# The presence of the ACA box in archaeal H/ACA guide RNAs promotes atypical pseudouridylation

MRINMOYEE MAJUMDER,[1,2] SHAONI MUKHOPADHYAY,[1] PARINATI KHAREL,[3] and RAMESH GUPTA

Department of Biochemistry and Molecular Biology, Southern Illinois University, Carbondale, Illinois 62901-4413, USA

## ABSTRACT

**Archaea and eukaryotes, in addition to protein-only enzymes, also possess ribonucleoproteins containing an H/ACA guide RNA plus four proteins that produce pseudouridine (Ψ). Although typical conditions for these RNA-guided reactions are known, certain variant conditions allow pseudouridylation. We used mutants of the two stem–loops of the *Haloferax volcanii* sR-h45 RNA that guides three pseudouridylations in 23S rRNA and their target RNAs to characterize modifications under various atypical conditions. The 5′ stem–loop produces Ψ2605 and the 3′ stem–loop produces Ψ1940 and Ψ1942. The latter two modifications require unpaired "UVUN" (V = A, C, or G) in the target and ACA box in the guide. Ψ1942 modification requires the presence of U1940 (or Ψ1940). Ψ1940 is not produced in the Ψ1942-containing substrate, suggesting a sequential modification of the two residues. The ACA box of a single stem–loop guide is not required when typically unpaired "UN" is up to 17 bases from its position in the guide, but is needed when the distance increases to 19 bases or the N is paired. However, ANA of the H box of the double stem–loop guide is needed even for the 5′ typical pseudouridylation. The most 5′ unpaired U in a string of U's is converted to Ψ, and in the absence of an unpaired U, a paired U can also be modified. Certain mutants of the Cbf5 protein affect pseudouridylation by the two stem–loops of sR-h45 differently. This study will help elucidate the conditions for production of nonconstitutive Ψ's, determine functions for orphan H/ACA RNAs and in target designing.**

**Keywords: pseudouridine; RNA modification; ribosomal RNA; Cbf5 protein; Archaea; *Haloferax volcanii***

## INTRODUCTION

Pseudouridine (Ψ), the post-transcriptional C5-ribosyl isomer of uridine (U), is the most commonly modified residue, observed in nearly all RNAs (Charette and Gray 2000; Hopper and Phizicky 2003; Grosjean 2009; Mueller and Ferré-D'Amaré 2009; Cantara et al. 2011; Yu and Meier 2014). Ψ residues are suggested to have possible roles in the stabilization and function of RNAs (Davis 1995; Charette and Gray 2000; Ofengand et al. 2001b; Decatur and Fournier 2002; Lecointe et al. 2002; Namy et al. 2005; Hamma and Ferré-D'Amaré 2006; Baudin-Baillieu et al. 2009; Karijolich and Yu 2010; Wu et al. 2011; Spenkuch et al. 2014). These residues are produced in RNAs by single-protein Ψ synthases in all three domains of life (Ofengand 2002; Mueller and Ferré-D'Amaré 2009; Spenkuch et al. 2014; Rintala-Dempsey and Kothe 2017). In addition, specific ribonucleoprotein (RNP) complexes can also produce Ψ at various sites in the RNAs of Archaea and Eukarya (Decatur and Fournier 2003; Henras et al. 2004b; Dennis and Omer 2005; Matera et al. 2007; Hamma and Ferré-D'Amaré 2010; Kiss et al. 2010; Watkins and Bohnsack 2012; Ge and Yu 2013; Yu and Meier 2014). Each complex contains a distinct box H/ACA guide RNA and four core proteins: the Ψ synthase Cbf5 (Dyskerin in human and NAP57 in rodents), Gar1, Nop10, and L7Ae (Nhp2 in Eukarya). In this RNP complex, the guide RNA determines the specificity for modification of a particular uridine in the target RNA.

Most eukaryal box H/ACA guide RNAs contain two stem–loop or hairpin regions, connected by a hinge and followed by a 3′-tail, and both stem–loops are needed for the in vivo activity and optimal in vitro activity of each of the two stem–loops of the RNP (Ganot et al. 1997b; Bortolin et al. 1999; Decatur and Fournier 2003; Li et al. 2011). The hinge and tail regions have H (consensus ANANNA) and ACA boxes, respectively, both of which are required for the accumulation of guide RNAs and conversion of their target U to Ψ

[1]These authors contributed equally to this work.

[2]**Present address:** Department of Biochemistry and Molecular Biology, College of Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA

[3]**Present address:** Pfizer, Pearl River, New York 10965, USA

**Corresponding author: rgupta@siumed.edu**

Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.073734.119.

(Balakin et al. 1996; Ganot et al. 1997b; Bortolin et al. 1999). Furthermore, binding of Cbf5 to the ACA box is essential for the formation of correct 3′ terminus of the guide RNA, probably by preventing the action of exonucleases, especially for the RNAs derived from introns. Archaeal box H/ACA RNAs generally have one to three stem–loops or H/ACA motifs (Tang et al. 2002; Muller et al. 2008) with no H box in one stem–loop H/ACA RNA. Archaeal H/ACA motif (see Fig. 1A) contains a proximal or lower (P1) and a distal or upper (P2) stem separated by an internal loop called a "pseudouridylation pocket" (Ψ pocket). The P2 stem is capped by a loop. The 5′ and 3′ sides of the P1 and P2 stems

