-GRDWT ER-GRDWT ER-GALLE DK-GKLYS VE-GKSLS EK-GKKYT EQ-GKTID EE-GESLT IE-KQLT IE-SKQLT	141 POLA-AELERWH PDFA-GRIGGMF MQLA-HALPGW(IEFA-KLIQD IQFS-KKVENII VKWA-AHLER VEFS-EYLQKKN SULFS-DYLNKQN VLFS-DYLNKQN VJFASRL(LGFANRLI EELA-AKIDGL(SELARRELEPL/	151 LDGR-DVSLL NQOGR-DVA NQOGR-DVA MAGURADAV MAGURADAV MAGURADAV MAGURADAV MAGURADAV SQUGSQLLAAV SLGSSRLLAAV SLGSSRLLAV SLGSSRLLAV	GAD LAPE GAD LAPE GAD LDPA GAD LDPA GAD LDPA GAD LDPA GAD VAE GAD VAE GAD VAE GAD TAE GAD TAE GAD TAE GAD LAPE GAL LAPE	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKREAAACWS VYKKANEKMS VYKKANEKMS VYKKANEKMS LKASAAWQLS LKASAAWQLS LKASAAWQLS LLQRADYLLS	β5 181 LSALTIPFPIL ISSMTIPGM LSKLTLAQF LSKLTLAQF USMTFSQMI LSKMTFSQMI LSMTFSQMI LSPMTFPELI LSPLTFPEI LSPLTFPEI FSRMTFPQI LSPLTFPEI	191 YRVIDAS LY YRVIDAS LY YRVIDAS LY YRVIDAS LY YRITHS LY YRITHS LY YRITHS LY YRLFY LY YRVIDAS LY YRVY LY	201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IQGHP GFT LKNEP GFT LKNEP AFK AEGSS AFS LKGEP AQT RAGSP ANN CEGGS AYK NAGEP	R 154 R 155 R 155 K 153 N 155 R 155 R 155 R 155 R 160 H 156 K 160 H 156 K 143 R 144 K 159 W 158
	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607</pre>	74 75 77 74 95 75 75 65 66 78 75 71	131 IP-&KPWD ER-&RVES ER-&RDWT ER-&SVLD ER-&SVLD ER-&SLLE DK-&KHYS VE-&KSLS EK-&KKYT EQ-&KTID EE-&SLT IE-&KQLT RKT&RRYG KR-&EEVS	141 POLA-AELERWH PDFA-GRLGGWF TMOLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VEFS-EYLQKKN SELS-RRVTDWR VLFS-DYLNKQN VUFSSRLC IGFANRLF SEELA-AKIDGLC EELA-RRLEPL/ SEEFA-DFLKDLF	α4 151 KLDGR-DVSLL RDQGRGDLAFV QDGR-DVXHM KNELSFF GSSKKRAIII KNELSFF GSSKKRAIII KNELSFF GSSKLLFV ANAGIKQLDFV MIGSVQHLMFI GSYGSQLLAFV KSLGSSRLLFV SLHGTSHIQLI AVSGRGHVAFV MKGK-DITIL	GAD TAE GAD TAE GAD TAE GAD TAE GAD GAD GAD TAE GAD GAD GAD GAD GAD GAD GAD G	-α5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS VKERADLKLS LKASAANQLS IKAMANWSIS ILQRADYLLS VLERADERWS IFAKAHRVFS	β5 181 ISATIPE PIL ISSMIPE GM ISSMIPE GM ISSMIPE GM ISSMIPE GM ISSMIPE GM ISSMIPE ON ISSMIPE IL ISPMIPE EL ISPLIPE IL FSRMIPE ON FGEITLP AL ISSMITE GM	191 YRVIDAS SL YRVIDAS CL YRVIDAS CL YRTILL CL YRLFVIS CL YRLFVIS CL YRLFVIS CL YRLFVIS CL YRLFUS CL YRLLL CL YRLLL CL YRLLL CL YRVIL CL YRVIS CL Y	201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKNIP AFT IKNIP AFT ICHP AFT LKNEP AFK AEGSS AFS LKGEP AKN CEGGS AKN CEGGS AYK NAGEP AVK LRGEP	R 154 R 155 R 155 R 155 R 148 H 156 R 143 R 144 K 159 W 158 Y 151
B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 140279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798</pre>	74 75 77 75 79 75 65 66 75 75 75 75 71 74	131 IP-CKPWD ER-CRVES ER-CRUT EK-CKULT ER-CALLE DK-CKHYS VE-CKSLS EK-CKYT EQ-CKTID EE-CESLT IE-SKQLT RKT-CESUS LE-CKEN	141 POLA-AELERWH PDFA-GRLGGWF TMOLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VFS-SYLQKKN SEELS-RRVTDWA VJFS-DYLNKQD VTFASRLC SELA-AKIDGLC SELARRLEPLJ SEEFA-DFLKDLF SEMS-KYIENLH	α4 151 KLDGR-DVSLL RDQGRGDLA V RDQGRGDLA V RDQGRGDLA V GUDGR-DVA M KNELS F GSSKKRAI I 	GAD	-a5 - 171 CKAAAEQSWS LRAEADFALS KRKADLLLR FNQSLDFRLS VRERADLKVA VRERADLKVA VRERADLKVS VKERADLKS I VERANGSLS LKASAAWQLS I LQRADYLLS VLERADERWS I LQRADYLLS VLERADERWS VLERA	β5 181 ISALTPP PIL ISSMTLP GM ISSMTLP GM ISSMTLP GM ISSMTFP GM ISSMTFP GM ISSMTFP FIL ISPMTFP FIL ISPMTFP FIL ISPTFP FIL ISSMTFP GM FGEITLP AL ISSMTFP GM	191 TRVINAS TRVILAS TRVILAS TRVILAS TRIFI	201 AWS TTNHP AWS TQNHP AFC NNNHP AFT IKKIP AFT IKKIP AFT IKKEP AFK AEGSS AFS LKGEP AFK AEGSS AFS LKGEP AQT RAGSP AVK NAGEP AVK LRGER AFK IHGEN	R 154 R 155 R 155 R 155 R 148 H 156 K 160 H 156 R 144 K 159 W 158 W 158 K 155
В	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 14776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 15903250 Bifden.gi 171741054 Rubxyl.gi 100804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646</pre>	74 75 77 75 79 75 65 66 78 75 71 74 64	131 IP-SKPWD ER-RVES ER-RDWT EK-KDLT ER-SVLD ER-SVLD EK-KLS VE-KSLS VE-KSLS EK-KKYT EQ-KTID EE-ESLT IE-KQLT RKT RRYG KR-ZEEVS LE-KEIN PR-KQFS	141 POLA-AELERWH PDFA-GRIGGWF IMQLA-HALPGWC IQFS-KKVENII VKWA-AHLER VEFS-EYLQKKN EELS-RVTDWA VUFS-DYLNKQN SUFS-SYLNKQN SEELA-ARIDGLC SEELA-RRLEPL/ ZEFA-DFLKDLI SENS-KVIENLL SEKS-EYLDAE/	α4 151 XLDGR-DVSLL XDQGR-DVA DQGR-DVA MQDGR-DVA MAGIKQUFV QAGNAQLD V AGSSKKRAI I QAGNAQLD V AGGGSQLC V AGGGSQLL V XSLGSQLL V XSLGSQLL V XSLGSQLL V XSLGSQLA V XSLGSXLL V XSGGHVA V MKGK-DTTIL ALGGHGELA A	β4 161 CGPE LSPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD VADE' CGAD VADE' CGAD VADE' CGAD VADE' CGAD VADE' CGAD VADE' CGAD TAE	-a5 - 171 CKAAAEQSWS LKARADILLR FNQSLDFRLS VRRADLKVA VRRAAACWS VKRANDERLS VKERADLRLS IVERANGSLS LKASAAWQLS ILQRADYLLS VLERADERWS ILQRADYLLS VLERADERWS IFAKAHRVFS IFAKAH	B5 B5 B5 B5 B5 B5 B5 B5 B5 B5	191 RVLVAS SLO RVMLAS CLY RVLLAS CLY RVLLAS CLY RVLLBS CLY RAIVLS CLY RAIVLS CLY RVLLL CLY RVLLL CLY RVVLLS CLY RVVLL CLY RVVL C	201 AWS TTNHP AAS LAGAP AWS TQNHP AFT IKKIP AFT IKKIP AFT ICHP AFT AEGSS AFS LKGEP AFK AEGSS AFS LKGEP AQT RAGSP AVK LRGER AVK LRGER AFF IHGEN AFF SNNIK AAT SAGEF	R 154 R 155 R 155 R 155 R 155 R 148 H 156 K 160 H 156 K 143 R 144 K 159 W 158 Y 151 R 155 R 155 R 145 R
8	<pre>Ins5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 14776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococccus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646 Akkmuc.gi 16834251</pre>	74 75 77 75 79 75 65 66 78 75 75 74 64 64	131 IP-GKPWD ER-GRVES ER-GRDWT ER-GXLLE DK-GXLLE DK-GXLLE DK-GXLLE DK-GXLT ER-GXLLE DK-GXLT IE-SELT IE-SELT IE-SELT RKT GRYG KR-GEEVS LE-KIN PR-KOFS ER-GELWT	141 PQLA-AELERWH PDFA-GRLGGMF IMQLA-HALPGWC IEFA-KLIQD IQFS-KKVENII VUFS-EYLQKKN EELS-RRVTDWA VUFS-DYLNKQN VUFSSRLC LGFANRLH EELA-AKIDGLC SELA-RKLEPLJ SEKS-SYLDAEJ TRKLV-DLAKDWC	151 151 LDGR-DVSLL NDQGRGDLANV NDGGRGDLANV NDGGRGDLANV ANAGIKQLV V ANAGIKQLV V ANAGIKQLV V AGQGRSQLCAV ANAGIKQLV V AGQGSQLANV SLGSSRLLV VSGGRHVANV SLGSSRLLV NIGISSINII LIGGRGELANA MHSVRRIAGL	β4 161 CGPE LSPAC GAD IDPAI CGAD IDPAI CGAD CLDPEI CGAT VREE CGAD CVADE CGAPY FSDDI CSY FSDDI CSY FDER CGAL CLAPE CGAD CLAPE CGAD CLAPE CGAL CLAPE CGSL CLAPE CGPY FDER CGAD CLAPE CGSL CLAPE CGPL CLAPE CGPD FNDA CGPD FNDA CGPD FNDA CGPD HTDA CGPD HTDA	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFALS VYKKANEKMS VYKKANEKMS VYKKANEKMS IVERANAGSLS LKASAAWQLS LKASAAWQLS LKASAAWQLS IFAKAHRVFS VKNSVDMLKL LRGQAKLLWS LRSRCNHILS	β5 181 LSALTEP PLI ISSMTLP GM LSKITLA OFF GMUMNELM LSKITLA OFF GMUMNELM LSKITLA OFF GMUMNELM LSMTFF QMU LSPMTFP GLI LSPMTFP GLI LSMTFT GMU FSHTFP QLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITFP DLI LSKITPP DLI LS	191 YRVLVASSL YRVLLASCL YRVLLASCL YRVLLASCL YRVLLSCL YRLFISCL YRLFISCL YRLFISCL YRLFISCL YRLFUSCL YRLFY Y	201 AWS TTNHPY AAS LAGAPY AFC NNNHPY AFT ICALPY GFT LKNEPY AFT ICALPY AFT ACCSS AFK ACCSS AFK ACCSS AYK ACCSS AYK NAGEP AYK ACCSS AYK ACCE AFK IHGEN AFK IHGEN AT SAGEP CHT LAGTP	R 154 RR 155 RR 155 KK 153 NR 145 RH 156 KK 160 HR 143 KK 159 W 158 KR 145 KK 155 R 145 R 145 R 145 R