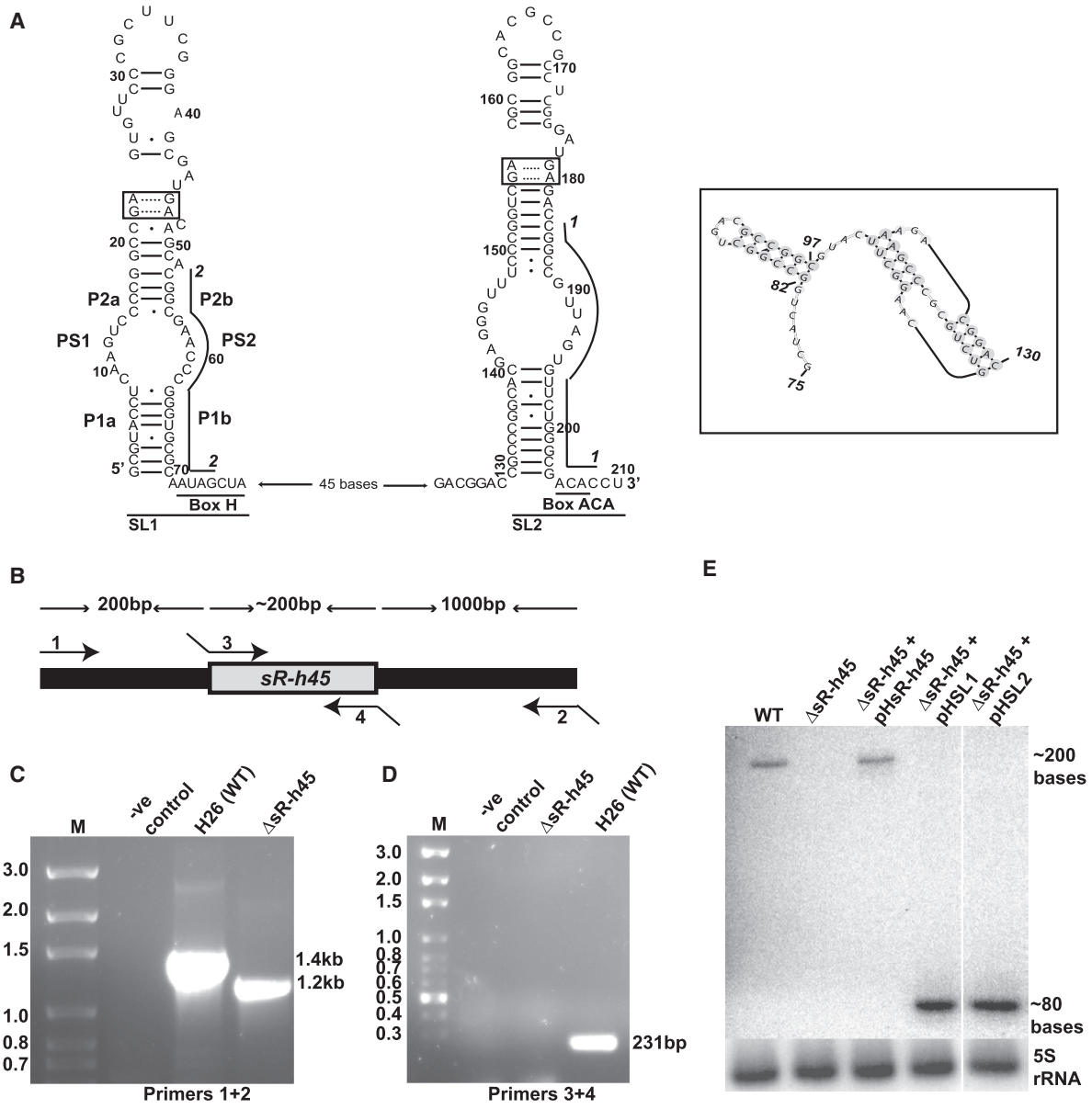


**FIGURE 1.** Confirmation of *sR-h45* deletion and its rescue. (*A*) The sequence of sR-h45 RNA. The *inset* shows the 56-base region that separates the two stem–loops (SL1 and SL2). This can potentially form a pseudoknot. Conserved H (AUAGCUA) box and ACA box are underlined. The G:A pairs of K-turns are boxed. The positions for probe 1 (HV1940/42HA-R) and probe 2 (HV2605HAR) are indicated by lines alongside SL2 and SL1, respectively. The 5′ and 3′ sides of the lower stem, upper stem and Ψ pocket are labeled (only in SL1) as P1a and P1b, P2a and P2b, and PS1 and PS2, respectively. (*B*) Schematic of sR-h45 and its flanking regions. Positions of the PCR primers HVHAMID-F, HVHA-ER, 2605HA-F, and 1940/42HA-R are marked as 1, 2, 3, and 4, respectively. The bp shown are approximate. (*C*) The PCR product from strain Δ*sR-h45* using external primers 1 and 2 is about 200 bp shorter than that of the WT. (*D*) Internal primers 3 and 4 show a PCR product in WT, but not in the Δ*sR-h45* strain. (*E*) Northern analysis of total RNA isolated from WT and Δ*sR-h45* strains, and Δ*sR-h45* transformed with pHsR-h45, pHSL1 and pHSL2. The RNAs were separated by 6% PAGE. $^{32}$P-labeled probe 1 (see *A*) was hybridized to RNA of the pHSL2 transformant and probe 2 (see *A*) to all others. An RNA of about 200 bases was observed in the WT and pHsR-h45 transformant and an ~80-base RNA in pHSL1 and pHSL2 transformants. 5S rRNA was used as a loading control.

are referred to as P1a and P1b, and P2a and P2b, respectively. The 3′ tail of the P1 stem includes an ACA (ACA includes all ANA sequences) box. Bases on the two sides of the Ψ pocket (PS1 and PS2 on the 5′ and 3′ sides, respectively) pair with the bases of the RNA on each side of the target U that is to be modified. During the guide-target pairing in a typical pseudouridylation reaction, the target U and one base on its 3′-side ("UN") remain unpaired and positioned at the base of the P2 stem. The distance from the ACA box to the target U is conserved to 15 ± 1 residues, also called the "*n* + 15" spacing rule (Ganot et al. 1997a; Ni et al. 1997; Baker et al. 2005; Xiao et al. 2009). This spacing includes only the P1b side of the P1 stem and the section of the PS2 side of the Ψ pocket up to the last base that pairs with the target RNA. The P2 stem in Archaea has either a K-turn or its variant, a K-loop (Rozhdestvensky et al. 2003; Hamma and Ferré-D'Amaré 2004; Muller et al. 2008), both of which contain two tandem sheared G:A pairs. Generally, the distance from the distal G:A pair to the base of the P2 stem is 9–10 bases (or base pairs) (Li and Ye 2006).

Several structural and in vitro modification studies have determined the importance of specific regions of the guide RNAs and their association with the core proteins (Rozhdestvensky et al. 2003; Hamma and Ferré-D'Amaré 2004, 2010; Henras et al. 2004a; Baker et al. 2005; Charpentier et al. 2005; Hamma et al. 2005; Li and Ye 2006; Manival et al. 2006; Rashid et al. 2006; Jin et al. 2007; Wu and Feigon 2007; Ye 2007; Li 2008; Duan et al. 2009; Liang et al. 2009; Li et al. 2011; Zhou et al. 2011; Fourmann et al. 2013). The crystal structure of a complete archaeal RNP complex with a single stem–loop box H/ACA RNA shows that the P2 stem binds to a composite surface formed by L7Ae, Nop10 and the D1 subdomain of the catalytic domain of Cbf5, while the PUA domain of Cbf5 binds to the P1 stem and ACA tail region. Archaeal L7Ae specifically binds the K-turn of the P2 stem and is required for correct positioning of the target RNA and a functional RNP formation. The Cbf5 creates an "L"-shaped conformation of the P1 stem and the ACA motif and specifically recognizes the two A's of the ACA box. The binding of Cbf5 to the two stems places the Ψ pocket near the active site of the enzyme. The PS1 side of the pocket resides away from the Cbf5 surface, while the PS2 side is toward the protein, making some intermolecular contacts with it. A stack is formed by the P1 stem and the helix formed by the PS2 side of the Ψ pocket with the target RNA. This stack constrains the distance between the ACA box and the target U to ~15 bases ("*n* + 15" spacing), both of which bind to Cbf5 during Ψ formation. The thumb loop of Cbf5, along with Gar1 clamps the target RNA in the correct position and the two proteins are also required for product release after Ψ formation.

Although basic requirements of H/ACA RNA-guided pseudouridylation under typical conditions are known, the exact nature of most of these requirements is not clear. There have been conflicting reports about the need for the ACA motif for the binding of proteins as well as for Ψ formation (Baker et al. 2005; Charpentier et al. 2005; Caton et al. 2018). A recent study showed that a minimum of 8 bp between guide and substrate are required for RNA-guided pseudouridylation and the Ψ pocket can accommodate more than two unpaired residues at the base of the P2 stem (De Zoysa et al. 2018). Another recent guide-target pairing-related study showed that the initial velocity of pseudouridylation decreased upon reduction of base pairs between the guide and target (Kelly et al. 2019). We wanted to characterize the requirements for pseudouridylation under atypical conditions, which would help in identifying targets for orphan H/ACA RNAs.

Bioinformatically, two single stem–loop box H/ACA RNAs were predicted to guide Ψ formation in the 23S rRNA of the archaeon *Haloferax volcanii*: one guiding U2605 (in between helix 90 and 93 of domain V) and the other for both U1940 and U1942 (in helix 69 of domain IV) (Grosjean et al. 2008). These regions of the 23S rRNA are involved in its peptidyl transferase activity (Decatur and Fournier 2002; Ofengand 2002). Previously we showed that both stem–loops are part of the same RNA molecule (Blaby et al. 2011), located in the HVO_2651s region, the intergenic region between HVO_2651, and HVO_2652 (Straub et al. 2009). (*H. volcanii* ORFs are as annotated at http://archaea.ucsc.edu) We also showed that a *cbf5*-deleted strain of *H. volcanii* completely lacks these three Ψ modifications and contains reduced steady-state levels of this box H/ACA guide RNA (Blaby et al. 2011).

The aim of this work was to determine under what conditions the ACA box was required, whether two alternate Ψ's, such as Ψ1940 and Ψ1942, can be guided by the same box H/ACA RNA, and under which atypical conditions Ψ can be produced. Here we show that the previously identified (Blaby et al. 2011) double stem–loop box H/ACA RNA (called sR-h45 in this study and its 5′ and 3′ stem–loops are called SL1 and SL2, respectively) (Fig. 1A) is indeed responsible for guiding the formation of three Ψ's (at positions 1940, 1942, and 2605) of *H. volcanii* 23S rRNA. SL1 is the guide for Ψ2605 in the unpaired "UN" configuration of the target and SL2 guides Ψ1940 and Ψ1942 formation in the unpaired "UNUN" configuration. The two stem–loops can function independently, both in vivo and in vitro. Ψ1942 is not produced by SL2 in the absence of U1940 (or Ψ1940) and Ψ1940 is not produced if Ψ1942 is already present in the substrate. SL1 produces Ψ2605 under typical conditions, but SL2 produces the two Ψ's in an atypical manner. The ACA box, although needed for the binding of Cbf5, is not required either in vivo or in vitro for the activity of the typical archaeal single stem–loop H/ACA guide RNA. However, it is needed for Ψ formation under atypical conditions. Conversely, ANA as part of the H box of the double stem–loop-containing sR-h45 guide is needed for in vivo Ψ2605 production, even under typical conditions.