A	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Poolirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15805646 Akkmuc.gi 168054251 Lenara.gi 149195686</pre>	74 75 77 79 75 75 666 78 75 71 64 64 64	131 IP- KRWD ER- RVES ER- ROWT ER- SVLD ER- GALLE DK- KHYS VE- SKLSS EK- GKKYT EQ- KTID EE- CSLT IE- CQLT RKT RRYG KR- EEVS LE- KEIN PR- KQFS ER- SLEWT EK- SLEFA	141 PQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGW(IEFA-KLIQO- IQFS-KKVENII VWFS-EYLQKKN SELS-RRVTDWR VLFS-DYLNKQM VJFSSRLG SELA-AKIDGL(SELA-KKIDGL(SELA-KKIENLE SEKAS-KVIENL SEKAS-KVIENL SEXAS-KVIENC SVEGA-KWLEKKX	α4 1,51 KLDGR-DVSLL NDQGR-DVA DQGR-DVA GGSCALA GSSKKRAI GSSKKRAI GSSKKRAI GSSKLD VA NAGIKQLV VA GQGRSQLC VA SGSSKLA VA SGSSKLA VA SGSSKLA VA SGSSKLA VA SGSSKLA VA SGSSKLA VA SGSSKLA VA SGSSKA SGSSS SGSS	β4 161 CGPDE LSPAC GADG IDPAI CGADC IDPEI GAT LDPEI GAT LREE CGAD VADE CGPY FSDD CSY LHPA CGPY FDER CGAD IDSE CGPY FDER CGPU FDER CGPU LAPE CGPU LAPE CGPU LAPE CGPU LAPE CGPU LAPE CGPU LAPE CGPU CAPE CSNG VSKN CAPE HTEE CAPE TEE	-α5 -171 CKAAAGSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS IYERANGSLS LKASAANQLS IKAMANWSIS IIAQADYLLS IFAKAHRVFS VKNSVDMKLK LRSQCMHLLS EYSRANHTLS	β5 181 ISALTPE PLI ISSMTIPE GMI ISSMTIPE GMI ISSMTIPE GMI ISSMTIPE OLI ISSMTIPE OLI ISSMTIPE TAL ISSMTIPE TAL ISSMTIPE TAL ISSMTIPE OLI ISSMTIPE OLI I	191 YRVIDAS SLS YRVIDAS CLS YRVIDAS CLS YRTLIS CLS YRTLIS CLS YRTLIS CLS INTERNO INTE	201 Aws TTNHP AAS LAGAP AWS TONNP AFT ICNNP AFT ICN AFT ICN AFT ICN AFT ICN AFT ICN AFT LKNEP AFT LKNEP AFT ACG AFS LKGEP AFS LKGEP AKN CEGGS AYK INACEP AYK INGER AFK IHGEN AFK IHGEN AAT SAGEP AAT SAGEP CHT LAGTP	R 154 RR 155 K 153 K 155 K 153 R 148 K 160 HR 143 R 144 K 159 W 158 R 145 K 155 R 145 K 155 R 145 K 145 K 145 K 155 K 155 K 155 K 155 K 156 K 166 K 16
B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646 Akkmuc.gi 166834251 Lenara.gi 149155686 Acibac.gi 94967055 </pre>	74 75 77 75 65 65 65 65 65 65 71 74 64 64 64 63	131 IP-KFWD ER-KRUT ER-KOUT ER-KULT ER-KULT ER-KILE DK-KHYS VE-KSLS EK-KKYT RT RT RRYG KR-CEVS LE-KEIN PR-KOFS ER-SELT EK-CEFA ER-KOVG	141 POLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VFS-DYLQKKN SELS-RRVTDWR VLFS-DYLNKQN SUFFASRLC SUFFANRLF SELA-AKIDGLC SEFA-DFLKDLF SENS-KYIENLF SENS-KYIENLF SENS-KYIENLF SENS-KYIENLF SELS-EXLDAPF SELA-NRLEWKK SELA-NFLGYHF	α4 151 KLDGR-DVSLL RDQGRGDLAV QDGR-DVXAM KNELSIF GSSKKRAIII QOAGNAQLDAV MAGIKQLVAV MAGIKQLDAV MIGSVQLLAV KSLGSSRLLV VSGRGHVAV SLGSSRLLV SLGSSRLLV NGGGELAA MKGK-DITIL KNIGISSINIII ALGGHGELAA DMHSVRRIAL CMHSVRIAL	GAD	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS VKRANGLS LKASAAWQLS IKASAAWQLS IKASAAWQLS IFAKAHRVYS VVLERADERWS IFAKAHRVYS VKNSVDMKLK LRQQAKLLWS LRSRCNHILS EVSRANHTLS	Antipe β5 181 LSALTLP PLI SSMTLP GM LSKLTLA QFV LSKLTLA QFV LSKLTLA QFV LSKLTLP QLV LSKLTLP QLV LSKTTS QM MSKMTFP HLI LSMTFP ELI LSPTFP ELI LSPTFP ELI LSFTLP ALI LSKTTFT GM FSHFTFP QLI LSKTTTP QLI LSKTTTP M	191 YRVIDAS SL: YRVIDAS CLY YRVIDAS CLY YRVIDAS CLY YRLFULS CLY YRLFYLS CLY YRLFYLS CLY YRLFYS CLY YRLFYS CLY YRLFYS CLY YRLFYS CLY YRLFYLS CLY YRVIDLS CLY YRVIDLS CLY YRVIDLS CLY YRVFFLS CLY	201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKKIP AFT IKKIP AFT LKNEP AFK AEGSS AFS LKGEP AKN CEGGS AKN CEGGS AK	R 154 RR 155 K 153 N 155 RH 156 K 160 HH 156 RR 143 K 159 W 158 K 159 W 158 K 155 R 145 K 155 R 145 R 145 R 145 R 145 R 145 R 145 R 155 R
B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 8880363 Baccap.gi 154496020 Pedobacter_sp.gi 1140279619 Synechococcus_sp.gi 116074639 Promar.gi 15903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 1960804498 Themar.gi 15643607 Fusnuc.gi 1960878 Deirad.gi 15805646 Akkmuc.gi 16684251 Lenara.gi 14915686 Akibac.gi 94967055 Solusi.gi 116624738 Europi 1116624738</pre>	74 74 75 77 75 65 66 78 75 75 65 66 78 75 71 74 64 64 64 63 59	131 IP- KPWD ER- RDWT ER- RDWT EK- KULT ER- SVLD ER- SVLD ER- SVLD ER- SVLD ER- SVLD ER- SVLD ER- SVLD E- CSLT IE- SKQLT RKT REYS KR- GEVS ER- SLWT RK- SEFS ER- SLWT EK- LEFA ER- KQVG PA- SKPMD	141 POLA-AELERWH PDFA-GRLGGWF TMOLA-HALPGWC IEFA-KLIQD IGFS-KKVENII VKWA-AHLER VFS-DYLNKQD VFS-DYLNKQD VTFASRLC SELA-AKIDGLC SELARRLEPLJ SENAS-KYIENLI SEKA-DFLKDLF SEKAS-EYLDAEJ TRKLV-DLAKDWC VEGA-KWLEKKZ SELA-NFLGYHF SAFA-AMISKCI SENAS-2777	4 4 151 CLDGR-DVSLL CDQGRGDLAFV QQGR-DVAFM KNELSF CGSSKKRAFI CGSSKKRAFI CGSSKKRAFI CGSSKKRAFI CGSSKRAFI CG	β4 161 CGPE LSPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD VADE' CGAD VADE' CGAD VADE' CGAD VADE' CGAD VADE' CGAD TAE CSA	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS VKKANEKMS VKERADLLS LKASAAWQLS LKASAAWQLS LKASAAWQLS ULERADENWS LIQRADYLLS VLERADENWS LRSQAKLLWS LRSQAKLLWS LRSQAKLLWS LRSQAKLLWS LRSQAKLLWS LRSACNHILS TLKSATQULS WKARADLLVS	β5 181 LSALTLP PLI SSMTLP GM LSKLTLA QF LSKLTLA QF LSKLTLA QF LSKLTLP QL LSKLTP QL MSKMTFP EL LSPTFP EL LSPTFP QL SFMTFP QL LSKMTFT GM SFSHTFP QL LSKMTFT GM LSKMTFT EM MGKMTFP EL LSKMTFT EM	191 YRVIVAS IRVMLAS IVVILAS VRVILAS IV IVILAS IV INVILS IV IVILIS IV IVILS IVILS IVILS IVVILS	201 AWS TTNHP AAS TQNHP AFC NNNHP AFC NNNHP AFT IKKIP AFT IKKIP AFK AEGSS AFS LKGEP AFK AEGSS AFS LKGEP AQT RAGSP AKN CEGGS AYK NAGEP AKN CEGGS AYK NAGEP AKN LGEGS AYK INGEP AFK IHGEN MFA SNNIK AAT SAGEP CHT LAGTP GYT LTGHP	R 154 R 155 K 153 K 155 K 153 R 148 K 160 K 160 K 160 K 159 K 151 K 155 K 153 K 155 K 153 K 155 K 160 K 155 K 153 K 155 K 153 K 155 K 155
8	<pre>Ins5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 16643607 Fusnuc.gi 16703798 Deirad.gi 15805646 Akkmuc.gi 166834251 Lenara.gi 149195686 Acibac.gi 94967055 Solusi.gi 116624738 Synwol.gi 114567973 Bacsub.gi 16624738 Synwol.gi 1165795</pre>	74 74 75 77 74 69 75 79 65 66 78 75 71 74 64 67 8 71 74 64 67 78 78	131 IP-KFWD ER- RAVES ER- RAVES ER- RAVES ER- KAUE ER- KALLE DK- KHYS VE- KSLT EQ- KTID EE- ESLT IE- KKYT RKT RRYG KR- EEVS IE- KEIN PR- KKPS ER- ELWT EK- CLEFA ER- KQVG PA- KPMD LE- QVRS	141 PQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGWC TEFA-KLIQD IQFS-KKVENII VKWA-AHLER VEFS-EYLQKKN BELS-RVTDK VLFS-DYLNKQN VLFS-DYLNKQN VLFS-DYLNKQN SELA-AKIDGLC SELA-AKIDGLC SELA-KIDGLC SELS-EYLDAEZ TRKLV-DLAKDWC VEGA-KWLEKKN SELA-NFLGYH SAFA-AMISKGI SEEMA-RQLEKWI	α4 151 XLDGR-DVSLL XDQGRG-DVA QDGR-DVA QDGR-DVA MAGKALA GSSKKRAI I QAGNAQLDGV MAGKQLC V AGQGRSQLC AG AG AG AG AG AG AG AG AG AG	β4 161 CGPE LSPAC GGAD IDPAI GGAD IDPAI GGAT UDPAI GGAT VNEET GGAT VNEET GGAT VNEET CGPY FSDD CSY FDER GGAD ITAE: CSAT LAPE CGPY FDER CGAD ITAE: CGSL LAPE CGPL LAPE CGPD ILAPE CGPD TAA CASD HTEE SAN FTEE PSD WCKE CHD CASD TAP CASD TAP	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFALS VYKKAALKMS VYKKAALKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS LKASAAWQLS LKASAAWQLS LKASAAWQLS LKASAAWQLS LKASAAWQLS LKASAKLWS LRSPALLWS LRSPALLWS LRSPALLWS LKSACQULS WRAADLLVS LKQAQETIS LKQAQETIS	β5 181 LSALTPE PLI LSALTPE GM LSKITPE GM LSKITPE QM LSKITPE QM LSKITPE QM LSKITPE QM LSKITPE QM LSPTFFE QM LSPTFFE QM FGEITPE AL LSKITFFE QM FGEITPE DL LSKITFFE QM LSKITFFE QM	191 YRVLVAS SL YRVLLAS CL YRVLLAS CL YRVLLAS CL YRVLLAS CL YRVLLS CL YRVLLS CL YRLFY YRLFY YRLFY YRLFY YRLFY YLLLS CL YRLFL CL YRLFL CL YRLFL CL YRLFL CL YRLFL CL YRLFY YLLLS CL YRVLLS CL YRVLLS CL YRVLLS CL YRVFL CL YRVFC CL YRVF	201 AWS TTTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKKIP AFT IKKIP AFT ICGHP GFT LKNEP AFK AEGSS AFS LKGEP AQT RAGSP AVK RGEP AVK RGEP AFK IHGEN GFC LRGHP GFC LRGHP	RR 154 RR 155 KK 153 KK 153 KK 155 KK