Mutating ANA in this case does not affect Ψ1940 and Ψ1942 production. In the single stem–loop RNA, the pairing of N in "UN" and the N at the 3′ end of "UNUN" with a base at the proximal end of the P2 stem (on the P2a side) does not eliminate Ψ formation. Only the most 5′ unpaired U is converted to Ψ when another U is present on its 3′ side. Increasing the "*n* + 15" spacing by three bases reduced Ψ formation both in vivo and in vitro. The absence of sheared G:A pairs in the P2 stem abolished Ψ formation in vivo and drastically reduced it in vitro. However, their absence in one stem–loop did not affect the activity of the other stem–loop in the double stem–loop-containing sR-h45. Ψ is still formed, both in vivo and in vitro, if the G:A pairs are present, but are not in a strict K-turn configuration. The absence of guide-target pairing on the PS1 side of the Ψ pocket allows some Ψ formation in vitro, but not in vivo. Previously we showed that certain *H. volcanii* Cbf5 mutants produced partial Ψ at one or more of the three sites in vivo (Majumder et al. 2016). Most equivalent *Methanocaldococcus jannaschii* Cbf5 mutants showed reduced in vitro activity with the typical (SL1) guide but hardly any activity with the atypical (SL2) guide.

## RESULTS

### Genomic region HVO_2651s encoding the double stem–loop box H/ACA RNA sR-h45 is not required for *H. volcanii* viability

The intergenic HVO_2651s region of the *H. volcanii* genome contains the gene for sR-h45 RNA. The *sR-h45*-deleted strain of *H. volcanii* (ΔsR-h45) used here lacks the residues from genomic positions 2499857–2500038 (G12 to G193 in Fig. 1A). We confirmed the *sR-h45* deletion by PCR using primers that hybridized outside and within the *sR-h45* region (Fig. 1B–D). These data indicate that *sR-h45* is dispensable in *H. volcanii*.

To rescue the effects of the *sR-h45* deletion, we independently transformed the Δ*sR-h45* strain with plasmid-borne copies of *sR-h45* (pHsR-h45, genomic position 2499846–2500055, G1 to U210 in Fig. 1A) and its two stem–loops, SL1 and SL2 (pHSL1 and pHSL2, respectively) reflecting positions G1-A78 and C131-U210, respectively, in Figure 1A. The absence of sR-h45 RNA in the Δ*sR-h45* strain and the presence of RNAs derived from the plasmid-borne genes were confirmed by Northern hybridizations by using probes specific for SL1 and SL2 (Fig. 1E).

### SL1 of sR-h45 RNA guides Ψ2605 formation and SL2 guides Ψ1940 and Ψ1942 formation in *H. volcanii* 23S rRNA, and the two stem–loops can function independently in vivo

We used both U-specific (Fig. 2B,D) and CMCT (Fig. 2C,E) reactions, as described in the Materials and Methods, to determine the role of sR-h45 in Ψ1940, 1942, and 2605 formation. Ψ's are present at all three positions in the WT strain, but unmodified U's are observed at these positions in the Δ*sR-h45* strain. These unmodified U's of the Δ*sR-h45* strain revert back to Ψ's when the strain is transformed with pHsR-h45. These results suggest that Ψ's at all three positions are guided by sR-h45 RNA and Ψ modifications at these three sites are not required for cell viability. These modifications were also absent in the *cbf5*-deleted strain of *H. volcanii* as shown by us previously (Blaby et al. 2011).

To determine which stem–loop is responsible for guiding which Ψ modification, we transformed Δ*sR-h45* with pHSL1 and pHSL2 individually. As seen in Figure 2B–E, Ψ1940 and Ψ1942 were observed in the pHSL2 transformant and Ψ2605 in the pHSL1 transformant, suggesting that SL2 is responsible for guiding Ψ1940 and Ψ1942, whereas SL1 acts as the guide for Ψ2605 formation.

The sR-h45 RNA contains a 56 base (G75-C130) linker sequence between the two stem–loops (Fig. 1A), which can potentially form a pseudoknot structure (inset in Fig. 1A). When the Δ*sR-h45* strain was transformed with pHΔLinker, a plasmid containing SL1 and SL2, but lacking the 56-base linker, all three positions showed Ψ's (Fig. 2B–E), suggesting that the linker is not required for the modifications.

### SL1 of sR-h45 RNA can guide the conversion of the U2605 equivalent of 23S rRNA into Ψ, in a short RNA target in vitro, and an adjacent U is converted to Ψ in the absence of this U2605

An in vitro pseudouridylation reaction containing SL1 guide RNA and an [α-$^{32}$P]UTP-labeled WT 40-mer target RNA containing the U2605 equivalent of 23S rRNA of *H. volcanii* (Fig. 3A), produced Ψ in nuclease P1 digests of the reaction product (Fig. 3B), suggesting that Ψ is produced in this reaction as in vivo. Nuclease P1 generates labeled modified or unmodified versions of the 5′ monophosphates of all U's in [α-$^{32}$P]UTP-labeled RNA. We did time course analyses of the reactions under different substrate RNA (S) to SL1-containing RNP (sRNP) ratios (Supplemental Fig. S1). The mole Ψ formation per mole RNA after 2 min reaction was 0.96 under single turnover conditions (S/sRNP = 0.1) and 0.8, 0.74, and 0.66 under different multiple turnover conditions (S/sRNP = 2, 5, and 10, respectively). The reaction rate of SL1-containing RNP under single turnover conditions is comparable to previous reports of single stem–loop box H/ACA RNPs (Charpentier et al. 2005; Gurha et al. 2007; Duan et al. 2009). As expected, the reaction progress curve under multiple turnover conditions showed an initial burst phase, followed by slower kinetics.