B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechcococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 10804498 Themar.gi 15805646 Aktmuc.gi 15805646 Aktmuc.gi 16834251 Lenara.gi 149195686 Acibac.gi 94967055 Solusi.gi 16081075 Solusi.gi 116081075 Clacze gi 158072</pre>	74 74 77 77 74 69 75 66 65 66 57 71 64 64 64 64 67 59 78 78 78	131 IP- KRWD ER- RNWT ER- RNWT ER- SVLD ER- SLLE DK- KHYS VE- SKLSS EK- KKYT EQ- KTID EE- ESLT IE- KQLT RKT RRYG KR- EEVS LE- KLEFA ER- SLWT EK- SLEFA ER- SLWT ER- SLWT	141 PQLA-AELERWH PDFA-GRIGGWF TMQLA-HALPGW() IEFA-KLIQD IQFS-KKVENII VUFS-EYLQKKN SELS-RRVTDWA VUFS-DYLNKQN VJFASRL(SELARRLEPL/ SELA-KNIDGL(SELARRLEPL/ SEKARRLEPL/ SELA-NFIGLA SELA-NFIGLA SELA-NFIGLA SELA-NFIGHH SAFA-AMISKG SEMA-RQLEKW SELA-DTIDKL/ SELA-DTIDKL/ SELA-TIDK/	α4 151 KLDGR-DVSLL KDQGRGDLAFV QQGR-DVAFM GQGRSQCLAFV MAGIKQLVFV MAGIKQLVFV MAGIKQLVFV MAGIKQLVFV MAGIKQLAFV MAGIKQLAFV MAGIKQLAFV MAGIKGSLLFV MAGGHGELAFA MHSVRTIAFL AHCSGKLIFV NAGGKSVTFV VASGKSVTFV MAGKSVTFV	β4 161 CGPE LSPAC GAD IDPE GAD IDPE GAT LPE GAT VNET GAD CAPA CGPY FSDD CSY LPPA CGPY FDER CGAD TAE CGPU LAPE CGPU FDER	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKREAAACWS VYKKANEKMS VYKKANEKMS VYKKANEKMS IVERADACWS ILARAANQLS ILARAANVSIS ILARAANVSIS IFAKAHRVFS VKNSVDMKLKS LKSRCHHLS EYSRANHTLS TLKSAAQLLVS IKQQAQETIS VMRAADLLVS IKQQAQETIS	β5 181 LSALTPE PLI ISSMTIPE GMI ISSMTIPE GMI LSKLTLA QEP LSKLTLA QEP LSKLTLA QEP LSKLTEP QLI LSKTFF GMI FSKMTFF GLI LSKTLO ELI LSKTTC ELI LSKTTC ELI LSKTTFF ELI LSKTTFF ELI LSKTTFF ELI LSKTTFF ELI LSKTTFF ELI SKMTFFF ELI LSKTTFF ELI SKMTFFF ELI SKMTFFFFF ELI SKMTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	191 /RVUVAS SLV /RVULAS CLV /RVULAS CLV /RVULAS CLV /RULAS CLV /RULAS CLV /RULAS CLV /RULAS CLV /RULAS CLV /RULAS CLV /RULLE CLV /RVULS CLV /RULAS CLV /RVFFL CLV /RVFL CLV /RVFFL CLV /RVFL CLV /RVFC CLV /RVFC CLV /RVFC CLV /RVFC CLV /RVFC CLV /RVFC C	201 Aws TTNHP AAS LAGAP AWS TONHP AFC NNNHP AFT IQGHP GFT LKNEP GFT LKNEP GFT LKNEP AFK AEGSS AFS LKGEP AAC RAGSP AVK LGER AVK LGER AYK IHGEN FA SNNIK AAT SAGEP CHT LAGTP MIS QKGTP GFT LKGEP GFT LKGEP GFT LKGEP	RR 154 RR 155 KK 153 KK 155 KK 155 KK 160 HR 143 KK 159 KK 155 KK
8	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15805646 Akkmuc.gi 1980545 Deirad.gi 15805646 Akkmuc.gi 1682451 Lenara.gi 16624738 Synwol.gi 114657973 Bacsub.gi 16081075 Cloace.gi 15896772 Mycnul.gi 1524725</pre>	74 74 75 77 74 69 75 66 78 75 66 78 75 66 78 71 74 64 64 67 63 59 78 78 78	131 IP- KRWD ER- RVES ER- ROWT ER- SVLD ER- SVLD ER- SLLE DK- KHYS VE- KSLS EK- KKYT EQ- KTID EQ- KTID EQ- KTID EQ- KTID EQ- KSLS ER- SELWT ER- SLWT EK- SLEFA ER- KOVG PA- SKPMO IE- CVTS IE- KMKT IK- KMFS	141 PQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGWC IEFA-KLIQD IQFS-KKVENI VUFS-EYLQKKN SELS-RRVTDWR VUFS-DYLNKQA VUFS-DYLNKQA VUFS-CHARRALEPL/ SELA-AKIDGUC SEFA-FILDAKDW VEGA-KWLEKK/ SEMS-KYIENLL SAFA-AMISKGI SEEMA-RQLEKWN SEEMA-RQLEKWN SEEMA-RQLEKWN SEEMA-RDICKL/ SEEMA-RDICK	α4 151 KLDGR-DVSLL KLDGR-DVSLL KDQGRGDLAFV QQGR-DVA GOGRADUA GGGSQLLAFV KSLGSSLLFV KSLGSSLLFV KSLGSSLLFV KSLGSSLLFV KSLGSSLLFV KSLGSSLLFV AVSGRGHVAFV MKGK-DITIL ALGGHGELAFA ALGGHGELAFA ALGGHGELAFA ALGGGELAFV ALGGHGELAFA ALGGGELAFV ALGGHGELAFA ALGGGELAFV ALGGHGGHG ALGGHGELAFV ALGGHGELAFV ALGGHGELAFV ALGGHGELA	GAD TAAE GAD TAAE GAD TAAE GAD GAD GAD GAD GAD GAD GAD GAD GAD GAD GAD GAD GAD GAD GA	-a5 -171 CKAAAACSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS IYERAAACWS IYERAADLLS VKNSVDMKLK LRGQAKLLWS VKNSVDMKLK LRGAKLLWS VKNSVDMKLK LRGACHILS EYSRANHTLS TIKSAAQLLS WKRAADLLVS VKRAADLLS VKRAADLLS VKRAADLLS VKRAADLLS VKRAADS VKRAADS VLARANYLC	β5 181 LSALTLP PLI SGKMVWPSMLX ISSMTLP GM LSKLTLAQFV LSKNTFS QM MSKMTFP GL LSKMTFS QM MSKMTFP ELI LSPMTFP ELI LSPMTFP GL LSKMTFT GM FSHTFP QL LSKTTP LL LSKTTP ELI LSKMTFP ELI LSKMTFP CM FSKMTFP QL FSKMTFP QL	191 /RVUVAS SL /RVULAS CL /RVULAS CL /RVULAS CL /RTULAS CL /RTULAS CL /RUVLAS CL /	201 Aws TTNHP AAS LAGAP H Aws TQNHP AFC NNNHP AFT IKIF AFT IKKFP AFT LKKEP AFT LKKEP AFK AEGSS AFS LKGEP AVK LRGEP AVK LRGEP AVK LRGEP AVK LRGEP AVK LRGEP GFT LKGEP GFT LKGEP GFT LKGEP GFK NRGEP AFR NRGEP AFR NRGEP	R 154 RR 155 K 153 K 155 R 148 R 148 R 143 R 144 F 143 R 145 R 145 R 145 R 145 R 145 R 145 R 145 R 145 R 145 R 155 R 155
B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 8800363 Baccap.gi 154496020 Pedobacter_gp.gi 1140279619 Synechococcus_gp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646 Akkmuc.gi 16634251 Lenara.gi 16915686 Acibac.gi 94967055 Solusi.gi 116624738 Synwol.gi 114567973 Bacsub.gi 16081075 Cloace.gi 1580872 Mycpul.gi 15828725 Metmar.gi 15828725</pre>	74 75 77 79 75 65 66 66 78 75 75 74 64 64 63 59 78 78 78 78 78 78 78	131 IP- KPWD ER- RDWT ER- RDWT ER- RDWT ER- SVLD DK- KHYS VE- KSLS EK- KHYS VE- KSLS EK- GKYT EQ- KHYS ER- SLST IE- KKLT RXT RRYG KR- GEVS IE- KEIN PR- KQFS ER- SLWT EK- LEFA ER- KQVG PA- KRMTS IE- KMKT LK- KMFS IL- KEIT	141 POLA-AELERWH PDFA-GRIGGWF IMQLA-HALPGWC IGFS-KKVENII VKWA-AHLER VEFS-EYLQKKN EELS-RVTDWA VLFS-DYLNKQN SUFS-DYLNKQN SUFS-STUCK GGFANRLF SELA-ARIGGL SELA-RAIDGLC SELA-RAIDGLC SELS-YLDAEJ WEGA-KWIENLF SAFA-AMISKG SELA-NFLGYHH SAFA-AMISKG SELA-NFLGYH SAFA-AMISKG SELA-DTIDKLZ SEFS-AFIDDLC SEFS-AFIDDLC SEFS-AFIDDLC SEFS-AFIDDLC SEFS-AFIDDLC SESLA-KWND10	α4 151 KLDGR-DVSLL KLDGR-DVALM QQGR-DVAM QQGR-DVAM GSSKKRAIFI QAGNAQLDOV MNAGKQLCVV GSSKKRAIFI QGGQGRSQLCVV KSLGSSRLLVV KSLGSSRLLVV KSLGSSRLLVV KSLGSSRLDV MKGK-DITIL ALGGGGELAFA AMHSVRRIADL ALGGGGELAFA AMHSVRRIADL ALGGGGELAFA MHSVRRIADL AHCSGKLIVV MXGKSVTDV SVRGNSSIDV DISISV	β4 161 CGPE LSPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD VADE' CGAD TAE CGAD TAE <td< th=""><th>-a5 - 171 CKAAAEQSWS LRAEADFALS LRAEADFALS VREAADFALS VREAADFALS VREAADFALS VREAADFALS VREAAACWS VREAAACWS VKEAADLLLR KAMANWSIS LLASAAWQLS LKASAAWQLS LRADATHS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADELLS VLEAANYLLC MFKGA-WKIN</th><th>Antipe β5 181 LSALTLP PLI SKATLP GM LSKLTLA QFV LSKLTLA QFV LSKLTLA QFV LSKLTLP QLV LSKLTP QLV LSKTTP QLV SKMTFP ELI LSPTTPP ELI LSKTTP QLV SKFTLQ ELI LSKTTP QLV LSKTTP QLV SKTTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV</th><th>191 RVULAS SLO RVMLAS CLY RVLLAS CLY RVLLAS CLY RVLLAS CLY RVLLBS CLY RAIVLS CLY RAIVLS CLY RVLLD CLY RVLLS CLY RVVLLS CLY RVVLS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVS R</th><th>201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFC NNNHP AFT IKKIP AFT IKKIP AFK AEGSS AFS LKGEP AFK AEGSS AFS LKGEP AQT RAGSP AKN CEGGS AKN CEGGS AKN CEGGS AKN CEGGS CHT LKGEP AVK LRGEP AFK IHGEN WFA SNNIK AAT SAGEP CHT LAGTP GYT LTGHP GYT LTGHP GFK LKGEP AFR NRGEP AFR NRGEP</th><th>R 154 R 155 K 153 K 155 K 155 R 155 K 160 K 160 K 160 K 159 K 152 K 152 K 152 K 153 K 153 K 155 K 155</th></td<>	-a5 - 171 CKAAAEQSWS LRAEADFALS LRAEADFALS VREAADFALS VREAADFALS VREAADFALS VREAADFALS VREAAACWS VREAAACWS VKEAADLLLR KAMANWSIS LLASAAWQLS LKASAAWQLS LRADATHS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADENUS VLEAADELLS VLEAANYLLC MFKGA-WKIN	Antipe β5 181 LSALTLP PLI SKATLP GM LSKLTLA QFV LSKLTLA QFV LSKLTLA QFV LSKLTLP QLV LSKLTP QLV LSKTTP QLV SKMTFP ELI LSPTTPP ELI LSKTTP QLV SKFTLQ ELI LSKTTP QLV LSKTTP QLV SKTTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV FSKMTPP QLV	191 RVULAS SLO RVMLAS CLY RVLLAS CLY RVLLAS CLY RVLLAS CLY RVLLBS CLY RAIVLS CLY RAIVLS CLY RVLLD CLY RVLLS CLY RVVLLS CLY RVVLS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVVS RVS R	201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFC NNNHP AFT IKKIP AFT IKKIP AFK AEGSS AFS LKGEP AFK AEGSS AFS LKGEP AQT RAGSP AKN CEGGS AKN CEGGS AKN CEGGS AKN CEGGS CHT LKGEP AVK LRGEP AFK IHGEN WFA SNNIK AAT SAGEP CHT LAGTP GYT LTGHP GYT LTGHP GFK LKGEP AFR NRGEP AFR NRGEP	R 154 R 155 K 153 K 155 K 155 R 155 K 160 K 160 K 160 K 159 K 152 K 152 K 152 K 153 K 153 K 155 K 155
A B I	<pre>Ins5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 140279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646 Akkmuc.gi 166834251 Lenara.gi 149195686 Acibac.gi 94967055 Solusi.gi 116624738 Synvol.gi 114567973 Bacsub.gi 10880772 Mycpul.gi 15828725 Metmar.gi 159906004 Methun.gi 8601833</pre>	74 75 77 75 79 75 65 66 78 75 71 74 64 64 64 65 78 78 78 78 78 78 78 78	131 IP-KFWD ER-GRVES ER-GRUMT ER-GRUMT ER-GLLE DK-KYT ER-GLLE DK-KHYS VE-GKLS EK-GKYT EQ-KTID EE-GELT IE-KKYT RKT RRYG KR-GEVS IE-KEIN PR-KQFS ER-ELWT EK-CKFS ER-ELWT EK-CKFS IE-KKUS IL-SKLT KMFS IN-AEQYS IL-KELT KT-KGVS	141 PQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGWC TEFA-KLIQD IQFS-KKVENII VUFS-DYLNKQN VUFS-DYLNKQN VUFSSRLC LGFANRLK EELA-AKIDGLC SELA-RKIDGLC SELA-RKIDELS SELS-SYLDAES TRKLV-DLAKDWC VUESA-WHEKKI SELA-NFLGYH SAFA-NISKGI SEELA-NTIDLC SEFA-DFLDDLC SEFA-LFLQE- SELA-KKINDLS SEGLA-KVIRCC	α4 151 LDGR-DVSLL ADQGRGDLANV ADQGRGDVAM COGROLANV AGQGRSQLCIV AGQGRSQLCIV AGQGRSQLCIV AGQGRSQLCIV AGQGRSQLCIV AGQGRSQLCIV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGRSQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGQGVQLLANV AGGRSVT ATYGKSKVT AGGRSDT AGGSSDT AGGSSDT AGGSSDT AGGS	β4 161 CGPE LSPAC CGAD IDPAI CGAD IDPAI CGAD CLDPEI CGAT VNEET CGAT CLAPE CGSL CLAPE CGSL CLAPE CGPT LAPE CGPD TTDA CASD HTEE VSAN FTEE PSD WCKE CGL LSAPT CGHD TAA CASD HTEE CGAD TAB CGAD TAB CGAD TAB CGSL LSTA CGAD TAB CGAD TAB CGSL LSTA CGSL LSTA CGSL LSTA CGSL LSTA CGSL LSTA CGSL LSTA	-a5 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFALS VYKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS IVERANGSLS LKASAAWQLS IKAMANWSIS IVERANGSLS IFAKAHRVFS VKNSVDMKLK LRSQAKLLWS LRSQCHILS EYSRANHTLS EYSRANHTLS WKRADEKLS VMKRADEKLS VMKRADEKLS VMKRADEKLS VMKRADEKLS MK	β5 181 LSALTIPE PLI LSALTIPE GMI LSKITLPE GMI LSKITLPE GMI LSKITLPE QLI LSKITLPE QLI LSKITFE QMI MSKMTFFE QMI LSPMTFFE QMI LSPMTFFE QMI LSPMTFFE QLI LSKITFTE GMI LSKITFTE QLI LSKITFFE QLI FSKMTFFE QLI FSKMTFFE QLI FSKMTFFE QLI LSAMTFFE QLI L	191 YRVLVAS SL YRVLLAS CL YRVLLAS CL YRVLLAS CL YRVLLAS CL YRVLLS CL YRVLLS CL YRVLS CL YRVLS CL YRVLS CL YRVVLS CL YRVS CL	201 Aws TTNHP AAS LAGAP AWS TQNHP AFC NNNHP AFT IKKIP GFT LKNEP GFT LKNEP AFK AEGSS AFS LKGEP AQT RAGSP AGT RAGSP AVK LRGEP AVK LRGEP AVK LRGEP CHT LAGTP MLS QKGTP GFT LTGHP GFT LTGHP GFT LRGEP AFR SRGEP AFR SRGEP AFR SRGEP AFR SRGEP AFR INGEP	R 154 R 155 R 155 R R 155 R R 155 R 145 R 155 R 145 R 155 R 145 R 155 R 145 R 155 R 145 R 155 R
A B	<pre>Ins5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Edebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 14074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 10804498 Themar.gi 15805646 Akfmuc.gi 16834251 Lenara.gi 149195686 Acibac.gi 9407055 Solusi.gi 116624738 Synwol.gi 114567973 Bacsub.gi 16081075 Cloace.gi 159906004 Methun.gi 88603833 Orysat.gi 115406408</pre>	74 75 77 74 65 66 78 75 71 74 64 64 64 67 63 78 78 78 78 78 78 78 78 78 78 78 78 78	131 IP- KRWD ER- RNWT ER- RNWT ER- SVLD ER- SLLE DK- KHYS VE- KSLS EK- KKYT EQ- KTID EE- ESLT IE- KQLT RKT RRYG KR- CEVS LE- KLIN PR- KQVS IE- KLIN PR- KQVS IE- KKMTS IL- KKMTS IL- KKUT RT - KQVS IL- KKUT	141 PQLA-AELERWH PDFA-GRIGGMF MQLA-HALPGW() IEFA-KLIQD IQFS-KKVENII VEFS-EYLQKKN SELS-RRVTDWA VIFSSRL(LGFANRLH EELA-AKIDGL(SELA-RKIEPL/ SELA-RKIEPL/ SELA-NFIGU(SELA-NFIGU(SELA-NFIGU(SELA-NFIGUH SAFA-AMISKG SEEMA-RQLEKWN SELA-DTIDKL] SEFS-AFIDDL(SEFS-AFIDDL(SEFA-FIDDL(SEFA-FIDDL(SEGLA-EVIRC(SEQVA-DLVGDA(SEQVA-DLVGDA(SEQVA-DLVGDA(SEQVA-DLVGDA(SECA-SULC) SEQVA-DLVGDA(SECA-SULC) SEQVA-DLVGDA(SECA-SULC) S	a4 151 LDGR-DVSLL KLDGR-DVSLL KDQGRGDLAFV QQGR-DVAFM GGSKKRAIFI GGSSKRAIFI GGSSKLAFV MAGIKQLVFV MAG	β4 161 CGPE LSPAC CGAD IDPAI CGAD IDPAI CGAD IDPAI CGAD IDPAI CGAD IDPAI CGAD IDPAI CGAD VADE CGAD YADE CGAD YADE CGAD TAE	-a5 -171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VKKANEKMS VYKKANEKMS VYKKANEKMS UKREAAACWS VYKKANEKMS IYAKANGSLS LKASAAWQLS IIAKAANWSIS IIAKAANWSIS IIAKAANWSIS IIAKAANVS VKNSVDMKLKS LNSQAKLLWS LNSQAKLLWS LNSQAKLLWS LNSQAKLLWS LNSQAKLLWS LNSQAKLMS LNSQAKLMS LNSQ	β5 181 LSALTIPF PLI ISSMTIPF GMI ISSMTIPF GMI LSKITIA QFF LSKITIFF QMI LSKITIFF QMI LSKITFF QMI LSPMTFF ELI LSPMTFF ELI LSPMTFF ELI LSFTTOF GMI FSKITFF QMI FSKITFF QLI FSKITFF QLI FSKITFF QLI SSMTFF QLI LSSMTFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFF QLI LSSMTFFFF QLI LSSMTFFFF QLI LSSMTFFFF QLI LSSMTFFFF QLI LSSMTFFFF QLI LSSMTFFFFFF QLI LSSMTFFFFF QLI LSSMTFFFFF QLI LSSMTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	191 //RVLVAS_SLV //RVLLAS_CLV //RVLLAS_CLV //RVLLAS_CLV //RVLLS_CLV //RLFIVS_CLV //RLFIVS_CLV //RVLLS_CLV //RVLLS_CV //RVLS_CV	201 Aws TTNHP I AAS LAGAP I Aws TONNP I AFC NNNHP AFT IQGHP GFT LKNEP I AFT IQGHP GFT LKNEP I AFK AEGSS AVS LKGEP AQT RAGSP AVS LKGEP AVK LGER AVK LGER AVK IHGEN AT SAGEP GFT LKGEP GFT LKGEP GFT LKGEP AFR SRGEP AFR SRGEP AFR INGEP AFR SRGEP AFR INGEP AFR INGEP AFR INGEP	RR 154 RR 155 K 153 K 155 K 155 K 155 K 156 K 166 K 156 K 155 K 156 K 155 K 15
A B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Edebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 8880363 Baccap.gi 154496020 Pedobacter_sp.gi 114074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15805646 Akfmuc.gi 19805483 Themar.gi 1560566 Acibac.gi 94967055 Solusi.gi 116624738 Synwol.gi 116624738 Synwol.gi 116681075 Cloace.gi 15990722 Mycpul.gi 15896772 Mycpul.gi 15896772 Mycpul.gi 15896772 Mycpul.gi 15896772 Mycpul.gi 15906004 Methun.gi 88003833 Orysat.gi 115460408 Aratha.gi 15540408 Aratha.gi 152424</pre>	74 75 77 74 65 76 76 76 76 76 76 76 76 76 76 76 76 76	131 IP- KFWD ER- RVWJ ER- RDWT ER- SVLD ER- SVLD ER- SLLE EK- KKYT EQ- KIND EE- CSLT IE- CQLT RKT RRYG KR- CEVS LE- KEIN PR- KQFS ER- SLWT EK- LEFA ER- SLWT EK- KMFS IL- KQVS IL- KELT RKT RRYG RK- CEVS IL- KKFS IL- KQVS ER- SLWT ER- SRUD	141 PQLA-AELERWH PDFA-GRLGGWF TMQLA-HALPGW(IEFA-KLIQD IQFS-KKVENII VWFS-EYLQKKN SELS-RRVTDWR VLFS-DYLNKQN VJFS-DYLNKQN SELA-KKIDGL(SELA-KKIDGL(SELA-KKIENL SEKA-DFLKDLH SEKA-NLGYLN SELA-NLGYLN SELA-NLGYLN SELA-RULEKKN SELA-NLGYLN SELA-TLDLAC SELA-RULEKN SELA-TLDLC SELA-KINDL(SELA-TLDLC