There are three consecutive U's at the 2603–2605 equivalent positions in the WT target RNA (Fig. 3A). U2603 and

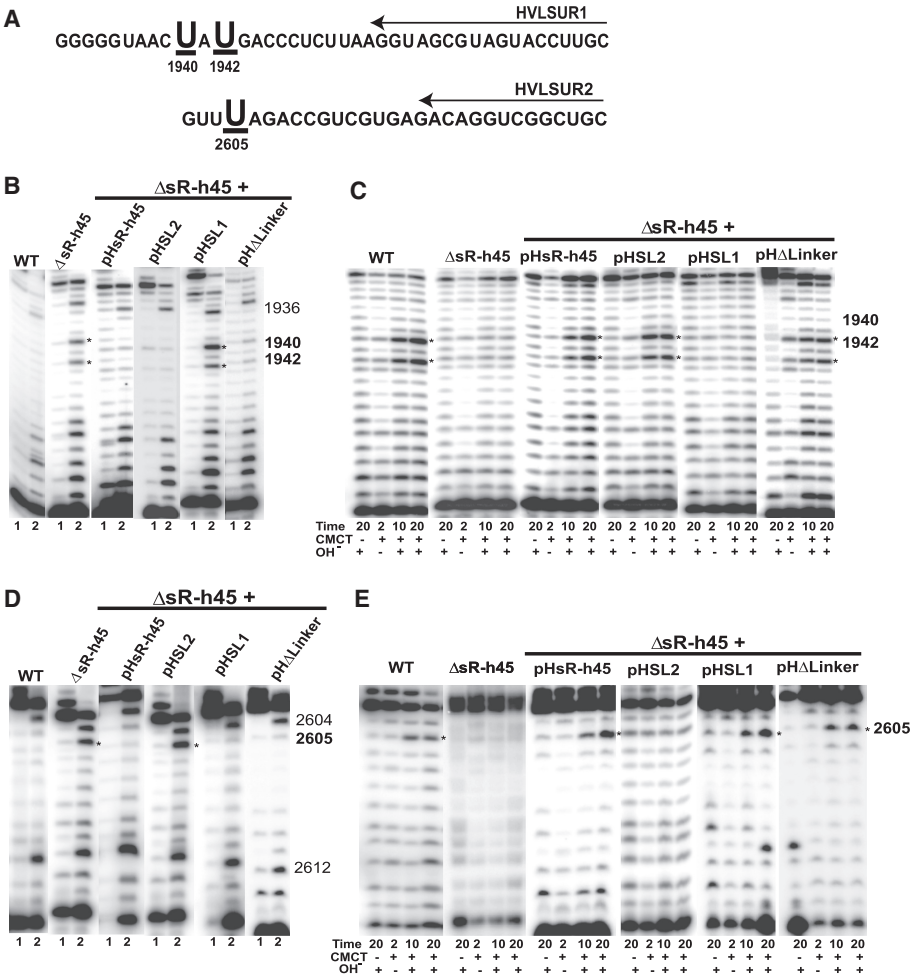


**FIGURE 2.** The sR-h45 guides the formation of Ψ1940, Ψ1942, and Ψ2605 in *H. volcanii* 23S rRNA. (*A*) Positions of the primers (arrows) HVLSUR1 and HVLSUR2 relative to the 23S rRNA sequence are shown. Residues U1940, 1942, and 2605 are underlined and are in bold large font. (*B*) U-specific analyses to determine the modification status of U1940 and U1942 were done using primer HVLSUR1 and total RNA of strain Δ*sR-h45* transformed with pHsR-h45 (full-size sR-h45) and its variants containing one stem–loop each (pHSL1 and pHSL2) and one lacking the 56 base-linker between the stem–loops (pHΔLinker). Lanes *1* and *2*: Primer extensions on untreated RNA and on RNA following U-specific reactions, respectively. The position of the primer extension band in the gel is one residue before the expected U. A dark band at a position in lane *2*, but not in lane *1* indicates an unmodified U. In WT cells, U1940 and U1942 are converted to Ψ, and U1936 (used as an indicator for the position) remains unmodified. (*C*) The primer and total RNA used in *B* were also used for CMCT-primer extension analyses. Total RNAs were either untreated (–) or treated with CMCT (+) for the indicated times (in min), followed by alkali ($OH^-$) treatment (+) or no treatment (–). Positions 1940 and 1942 are indicated on the *side*. The extension stops one residue before the CMCT modified Ψ. A dark band in the CMCT followed by alkali treatment lanes, with an increased intensity in the 20 min lanes, indicates Ψ. (*D*,*E*) Analyses similar to those in *A* and *B*, respectively, using primer HVLSUR2, were done to determine the modification status of U2605. Unmodified U2604 and U2612 served as indicators for positions in *D*. Dark bands across all lanes at position 2603 in *D* and *E* are due to strong primer-extension stops probably caused by 3-methyluridine at this position, which is observed at equivalent positions 2619 of *Haloarcula marismortui* 23S rRNA (Kirpekar et al. 2005) and 4500 of human 28S rRNA (http://people.biochem.umass.edu/fournierlab/3dmodmap/heneqlsu.php). Asterisks *next* to the bands indicate unmodified U1940 and U1942 in *B*, Ψ1940 and Ψ1942 in *C*, unmodified U2605 in *D*, and Ψ2605 in *E*.

U2604 normally pair with two A's on the 3′ side (PS2 side) of the Ψ pocket. To determine the source of the U that is converted to Ψ in the nuclease P1 digests shown in Figure 3B, we digested the same products with RNase T2. RNase T2 generates ribonucleoside-3′-monophosphates (Np). Consequently, in this nearest-neighbor analysis, the labeled 5′ phosphate, originally derived from the [α-$^{32}$P]-labeled NTP used to produce the transcript, is transferred to the 3′ side of the preceding residue. Therefore, the labeled U (or modified U) in RNase T2 digest of [α-$^{32}$P]UTP-labeled WT (U2605-containing) target would be derived from U2603 and/or U2604 (see the positions of labeled phosphates in the sequence shown below Fig. 3B). No labeled Ψ is observed in this case (WT + guide panel in Fig. 3C). This suggests that in the WT target, U2603 and U2604 are not converted to Ψ and U2605 most likely is the only source of the observed 0.8 mole Ψ/mole RNA in the nuclease P1 digest (Fig. 3B).

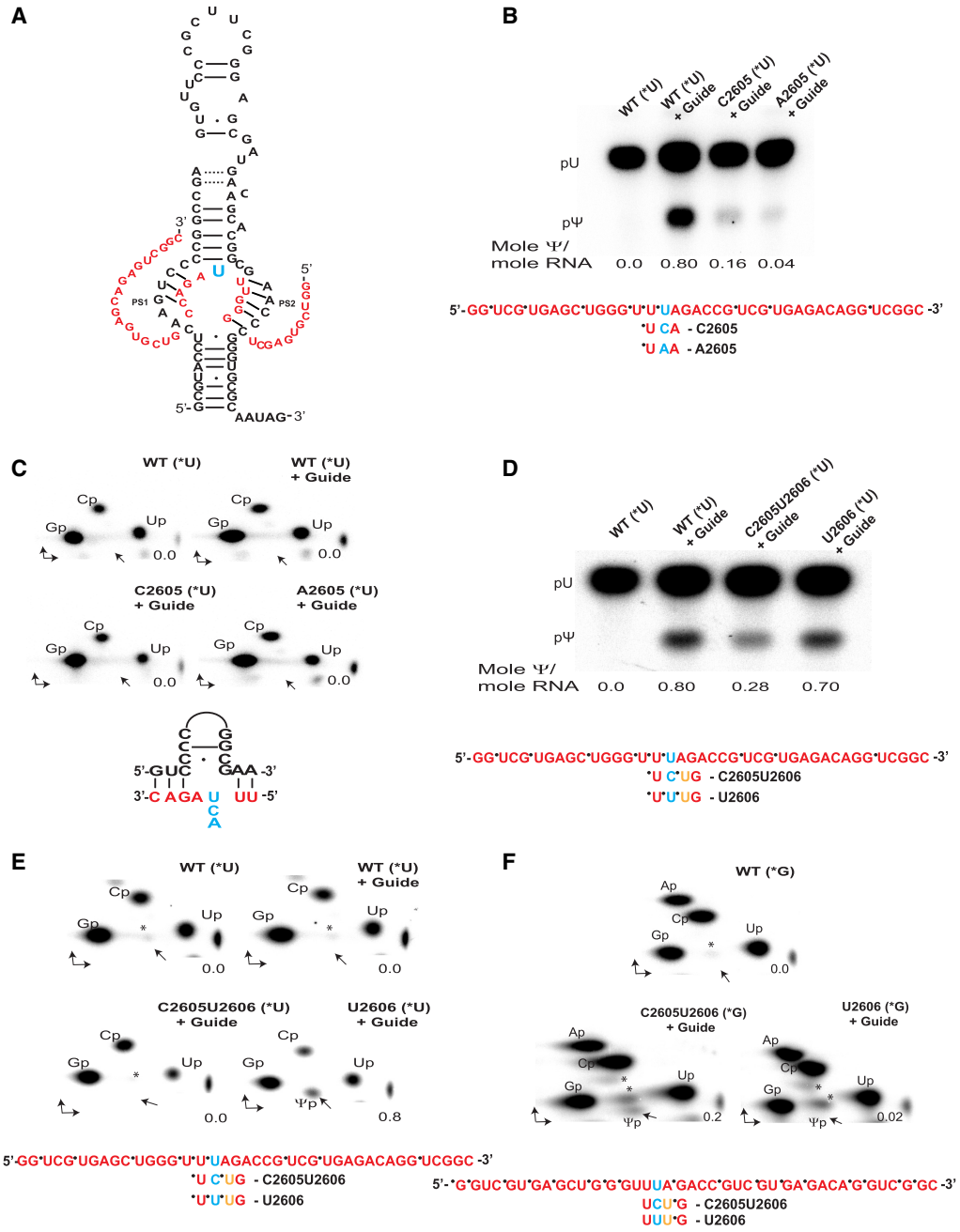


**FIGURE 3.** SL1 can guide in vitro modification to Ψ of both U2605 and the adjacent U if U2605 is absent. (*A*) Schematic of WT SL1 guide and WT 40-mer target RNAs used for in vitro assays. The target RNA (red) sequence is part of *H. volcanii* 23S rRNA sequence (positions 2589–2628) containing U2605 (blue), which is converted to Ψ2605. The PS1 and PS2 sides of the guide are indicated. (*B*,*C*) [α-$^{32}$P]UTP-labeled (indicated as *U) WT target RNA and its C2605 and A2605 mutants were separately incubated with SL1 guide RNA and four recombinant box H/ACA core proteins of *M. jannaschii*. The reactions without the guide RNA served as control as well as to rule out any protein-only (RNA-independent) Ψ formation. The products were separately digested with nuclease P1 and RNase T2, separated by 1D-TLC (in *B*) and 2D-TLC (in *C*), respectively, and phosphorimaged. Target RNAs with or without a guide are indicated *above* the lanes in *B* and *above* the panels in *C*. Radiolabeled Ψp is not observed in any panel in *C* and its expected position in each panel is indicated by an arrow. Mole Ψ/mole RNA were determined from TLC analyses and are indicated *below* each lane in *B*. The sequence of the [α-$^{32}$P]UTP-labeled WT target (and sections of its mutants) indicating the positions of the radiolabeled phosphates (dots) is shown *below B*. The pairing of WT target with upper part of Ψ pocket of SL1 is shown *below C*. (*D*,*E*) [α-$^{32}$P]UTP-labeled WT (U2605A2606-containing) 40-mer target RNA and its C2605U2606 and U2606 (U2606 in orange) mutants were individually incubated and separately digested with nuclease P1 and RNase T2, and separated by TLC, as in *B* and *C*. Mole Ψ/mole RNA are indicated *below* the lanes in *D* and *above* the panels in *E*. (*F*) Incubation and analyses were done similarly to those in *E* using [α-$^{32}$P]GTP-labeled (*G) target RNAs, instead of [α-$^{32}$P]UTP-labeled RNAs. The sequence *below* the panel indicates the position of radiolabeled phosphates (dots) in this case. For unknown reasons, extra unidentified spots (marked by [*], *below* Cp and *above* Gp) are observed in reactions containing mutant targets. Mole Ψ/mole RNA are indicated in the panels.

When U2605 of the target was mutated to C2605 or A2605, the reactions showed Ψ production in nuclease P1 digests, although much less than that for the WT target, which suggested that some U's other than the one at position 2605 were converted to Ψ in this case, although less efficiently (Fig. 3B). Labeled Ψ is also not observed in RNase T2 digests of these [α-$^{32}$P]UTP-labeled C2605 and A2605-containing substrates (lower panels in Fig. 3C), suggesting that U2603 is also not converted to Ψ in these mutants and that U2604 is the source of labeled Ψ in the nuclease P1 digests of these mutant target RNAs (Fig. 3B).