SELA-KINDL(SELA-KINDL(SELA-LFLQE SELA-KINDL(SELA-LFLQE SELA-KINDL(SECA-EVIRC(SEQA-ELLGDA(SEQMA-ELLGDA(SEMA-SEMA-SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA-SEMA(SEMA) SEMA SEMA-SEMA(SEMA) SEMA	α4 1,51 KLDGR-DVSLL KLDGR-DVSLL NDQGR-DVA GOGR-DVA GGGR-DVA GGSVGLA MAGIKQLV VA GGGSVGLA MAGIKQLV VA GGGSVGLA MAGIKQLV VA SGGSSRLL VA SGGSSRLL VA SGGSCA MAGGHGELA ALGGHGELA ALGGHGELA MHSVRTA LA GGGRGLA MHSVRTA LA MHSVRTA SI SSGSSTT I SSGSSTT ST SSGSSTT SN SSGSSTT SSGS S	β4 161 CGPE LSPAC GADG IDPAI GADG ITAE: GADG ITAE: GSNG VSKNI GPD HTDA ASD HTEE: CSAY FTEE: CGSL LSDT GSL LSDT GSH HOPE	-a5 -171 CKAAAAGSWS LRAEADFALS LRAEADFALS LKARADLLLR FNQSLDFRLS VKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS IYAKANEKMS IYAKANESI IYAKANESI IYAKANESI IYAKANESI IYAKANESI IYAKANESI IYAKANESI VKNSVDMKLK LRSQCHILS VKRAADELLS VLERADELLS VLERADELLS VLERADELLS VMKRANELS VLERADELLS VMKRANESI VLERADELLS VLERADELLS VLERADELLS VLERADELLS VLERADELLS VLERADELLS VLERADELLS	β5 181 LSALTIP PIL ISMTIP GM LSKLTLA QFV LSKNTFB GM LSKLTLA QFV LSKMTFS QM MSKMTFP GL LSKMTFP GL LSPMTFP GL LSPMTFP GL LSKMTFT GM FSKMTFP QL LSKTLO ELL LSKTTC ELL LSKTTP QL LSKTTP QL FSKMTFP QL FSKMTFP QL ISSMVIN QVI LSSMVIN QVI LSSMVIN QVI LSSMVIN QVI	191 /RVUVAS SLS /RVULAS CLS /RVULAS CLS /RVULAS CLS /RVULAS CLS /RVULAS CLS /RUTIS CLS /	201 Aws TTNHP AAS LAGAP H Aws TONHP AFC NNNHP AFT IKKIP GFT LKNEP GFT LKNEP AFK AEGSS AFS LKGEP AKN CEGGS AKN CEGGS AKN CEGGS AKN CEGGS AKN LRGEP AKK HGEN FA SNNIK FA SNNIK GFCTLRNHP GFCTLRNHP GFCTLRNHP GFC LKGEP AFR NRGEP AFR SRGEP AFR SRGEP AFR SRGEP AFR SRGEP AFR INGEP AFT IRGEP AFT IRGEP	RR 154 RR 155 K 153 RH 155 RH 160 HR 143 F RR 143 F RR 143 F RR 143 F RR 143 F RR 145 F RR 145 RR 145 F RR 145 RR 159 RR 159 RR 159 RK 159 RR 159 RR 159 RR 159 RR 159 RR 159 RR 155 RR 155 RH
E A B	<pre>lns5 Esccol.gi 16128619 Roslit.gi 63732522 Burdol.gi 84362417 Camjej.gi 15791514 Edebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 8800363 Baccap.gi 154496020 Pedobacter_sp.gi 149279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108004498 Themar.gi 15805646 Akkmuc.gi 1940575 Solusi.gi 16624738 Synwol.gi 16081075 Cloace.gi 15896772 Mycpul.gi 15406014 Methun.gi 8807833 Orysat.gi 15400408 Aratha.gi 1540408 Aratha.gi 15428214 Phypat.gi 154060349</pre>	74 74 75 77 79 75 66 78 75 66 78 75 66 78 75 71 74 64 64 67 63 78 78 78 78 78 78 78 78 78 78 78 72 74	131 IP- KPWD ER- ROWT ER- ROWT ER- ROWT ER- SVLD ER- SVLD ER- SVLD ER- SLD ER- SLD ER- SLST IE- SKLT IE- SKLT RTT RYG KR- GEVS IE- KEIT RTT RYG KR- GEVS IE- SLFA ER- SCUST IE- SKNT IE- SKNT IE- SKNT IL- SKNT RT- SKOVS EN- SKNT IL- SKNT RT- SKOVS ER- SKLT RT- SKNT IE- SKLT	141 PQLA-AELERWH PDFA-GRIGGWF IMQLA-HALPGWC IEFA-KLIQD IQFS-KKVENII VKWA-AHLER VEFS-EYLQKKN EELS-RVTDWA VUFS-DYLNKQN SUFS-DYLNKQN SUFS-DYLNKQN SELA-ARIGGLA SELA-RAIDGLA SELA-RAIDGLA SELA-RAIDGLA SELA-NIGYH SAFA-AMISKG EELA-RULAKDW VEGA-KWLEKKA SELA-NIGYH SAFA-AMISKG EELA-DTIDKLA SEFS-AFIDDLA SEFS-AFIDDLA SELA-VIRCI SELA-VIRCI SELA-VIRCI SELA-VIRCI SELA-VIRCI SELA-VIRCI SELA-VIRCI SELA-DUXDAA	a4 151 XLDGR-DVSLL XDQGRGDLAW QQGR-DVAM QQGR-DVAM QQGR-DVAM QQGRSQLC VANAGKQLC VANAGKQLC VANAGKQLC VANAGKQLC VANAGKQLC VANAGKQLC VANAGKQLC VANAGKAL VANAGKA	β4 161 CGPE LSPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD LDPA(CGAD VADE' CGAD TAE	-a5 - 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFRLS VRKRADLKVA VRREAAACWS VKRANCSLS IVERANGSLS IVERANGSLS IVERANGSLS IFAKAHRVFS VKNSVDMKLK LRGQAKLLWS IFAKAHRVFS VKNSVDMKLK LRGQAKLLWS ULRSRCHHLS EVSRANHLS EVSRANHLS EVSRANHLS EVSRANHLS CLQKGNILS VLRAAVKIN ISEISDFKLS CLQKGNILS VERAAVTIR VERAAVTIR VERAAVTIR	β5 181 LSALTLP PLI SSMTLP GM LSKLTLA QFV LSKLTLA QFV LSKLTLA QFV LSKLTLA QFV LSKLTLP QLV LSKLTLP QLV LSKTTP QLV SSMTFP ELI LSPTFP ELI LSPTFP ELI LSKTTP QLV FSKTFP QLV FSKMTFP QLV FSKMTFP QLV FSKMTFP QLV FSKMTFP QLV SSMVLN QVI LSSMVLN QVI LSSMVLN QVI LSSMVLN QVI	191 (RVULAS SLO (RVULAS CLY (RVULAS CLY (RVULAS CLY (RVULAS CLY (RVULAS CLY (RULE) CLY (RUMILS CLY (RUMILS CLY (RUMILS CLY (RVULS CLY (RVU	201 Aws TTNHP AAS LAGAP AWS TQNHP AFT IKITP AFT IKITP AFT IKITP AFT IKKEP AFT LKNEP AFT LKNEP AFK AEGSS AFS LKGEP AVK CEGGS AKN CEGGS AFX INGEP AVK LRGEP AFK IHGEN WFA SNNIK AAT SAGEP GFT LTGHP GFT LTGHP GFT LTGHP AFK NRGEP AFK NRGEP AFK NRGEP AFF NRGEP AFF SRGEP AFF SRGEP AFF SRGEP AFF SRGEP AFF SRGEP AFF INGEP AFT INGEP AFT INGEP	RR 154 RR 155 K 153 RH 155 RH 155 RH 160 F 160 RH 160 F 160 RR 143 F 160 RR 143 F 160 RR 143 F 155 RH 155 RH 145 F 155 RH 145 F 155 RH 145 F 155 RH 155 RH 155 RH 155 F
E A B	<pre>Ins5 Esccol.gi 16128619 Roslit.gi 163732522 Burdol.gi 84362417 Camjej.gi 15791514 Bdebac.gi 42525168 Marfer.gi 114776729 Polirg.gi 88803363 Baccap.gi 154496020 Pedobacter_sp.gi 140279619 Synechococcus_sp.gi 116074639 Promar.gi 159903250 Bifden.gi 171741054 Rubxyl.gi 108804498 Themar.gi 15643607 Fusnuc.gi 19703798 Deirad.gi 15805646 Akkmuc.gi 16624738 Synvol.gi 114567973 Bacsub.gi 10681075 Cloace.gi 15896772 Mycpul.gi 15828725 Metmar.gi 15406004 Methun.gi 86603833 Orysat.gi 11540408 Aratha.gi 15238214 Phypat.gi 16806349</pre>	74 75 77 76 69 75 65 66 78 75 71 74 64 64 64 64 67 78 78 78 78 78 78 78 78 78 78 78 78 78	131 IP-KFWD ER- RAVES ER- RAVES ER- RAVES ER- RAVES ER- RALLE DK- KHYS VE- KSIT EQ- KTID EE- ESLT IE- KKYT RKT RRYG KR- EEVS IE- KKIT RKT RAYG RK- EEVS IE- KKIT ER- SKUM ER- KAVS IL- KKIT IK- KHYS IN-AEYS IL- KKIT IK- KMFS IN-AEYS IL- KKIT ER- SKUT ER- SKLT ER- SKLT ER- SKLT	141 PQLA-AELERWH PDFA-GRLGGWF MQLA-HALPGWC TEFA-KLIQD IQFS-KKVENII VKWA-AHLER VEFS-EYLQKKN BELS-RVTDK VLFS-DYLNKQN VJFASRLC LGFANRLF SELA-AKIDGLC SELA-KIDGLC SELA-KLIGLC SELA-KLIGLC SELA-TIGLC SELA-TIGLC SELA-TIGLC SELA-KINDLC SELA-KINDLC SEGLA-EVIRCC SEQA-LFLCGA- SEGLA-CHIRCL SECA-KURCLS SECA-STIDLC SECA-KURCLS SECA-STIDLC SECA-	a4 151 LDGR-DVSLL AUDGR-DVSLL AUDGR-DVAM AUDGR-DVAM AUGGRGLANV AGOGRSQLCIV AGOGRSQLLIV AGOGRSQLIV AG	β4 161 CGPE LSPAC GAD IDPAI GAD IDPAI GAD IDPAI GAT UPEI GAT VNEE GAT VNEE CGPY FSDD CSY FDER GAD ITAE CSAN LDSE CGPY FDER CGAD ITAE CSAN LDSE CGPL LAPE CGPD ITAA CASD HTEE CSAN FTEE CSAN FTEE CSAN CLSE CGSL LST CGHD ITAA CASD HTEE CSAN FTEE	-a5 171 CKAAAEQSWS LRAEADFALS LKARADLLLR FNQSLDFALS VYKKADLKVS VYKKAALKMS VYKKAALKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS VYKKANEKMS VIERADENS VIERADENS VIERADENS VIERADENS VIERADENS VIERADENS VIERADENS VIERANVII SEISDFALS VIERANVII	β5 181 LSALTPE PLI LSALTPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSKITPE QMI LSMITPE QMI LSKITPE QMI LSKITPE QLI LSKITPE QLI FSKMTPE QLI FSKMTPE QLI LSSMITPE QUI LSSMITPE QUI LSSMITPE QUI LSSMITPE QUI LSSMITPE QUI LSSMITME QMI LSSMITME QMI LSS	191 (RVUAS SL (RVULAS CL (RVULAS CL (RVULAS CL (RVULAS CL (RVULS CL (RUFUS CL (R	201 AWS TTNHPY ARS LAGAPY ARS LAGAPY AFC NNNHPY AFT IKKIPY AFT IKKIPY AFT ICGHPY GFT LKNEP AFK AEGSS AYK ACGES AYK NAGEP AVK LRGEP AVK LRGEP AVK LRGEP AT SAGEP CHT LAGTP GFCTLRNHP I GFCTLRNHP I GFCTLRNHP I GFC LRMP ARS SNE AFF SC AFF ACGEP AFF ACGEP AFF ACGEP AFF ACGEP AFF ACGEP AFF ACGEP AFF ACGEP AFF ACGEP AFT INGEP AFT INGEP AFT INGEP AFT INGEP	R R 154 154 155 155 155 155 155 155 155 155