We also determined whether the guide RNA can shift the target Ψ production to the 3′ side in the absence of U2605 if a U is present at position 2606, that is, the unpaired "UN" is changed to "NU." For this, we used two mutants of the target. In one case both U2605 and A2606 were changed (C2605U2606) where unpaired "UA" is changed to "CU" and in the other case, only A2606 was mutated (U2606) where "UA" is changed to "UU." Both of these produced labeled Ψ in nuclease P1 digests (0.28 and 0.70 mole Ψ/mole RNA, respectively, Fig. 3D) when we used [α-$^{32}$P]UTP-labeled targets. RNase T2 digests of these products showed labeled Ψ only in the case of the mutant U2606 target (Fig. 3E). The amount of Ψ in this case, when compared to that of the nuclease P1 digests, suggests that nearly the entire observed Ψ is derived from U2605 whether it is associated with normal A2606 in WT or with U2606 in the mutant (second and fourth lanes in Fig. 3D). To further investigate whether Ψ in the nuclease P1 digests of C2605U2606 came from U2606 (in addition to Ψ2604, if any, as observed in the C2605 and A2605 mutants in Fig. 3B), we used these same targets, but labeled them with [α-$^{32}$P]GTP instead of [α-$^{32}$P]UTP. The results showed that labeled Ψ was mainly observed in the C2605U2606 case and only a trace in the U2606 mutant (Fig. 3F). This suggests that artificially introduced U2606 can be converted to Ψ to some extent, but only when U2605 is not present, that is, it is in a "VU" (V is A, C, or G, not U) sequence instead of "UN" sequence. We did not conduct in vivo studies using mutant targets because these would require mutations in the 23S rRNA genes.

Overall these results suggest that SL1 can produce the Ψ2605 equivalent in small RNA targets in vitro. Furthermore, U2605 is the normal site of the modification even in the presence of adjacent U's. However, in the absence of U2605, adjacent U's at positions 2604 and 2606 can partially be converted to Ψ.

## SL2 of sR-h45 RNA can guide the formation of both the Ψ1940 and Ψ1942 equivalents of 23S rRNA, because of the unpaired "UNUN" configuration during guide-target pairing

An in vitro reaction containing SL2 guide RNA and an [α-$^{32}$P]UTP-labeled 29-mer target RNA containing the U1940 and U1942 equivalents of 23S rRNA of *H. volcanii* (Fig. 4A) produced Ψ in nuclease P1 digests of the reaction product (Fig. 4B). To determine the source of U for the Ψ formation we used [α-$^{32}$P]ATP- and [α-$^{32}$P]GTP-labeled targets in separate reactions, and digested the products with RNase T2. A labeled Ψ was observed in 2D TLC in each case (Fig. 4C). It is derived from U1940 of [α-$^{32}$P]ATP-labeled target and from U1942 of [α-$^{32}$P]GTP-labeled target, suggesting that SL2 can produce both Ψ1940 and Ψ1942. The labeled U's produced in RNase T2 digests of [α-$^{32}$P]ATP-labeled substrate are derived from four different positions. However, only U1940 out of these four U's is converted to a Ψ, because as shown below, nuclease P1 digests of C1940 and G1940 mutant targets did not show any Ψ (see Fig. 4F). Labeled U in RNase T2 digests of [α-$^{32}$P]GTP-labeled substrate is derived only from position 1942.