FIGURE 5. Sequence alignment of proteins belonging to the COG1576 cluster of SPOUT methyltransferases superfamily. The domains of organisms are indicated on the *left* as follows: (B) Bacteria; (A) Archaea; (E) Eukaryota. Each class of organisms is represented by one sequence, denoted by species name and NCBI gene identification (GI) number. Columns of residues are shaded according to the percent of identity: (dark gray) 100% identity; (light gray) 90% identity. Secondary structure elements derived from *E. coli* YbeA protein crystal structure (1NS5) are shown on *top* of the alignment, together with predicted SAM binding motif (Anantharaman et al. 2002). The alignment was performed using MUSCLE (Edgar 2004), and the figure was generated with Jalview (Clamp et al. 2004).

methyltransferase, TrmH, was recently shown (Watanabe et al. 2005). When the set of aligned sequences was widened to include a representative member from each family of organisms having a putative COG1576 member (95

sequences in total), the number of totally conserved amino acids was reduced to four—G107, E138, R142, and Y152 (Supplemental Fig. 1). As one of the four universally conserved amino acids was R142, it further supports the possible importance of R142 in YbeA catalytic activity.

DISCUSSION

An extensive bioinformatical analysis predicted the product of the *ybeA* gene to be a member of the SPOUT superfamily of methyltransferases characterized by an intricate trefoilknot structure (Anantharaman et al. 2002; Tkaczuk et al. 2007). Several members of the SPOUT superfamily are confirmed to exhibit Gm (e.g., RlmB responsible for Gm2251 in 23S rRNA) and m¹G (e.g., TrmD responsible for m¹G37 in tRNA) methyltransferase activity in *E. coli*. RsmE protein, also a member of the SPOUT superfamily of methyltransferases according to recent classification (Tkaczuk et al. 2007), has an m³U methyltransferase activity (responsible for m³U1498 in 16S rRNA). Since corresponding enzymes for all of the Gm, m¹G, and m³U modifications in E. coli rRNA have been accounted for, the remaining putative RNA methyltransferases of the SPOUT superfamily are speculated to synthesize modifications of different chemical nature (such as $m^{3}\Psi$), or on different RNA substrates (Tkaczuk et al. 2007).

We have determined that the gene product of *ybeA* is responsible for the post-transcriptional modification of Ψ 1915 to m³ Ψ in *E. coli* 23S rRNA. Cells lacking the *ybeA* gene carried non-methylated Ψ at position 1915. Reintroduction of the ybeA gene into the knockout strain restored the corresponding $m^{3}\Psi$ methylation. Furthermore, purified YbeA protein was shown to synthesize $m^3\Psi$ methylation in vitro using ribosomes purified from the ybeA knockout strain as a substrate. Notably, YbeA protein was not able to methylate ribosomes isolated from a $\Delta ybeA/$ Δ rluD double deletion strain. YbeA-directed methylation was activated by prior treatment of ribosomes with pseudouridine synthase RluD, suggesting that the substrate for YbeA is pseudouridine at position 1915. The identity of $m^{3}\Psi$ was determined by RP-HPLC using 23S rRNA fragment 1778-1921. It was important to use a fragment of 23S rRNA that does not contain m⁵C (present at position 1962) due to the similar retention times during reversed phase chromatography. These results unambiguously establish YbeA as the methyltransferase that methylates Ψ 1915 in *E*. coli 23S rRNA to m³ Ψ .

Ribosome biogenesis is known to occur in a stepwise manner with some modifications synthesized on newly transcribed rRNA prior to its assembly into subunits (Liu et al. 2004). At the same time, synthesis of other modifications requires partially, or even fully assembled subunits (for review, see Ofengand and Del Campo 2004). RluD has been shown to be far more efficient in modifying H69 in structured 50S subunits, compared to free 23S rRNA; a low level of activity was seen even with 70S ribosomes (Vaidyanathan et al. 2007). Analysis of the pseudouridylation pattern of 23S rRNA of ribosome assembly precursor particles has shown that the pseudouridines in H69 are formed by RluD during the late assembly steps (Leppik et al. 2007). Given that synthesis of Ψ 1915 by RluD is likely to precede the methylation of the same position by YbeA, we propose that YbeA is also acting during the late step of ribosomes assembly. Furthermore, we show that both RluD and YbeA proteins are active on 70S ribosomes. YbeA does not methylate the pseudouridine on free 50S subunits. 70S ribosomes are formed during the initiation step of protein synthesis. Purta et al. (2008) have found that docking of the YbeA crystal structure onto the 70S ribosomes suggests extensive contacts with both ribosome subunits, without any clash with the P-site tRNA (Purta et al. 2008). These data suggest that the last steps of ribosome assembly overlap with the first step of translation. We speculate that YbeA is involved in the quality control of ribosome biogenesis. Interestingly, deletion of the *vbeA* gene has an effect on the modification level of nearby nucleosides. The absence of YbeA protein leads to slightly reduced amounts of all modified nucleosides in the 23S rRNA fragment 1778-1921 (Table 1). This points to possible interplay between modification enzymes RluD, RlmG, and RlmH during ribosome biogenesis.

The amino acid sequence of the YbeA protein has two C-terminal motifs that map to the SAM binding pocket, characteristic signatures of proteins belonging to the SPOUT superfamily. Proteins with highly significant similarity to YbeA exist in virtually all bacteria and are also found in genomes of green plants, as well as in the archaea phylum Euryarchaeota (Fig. 5). The YbeA homologs likely perform the same function in these organisms. The presence of the $m^{3}\Psi$ modification in 23S rRNA was experimentally determined in E. coli (Kowalak et al. 1996). It is worth noting that Bacillus subtilis and Zea mays chloroplast ribosome large subunit rRNAs are very likely to contain the equivalent of $m^{3}\Psi$ 1915 as revealed by a strong reverse transcriptase stop (Ofengand et al. 1995). The extent of conservation of the $m^{3}\Psi$ modification at the position equivalent to E. coli 1915 within the evolutionary domains is unknown.

Although both pseudouridines and methylated residues are widely distributed in rRNA, $m^3\Psi$ is the only derivative of pseudouridine thus far found in bacterial rRNA. Furthermore, to our knowledge, $m^3\Psi$ in 23S rRNA position 1915 (or equivalent) is the only example of methylated pseudouridine described in bacterial RNAs (including extensively modified tRNAs), making it in that sense "unique" (Rozenski et al. 1999; The RNA Modification Database, http://library.med.utah.edu/RNAmods/). The YbeA protein is hence a candidate for the only pseudouridine-specific methyltransferase in bacteria. Eukaryotic rRNAs also possess $m^1\Psi$ (Brand et al. 1978), Ψm (Gray and Keddy 1974; Maden and Salim 1974), and the hypermodified derivative $m^1 acp^3\Psi$ (Saponara and Enger 1974; Maden et al. 1975). To our knowledge, no pseudouridine-specific methyltransferase has been described in eukaryotes.

The crystal structure of the YbeA protein has been determined (PDB 1NS5) (Benach et al. 2003), and like other SPOUT methyltransferases, it displays the α/β knot fold consisting of six-stranded parallel β-sheets flanked by α -helices and an unusual C-terminal trefoil knot structure (Forouhar et al. 2003; Mallam and Jackson 2006). The deep trefoil knot in the YbeA backbone is formed by the threading of the last 35 residues (120-155) through a 45residue knotting loop (residues 74–119). YbeA crystallizes as an antiparallel homodimer, and the protein interface involves close-packing $\alpha 1$ and $\alpha 5$ from each monomer. Dimerization is consistent with other SPOUT methyltransferase crystal structures, although in some cases (e.g., RrmA and RsmE) the monomers are nearly perpendicular instead of being antiparallel. Dimerization is thought to be important for methyltransferase function as the knotted topology alone is not sufficient for maintaining the active conformation of the cofactor and substrate binding site (Nureki et al. 2004; Mallam and Jackson 2007a,b).

YbeA is composed solely of the core catalytic domain characteristic of SPOUT methyltransferases and lacks the extra N-terminal and C-terminal domains exhibited by many SPOUT methyltransferases (Benach et al. 2003; Tkaczuk et al. 2007). Comprehensive structural genomics data on the YbeA protein can be found in the accompanying paper by Purta et al. (2008).

The broad distribution of YbeA homologs among different species and the "uniqueness" of $m^3\Psi$ modification imply that there must have been a compelling reason for the cells to evolve an enzyme catalyzing this modification. What could be the functional purpose of this methylation remains to be determined.

MATERIALS AND METHODS

Strains and plasmids

Single gene knockout mutants of *E. coli* strain BW25113 (Keio collection) were obtained from the Nara Institute of Science and Technology. In each of the 11 mutant strains studied (JW5107, JW1123, JW0200, JW0203, JW4268, JW4366, JW2565, JW3581, JW0631, JW2777, JW0904), one putative RNA methyltransferase gene is replaced with a kanamycin-resistance cassette (Baba et al. 2006).

The MG Δ ybeA (RE111) strain was generated by transducing the *E. coli* wild-type MG1655 strain (Blattner et al. 1997) with lysate derived from a P1 phage-infected Δ ybeA strain (JW0631, Keio collection), followed by selection of colonies based on acquired kanamycin resistance. Similarly, the Δ ybeA/ Δ rluD doubleknockout strain (RE112) was generated by transducing the *E. coli* Δ rluD (Leppik et al. 2007) strain with lysate derived from the P1 phage-infected Δ ybeA strain, followed by selection of colonies based on both kanamycin and chloramphenicol resistance. Constructed strains were confirmed for *ybeA* and/or *rluD* deletion by PCR analysis.

For generation of the YbeA protein expression plasmid, the ybeA gene was amplified by PCR from genomic DNA of the E. coli MG1655 strain. The N-terminal primer was 5'-CCATCGGATCC AAGCTGCAACTTGTCGCCG-3', which introduced a BamHI restriction site. The C-terminal primer was 5'-GGCTCAAGCTTT CACTCACGGTGATAAGGATGG-3', and introduced a HindIII restriction site. The purified PCR fragment was digested with appropriate restriction enzymes and cloned into pQE-30 expression vector (QIAGEN) between the BamHI and HindIII restriction sites. The constructed plasmid pQE30-ybeA contained the ybeA gene coding for a protein with an N-terminal His-tag. Plasmid pQE60-rluD expressing RluD protein with a C-terminal His-tag was constructed as follows: The *rluD* gene was amplified by PCR from genomic DNA of the E. coli MG1655 strain. The Nterminal primer was 5'-GCCCATGGCACAACGAGTACAGCTC ACTGCA-3', which introduced an NcoI restriction site. The Cterminal primer was 5'-GCAGATCTTAACCAGTCCACTCCAT CCT-3', and introduced a BgIII restriction site. The purified PCR fragment was digested with appropriate restriction enzymes and cloned into the pQE-60 expression vector (QIAGEN) between the NcoI and BgIII restriction sites. Constructed plasmids were verified by sequencing.

For complementation analysis, the Δ ybeA strain was transformed with a pQE30-ybeA plasmid. Standard techniques were used for DNA manipulations, plasmid DNA isolation, and *E. coli* transformation (Sambrook et al. 1989).

Preparation of ribosomes and rRNA

Bacterial strains were grown at 37°C in 2xYT medium (16 g of tryptone, 10 g of yeast extract, 5 g of NaCl per liter) supplemented with kanamycin (50 µg/mL), chloramphenicol (20 µg/mL), or ampicillin (100 µg/mL). Ribosomes were isolated from exponentially growing cells at $OD_{600} = 0.8$ Š1.0. Bacteria were collected by low-speed centrifugation and resuspended in buffer TKNMS (16% sucrose [w/v] in 6 mM MgCl₂, 60 mM NH₄Cl, 60 mM KCl, 50 mM Tris/HCl at pH 8.0, and 6 mM β-mercaptoethanol). Cells were lyzed by five freeze-thaw cycles in the presence of lysozyme (final concentration 0.5 µg/mL) and DNase I (final concentration 50 units/mL). The S-30 lysate was prepared by centrifugation at 12,000g for 30 min in an SS34 rotor (Sorvall), and the volume of the lysate was increased twofold with TKNM buffer (12 mM MgCl₂, 60 mM NH₄Cl, 60 mM KCl, 20 mM Tris/HCl at pH 8.0, and 6 mM β -mercaptoethanol). A total of 150 U (OD₂₆₀) of S-30 lysate was layered onto a 15%-30% (w/w) sucrose gradient in TKNM buffer followed by centrifugation at 21,000 rpm for 17 h in a Beckman SW-28 rotor ($\omega^2 t = 2.96 \times 10^{11}$). Gradients were analyzed with continuous monitoring of absorbance at 254 nm. Ribosomal particles from gradient fractions were precipitated with 2.5 volumes of ice-cold ethanol and collected by centrifugation at 5000 rpm for 30 min in an HS4 rotor (Sorvall). Ribosomal pellets were dissolved in TKNM buffer and used for RNA preparation. 70S ribosomes were prepared as described above using sucrose gradient centrifugation in TKNM buffer. Gradient fractions containing 70S ribosomes were collected, and ribosomal particles were sedimented by centrifugation at 39,000 rpm for 20 h ($\omega^2 t$ = 1.2×10^{12}) in a Beckman Ti50.2 rotor. 50S subunits were obtained by dissociating 70S ribosomes. Ribosomal pellets were dissolved in TKNM buffer and stored at -80°C.

For primer extension analysis, the RNA was isolated from ribosome particles as described in Liiv et al. (2005). In the case of

HPLC analysis, the rRNA was purified by extraction with phenol and chloroform followed by ethanol precipitation. rRNA was dissolved in water and stored at -80° C.

Primer extension analysis

Primer U1 (CAGCCTGGCCATCATTACGCC) complementary to positions 1972–1992 of *E. coli* 23S rRNA was annealed to rRNA extracted from 70S ribosomes. The primer was extended by AMV reverse transcriptase (Seikagaku Corp.) according to the manufacturer's protocol using $[\alpha-^{32}P]dCTP$ (Amersham Biosciences). Primer extension products were precipitated with ethanol; dissolved in loading buffer containing formamide, bromophenol blue, and xylene cyanol; and separated on a 7% polyacrylamide/ 8 M urea denaturing gel. The gel was transferred to Whatman 3MM paper and vacuum-dried. Radioactivity was visualized by a Typhoon PhosphorImager (GE Healthcare).

Sequencing of *E. coli* 23S rDNA was done using a U1 primer and CycleReader DNA Sequencing Kit (Fermentas Life Sciences).

HPLC analysis

Preparation of rRNA fragment

A fragment of 23S rRNA corresponding to nucleotides 1778-1921 was excised by RNase H using oligonucleotides complementary to nucleotides 1760-1777 (primer C4: 5'-CAGTTGCAGCCAGC TGG-3') and 1922-1942 (primer U1 mini: 5'-TTTCGCTACCTT AGGACCG-3') of E. coli 23 S rRNA essentially as described by Douthwaite and Kirpekar (2007). In the denaturation step, 300 pmol of rRNA was mixed with a 10-fold molar excess of both oligodeoxynucleotides and heated for 3 min at 100°C in 270 µL of 1 mM EDTA. Denatured RNA probes were placed on ice, and 30 μL of 10× buffer (600 mM HEPES at pH 7.0 and 1.25 M KCl) were added. In the hybridization step, the reactions were heated for 1 min at 90°C and cooled in a water bath over the period of 2 h to 45°C. The resulting RNA-DNA hybrids were digested with 10 units of RNase H (Fermentas) in the presence of 8 mM MgCl₂ and 1 mM DTT for 30 min at 37°C to remove RNA/DNA heteroduplexes. Nuclease-treated RNA was phenol-extracted and recovered by ethanol precipitation, and the 143-nucleotide-long rRNA fragment was gel-purified using a 5% LE TOP agarose gel. The RNA fragment was excised and extracted from the gel by overnight incubation in 450 µL of 2M NH₄OAc (pH 6.0) at 4°C. The RNA was precipitated with two volumes of 1:1 mixture of ethanol and isopropanol, collected by centrifugation, and dissolved in water.

High performance liquid chromatography

For HPLC analysis, 100–200 pmol of gel-purified RNA fragment were digested with nuclease P1 (MP Biochemicals) and bacterial alkaline phosphatase (Fermentas Life Sciences) according to the method of Gehrke and Kuo (1989). Nucleoside composition was determined by RP-HPLC on a Supelcosil LC-18-S HPLC column (25cm \times 4.6mm, 5 µm) equipped with a pre-column (4.6mm \times 20mm) at 30°C on a SHIMADZU Prominence HPLC system.

The following buffers were used: buffer A (10 mM $NH_4H_2PO_4$, 2.5% methanol at pH 5.3), buffer B (10 mM $NH_4H_2PO_4$, 20% methanol at pH 5.1), and buffer C (10 mM $NH_4H_2PO_4$, 35% acetonitrile at pH 4.9). RP-HPLC analysis was performed using the gradient conditions of Gehrke and Kuo (1989): flow rate 1.0

mL/min held at 0%B 12 min, to 10%B over 8 min, to 25%B over 5 min, to 60%B over 8 min, to 64%B over 4 min, to 100%B over 9 min, 0%–100%C over 35 min, held at 100%C for 10 min, and equilibration with 0%B for 30 min. Nucleoside absorbance profiles were recorded at 260 nm, and peak areas were integrated. For quantitative calculations, the following molar extinction coefficients (at pH 4.1) were used: 13,100 for adenosine, 10,300 for guanosine, 6800 for cytidine, and 9000 for uridine (Holness and Atfield 1971).

Purification of recombinant proteins

Recombinant N-terminal His6-tagged YbeA and RluD proteins were prepared from E. coli M15 cells (QIAGEN), harboring the pQE30-ybeA or pQE60-rluD plasmid. Cells were grown in 400 mL of 2xYT liquid media containing 50 µg/mL kanamycin and 100 μg/mL ampicillin at 37°C to an OD₆₀₀ of 0.5. Isopropyl-β-Dthiogalactopyranoside (IPTG) at 1 mM was added, and incubation continued for 2 h. Cells were harvested by low-speed centrifugation; resuspended in lysis buffer (50 mM NaH₂PO₄, 1 M NaCl, 10% glycerol, 10 mM imidazole at pH 7.0); incubated on ice for 30 min with lysozyme (final concentration 1 mg/mL) and DNase I (final concentration 40 units/mL); and were passed through a French pressure cell at 18,000 psi. Cell debris was removed by centrifugation at 16,000 rpm for 30 min in a SS-34 Sorvall rotor. The recombinant YbeA protein was purified by nickel ion affinity chromatography on a Ni Sepharose 6 Fast Flow column (GE Healthcare Life Sciences) according to the manufacturer's instructions. Fractions containing YbeA protein were pooled; analyzed by SDS-PAGE; and dialyzed for 72 h against buffer: 10 mM Tris/HCl at pH 7.6, 50 mM NH₄Cl, 6 mM β-mercaptoethanol, and 1 mM EDTA. Purified protein was concentrated with an Amicon Ultra Ultracel-3k filter, frozen in liquid nitrogen, and stored at -80°C in dialysis buffer containing 50% glycerol.

In vitro methylation assay

Methyltransferase activity of YbeA protein in vitro was tested as follows: A reaction mixture (75 μ L) containing 24 pmol of 70S ribosomes or 50S subunits, 3 μ g of purified YbeA protein, 100 μ M [¹⁴C]-*S*-adenosyl-L-methionine (Amersham Pharmacia Biotech), and methylation buffer (50 mM Tris/HCl at pH 8.0, 100 mM NH₄Cl, 10 mM MgCl₂, and 1 mM DTT) was incubated for 1 h at 37°C. When indicated, pre-incubation of ribosomes with 1 μ g of purified RluD protein in the methylation buffer for 30 min at 37°C was performed. Reaction products were precipitated with 2 mL of ice-cold 5% TCA. Samples were incubated for 30 min on ice prior to collection on glass fiber filters (Whatman). Filters were washed with 10 mL of 5% TCA, followed by 2 mL of 70% ethanol, and allowed to dry. Radioactivity was determined by scintillation counting using Optiphase HiSafe III scintillator (PerkinElmer).

Preparative methylation of 300 pmol of 70S ribosomes for HPLC analysis was performed in the same conditions.

Multiple sequence alignment

Multiple sequence alignment was done with MUSCLE (Edgar 2004). Sequences for alignment were selected from COG (Tatusov et al. 2003) and the RefSeq database (Pruitt et al. 2007) as of March 2008 so that each class of organisms having the annotated

putative COG1576 member is represented by a single sequence. Taxonomy is based on The NCBI Entrez Taxonomy Homepage (http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy) as of March 2008.

SUPPLEMENTAL DATA

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We thank Dr. H. Mori from the Nara Institute of Science and Technology, Japan for knockout strains; Professor S. Douthwaite (University of Southern Denmark) for useful discussions and comments on the manuscript; Dr. Ü. Maiväli and Dr. T. Tenson (both University of Tartu) for critically reading the manuscript; and M. Leppik for help and advice. The work was supported by Estonian Science Foundation Grants 7509 (to J.R.) and 7072 (to A.L.).

Received May 16, 2008; accepted July 9, 2008.

REFERENCES

- Agrawal, R.K., Sharma, M.R., Kiel, M.C., Hirokawa, G., Booth, T.M., Spahn, C.M.T., Grassucci, R.A., Kaji, A., and Frank, J. 2004. Visualization of ribosome-recycling factor on the *Escherichia coli* 70S ribosome: Functional implications. *Proc. Natl. Acad. Sci.* 101: 8900–8905.
- Ali, I.K., Lancaster, L., Feinberg, J., Joseph, S., and Noller, H.F. 2006. Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Mol. Cell* 23: 865–874.
- Anantharaman, V., Koonin, E.V., and Aravid, L. 2002. SPOUT: A class of methyltransferases that includes spoU and trmD RNA methylase superfamilies and novel superfamilies of predicted prokaryotic RNA methyltransferases. J. Mol. Microbiol. Biotechnol. 4: 71–75.
- Andersen, N.M. and Douthwaite, S. 2006. YebU is a m5C methyltransferase specific for 16S rRNA nucleotide 1407. *J. Mol. Biol.* **359:** 777–786.
- Andersen, T.E., Porse, B.T., and Kirpekar, F. 2004. A novel partial modification at C2501 in *Escherichia coli* 23S ribosomal RNA. *RNA* 10: 907–913.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Methods Mol. Biol.* 416: 171–181.
- Basturea, G.N. and Deutscher, M.P. 2007. Substrate specificity and properties of the *Escherichia coli* 16S rRNA methyltransferase, RsmE. *RNA* 13: 1969–1976.
- Basturea, G.N., Rudd, K.E., and Deutscher, M.P. 2006. Identification and characterization of RsmE, the founding member of a new RNA base methyltransferase family. *RNA* **12**: 426–434.
- Benach, J., Shen, J., Rost, B., Xiao, R., Acton, T., Montelione, G., and Hunt, J.F. 2003. X-Ray structure of Ybea from *E. coli*. Northeast Structural Genomics Research Consortium (NESG), PDB: 1NS5. European Bioinformatics Institute, European Molecular Biology Laboratory, Grenoble, France.
- Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 5: 1453–1474.
- Brand, R.C., Klootwijk, J., Planta, R.J., and Maden, B.E.H. 1978. Biosynthesis of a hypermodified nucleotide in *Saccharomyces*

- Brimacombe, R., Mitchell, P., Osswald, M., Stade, K., and Bochkariov, D. 1993. Clustering of modified nucleotides at the functional center of bacterial ribosomal RNA. *FASEB J.* **7:** 161– 167.
- Caldas, T., Binet, E., Bouloc, P., Costa, A., Desgres, J., and Richarme, G. 2000. The FtsJ/RrmJ heat shock protein of *Escherichia coli* is a 23S ribosomal RNA methyltransferase. J. Biol. Chem. 275: 16414–16419.
- Clamp, M., Cuff, J., Searle, S.M., and Barton, G.J. 2004. The Jalview Java alignment editor. *Bioinformatics* **20**: 426–427.
- Cundliffe, E. 1989. How antibiotic-producing organisms avoid suicide. Annu. Rev. Microbiol. 43: 207–233.
- Cunningham, P.R., Richard, R.B., Weitzmann, C.J., Nurse, K., and Ofengand, J. 1991. The absence of modified nucleotides affects both in vitro assembly and in vitro function of the 30S ribosomal subunit of *Escherichia coli*. *Biochimie* **73**: 789–796.
- Decatur, W.A. and Fournier, M.J. 2002. rRNA modifications and ribosome function. *Trends Biochem. Sci.* 27: 344–351.
- Douthwaite, S. and Kirpekar, F. 2007. Identifying modifications in RNA by MALDI mass spectrometry. *Methods Enzymol.* 425: 1–20.
- Edgar, R.C. 2004. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.
- Ejby, M., Sørensen, M.A., and Pedersen, S. 2007. Pseudouridylation of helix 69 of 23S rRNA is necessary for an effective translation termination. *Proc. Natl. Acad. Sci.* **104**: 19410–19415.
- Elkins, P.A., Watts, J.M., Zalacain, M., van Thiel, A., Vitazka, P.R., Redlak, M., Andraos-Selim, C., Rastinejad, F., and Holmes, W.M. 2003. Insights into catalysis by a knotted TrmD tRNA methyltransferase. J. Mol. Biol. 333: 931–949.
- Fauman, E.B., Blumenthal, R.M., and Cheng, X. 1999. Structure and evolution of AdoMet-dependent methyltransferases. In S-Adenosylmethionine-dependent methyltransferases: Structures and functions (eds. X. Cheng and R.M. Blumenthal), pp. 1–38. World Scientific, River Edge, NJ.
- Forouhar, F., Shen, J., Xiao, R., Acton, T.B., Montelione, G.T., and Tong, L. 2003. Functional assignment based on structural analysis: Crystal structure of the yggJ protein (HI0303) of *Haemophilus influenzae* reveals an RNA methyltransferase with a deep trefoil knot. *Proteins* 53: 329–332.
- Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J., and Penczek, P. 2000. Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 100: 537–549.
- Gehrke, C.W. and Kuo, K.C. 1989. Ribonucleoside analysis by reversed-phase high-performance liquid chromatography. J. Chromatogr. 471: 3–36.
- Gray, M.W. and Keddy, G.M. 1974. Presence of O^{2'}-methylpseudouridine in the 18S + 26S ribosomal ribonucleates of wheat embryo. *Biochemistry* 13: 5453–5463.
- Green, R. and Noller, H.F. 1996. In vitro complementatation analysis localizes 23S posttranscriptional modifications that are required for *Escherichia coli* 50S subunit assembly and function. *RNA* **2**: 1011–1021.
- Green, R. and Noller, H.F. 1999. Reconstitution of functional 50S ribosomes from in vitro transcripts of *Bacillus stearothermophilus* 23S rRNA. *Biochemistry* 38: 1772–1779.
- Gustafsson, C. and Persson, B.C. 1998. Identification of the rrmA gene encoding the 23S rRNA m1G745 methyltransferase in *Escherichia coli* and characterization of an m1G745-deficient mutant. J. *Bacteriol.* 180: 359–365.
- Hirabayashi, N., Sato, N.S., and Suzuki, T. 2006. Conserved loop sequence of helix 69 in *Escherichia coli* 23 S rRNA is involved in A-site tRNA binding and translational fidelity. *J. Biol. Chem.* 281: 17203–17211.
- Holness, N.J. and Atfield, G. 1971. Chromatographic separation of mononucleotides derived from transfer ribonucleic acids. *Biochem.* J. 121: 371–377.

- Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P.J., and Santi, D.V. 1998. Identification of two *Escherichia coli* pseudouridine synthases that show multisite specificity for 23S RNA. *Biochemistry* 37: 15951–15957.
- Igarashi, K., Kishida, K., Kashiwagi, K., Tatokoro, I., Kakegawa, T., and Hirose, S. 1981. Relationship between methylation of adenine near the 3' end of 16-S ribosomal RNA and the activity of 30-S ribosomal subunits. *Eur. J. Biochem.* **113**: 587–593.
- Khaitovich, P., Tenson, T., Kloss, P., and Mankin, A.S. 1999. Reconstitution of functionally active *Thermus aquaticus* large ribosomal subunits with in vitro-transcribed rRNA. *Biochemistry* 38: 1780–1788.
- Kowalak, J.A., Bruenger, E., Hashizume, T., Peltier, J.M., Ofengand, J., and McCloskey, J.A. 1996. Structural characterization of U*-1915 in domain IV from *Escherichia coli* 23S ribosomal RNA as 3methylpseudouridine. *Nucleic Acids Res.* 24: 688–693.
- Krzyzosiak, W., Denman, R., Nurse, K., Hellmann, W., Boublik, M., Gehrke, C.W., Agris, P.F., and Ofengand, J. 1987. In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry* 26: 2353–2364.
- Leppik, M., Peil, L., Kipper, K., Liiv, A., and Remme, J. 2007. Substrate specificity of the pseudouridine synthase RluD in *Escherichia coli. FEBS J.* 274: 5759–5766.
- Liiv, A., Karitkina, D., Maiväli, U., and Remme, J. 2005. Analysis of the function of *E. coli* 23S rRNA helix-loop 69 by mutagenesis. *BMC Mol. Biol.* 6: 18.
- Liu, M., Novotny, G.W., and Douthwaite, S. 2004. Methylation of 23S rRNA nucleotide G745 is a secondary function of the RlmAI methyltransferase. *RNA* **10**: 1713–1720.
- Maden, B.E.H. and Salim, M. 1974. The methylated nucleotide sequences in HELA cell ribosomal RNA and its precursors. *J. Mol. Biol.* 88: 133–164.
- Maden, B.E.H., Forbes, J., de Jonge, P., and Klootwijk, J. 1975. The methylated nucleotide sequences in HELA cell ribosomal RNA and its precursors. *FEBS Lett.* **59**: 60–63.
- Mallam, A.L. and Jackson, S.E. 2006. Probing nature's knots: The folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359**: 1420–1436.
- Mallam, A.L. and Jackson, S.E. 2007a. A comparison of the folding of two knotted proteins: YbeA and YibK. J. Mol. Biol. 16: 650–665.
- Mallam, A.L. and Jackson, S.E. 2007b. The dimerization of an α/β -knotted protein is essential for structure and function. *Structure* **15:** 111–122.
- Mann, P.A., Xiong, L., Mankin, A.S., Chau, A.S., Mendrick, C.A., Najarian, D.J., Cramer, C.A., Loebenberg, D., Coates, E., Murgolo, N.J., et al. 2001. EmtA, a rRNA methyltransferase conferring high-level evernimicin resistance. *Mol. Microbiol.* 41: 1349–1356.
- Mitchell, P., Osswald, M., and Brimacombe, R. 1992. Identification of intermolecular RNA cross-links at the subunit interface of the *Escherichia coli* ribosome. *Biochemistry* **31**: 3004–3011.
- Noller, H.F. and Woese, C.R. 1981. Secondary structure of 16S ribosomal RNA. *Science* 212: 403–411.
- Nureki, O., Watanabe, K., Fukai, S., Ishii, R., Endo, Y., Hori, H., and Yokoyama, S. 2004. Deep knot structure for construction of active site and cofactor binding site of tRNA modification enzyme. *Structure* **12**: 593–602.
- Ofengand, J. and Del Campo, M. 2004. *EcoSal*—Escherichia coli and Salmonella: *Cellular and molecular biology* (ed. R. Curtiss), Chap. 4.6.1. ASM Press, Washington, DC.
- Ofengand, J. and Fournier, M.J. 1998. The pseudouridine residues of rRNA: Number, location, biosynthesis, and function. In *Modification and editing of RNA* (eds. H. Grosjean and R. Benne), pp. 229–253. ASM Press, Washington, DC.
- Ofengand, J., Bakin, A., Wrzesinski, J., Nurse, K., and Lane, B.G. 1995. The pseudouridine residues of ribosomal RNA. *Biochem. Cell Biol.* **73**: 915–924.

- Ofengand, J., Malhotra, A., Remme, J., Gutgsell, N.S., Del Campo, M., Jean-Charles, S., Peil, L., and Kaya, Y. 2001. Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harb. Symp. Quant. Biol.* 66: 147–159.
- Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2007. NCBI Reference Sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35: D61– D65.
- Purta, E., Kaminska, K.H., Kasprzak, J.M., Buinicki, J.M., and Douthwaite, S. 2008. YbeA is the $m^{3}\Psi$ methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA* (this issue). doi: 10.1261/rna.1198108.
- Raychaudhuri, S., Conrad, J., Hall, B.G., and Ofengand, J. 1998. A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of *Escherichia coli*. *RNA* **4**: 1407–1417.
- Rozenski, J., Crain, P.F., and McCloskey, J.A. 1999. The RNA Modification Database: 1999 update. *Nucleic Acids Res.* 27: 196– 197.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning:* A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Saponara, A.G. and Enger, M.D. 1974. The isolation from ribonucleic acid of substituted uridines containing α -aminobutyrate moieties derived from methionine. *Biochim. Biophys. Acta* 349: 61–77.
- Schubert, H.L., Blumenthal, R.M., and Cheng, X. 2003. Many paths to methyltransfer: A chronicle of convergence. *Trends Biochem. Sci.* 28: 329–335.
- Sergiev, P.V., Bogdanov, A.A., and Dontsova, O.A. 2007. Ribosomal RNA guanine-(N2)-methyltransferases and their targets. *Nucleic Acids Res.* 35: 2295–2301.
- Sergiev, P.V., Serebryakova, M.V., Bogdanov, A.A., and Dontsova, O.A. 2008. The ybiN gene of *Escherichia coli* encodes adenine-N6 methyltransferase specific for modification of A1618 of 23 S ribosomal RNA, a methylated residue located close to the ribosomal exit tunnel. *J. Mol. Biol.* **375**: 291–300.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. 1997. A genomic perspective on protein families. *Science* 278: 631–637.
- Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., et al. 2003. The COG database: An updated version includes eukaryotes. *BMC Bioinformatics* 4: 41. doi: 10.1186/1471-2105-4-41.
- Tkaczuk, K.L., Dunin-Horkawicz, S., Purta, E., and Bujnicki, J.M. 2007. Structural and evolutionary bioinformatics of the SPOUT superfamily of methyltransferases. *BMC Bioinformatics* 8: 73. doi: 10.1186/1471-2105-8-73.
- Toh, S.M., Xiong, L., Bae, T., and Mankin, A.S. 2008. The methyltransferase YfgB/RlmN is responsible for modification of adenosine 2503 in 23S rRNA. *RNA* 14: 98–106.
- Vaidyanathan, P.P., Deutscher, M.P., and Malhotra, A. 2007. RluD, a highly conserved pseudouridine synthase, modifies 50S subunits more specifically and efficiently than free 23S rRNA. *RNA* 13: 1868–1876.
- Watanabe, K., Nureki, O., Fukai, S., Ishii, R., Okamoto, H., Yokoyama, S., Endo, Y., and Hori, H. 2005. Roles of conserved amino acid sequence motifs in the SpoU (TrmH) RNA methyltransferase family. J. Biol. Chem. 280: 10368–10377.
- Weisblum, B. 1995. Erythromycin resistance by ribosome modification. Antimicrob. Agents Chemother. 39: 577–585.
- Xu, Z., O'Farrell, H.C., Rife, J.P., and Culver, G.M. 2008. A conserved rRNA methyltransferase regulates ribosome biogenesis. *Nat. Struct. Mol. Biol.* **15:** 534–536.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H., and Noller, H.F. 2001. Crystal structure of the ribosome at 5.5 Å. *Science* **292**: 883–896.