We hypothesized that the modification of alternate U1940 and U1942 is due to an unusual interaction/pairing of the target with the Ψ pocket of the SL2 guide. Here, two unpaired adjacent "UN" motifs, that is, "UNUN" (UAUG at position 1940–1943, see Fig. 4A) are accommodated in the Ψ pocket of the guide RNA. The 3′ G of this UAUG can potentially form a G·U pair with the Ψ pocket but not a Watson–Crick pair (Fig. 4A). In a typical case, only one unpaired "UN" locates there. We believed that both U1940 and U1942 could independently be converted to Ψ when placed in unpaired "UN" configuration. To show this we prepared three derivatives of SL2 guide RNA that could be considered as "perfect" guides for independently modifying U1940 and U1942. These were the 1940 guide, 1942 guide, and a modified 1942s guide where two residues at the 3′ end of the PS2 side of the Ψ pocket of the 1942 guide are deleted (see schematics of the Ψ pockets in Figs. 4D, 6B). By using [α-$^{32}$P]ATP- and [α-$^{32}$P]GTP-labeled WT target RNA in separate reactions with these three guides, we could show that the 1940 guide produced only Ψ1940, and both 1942 and 1942s guides produced only Ψ1942 in the WT target RNA (Fig. 4E), confirming our hypothesis.

We did not get consistent in vivo data using the 1940 guide and 1942 guide (Supplemental Fig. S2). Therefore, we did not conduct in vivo experiments using mutants of these guide RNAs.

## Ψ1942 modification requires the presence of U at position 1940 and cannot be followed by conversion of U1940 to a Ψ

To determine the relationship between Ψ1940 and Ψ1942 formation, we produced C1940 and C1942 transition mutants, and G1940 and G1942 transversion mutants of the 29-mer target RNA by individually mutating U1940 and U1942. Nuclease P1 digests of the reaction containing [α-$^{32}$P]UTP-labeled C1942 and G1942 produced Ψ, but

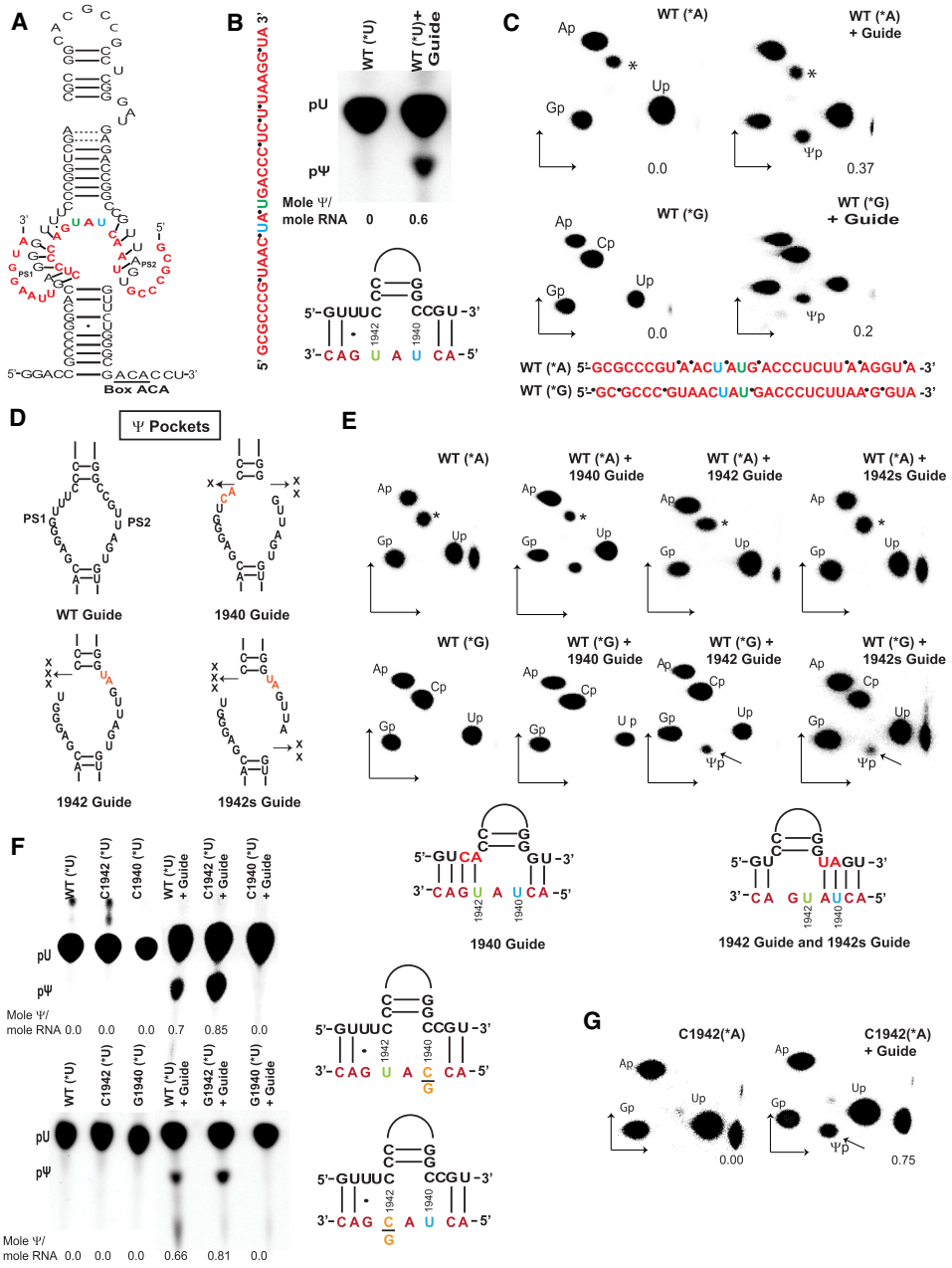


**FIGURE 4.** SL2 guides Ψ1940 and Ψ1942 equivalent modifications in vitro. (*A*) Schematic of WT SL2 guide and 29-mer WT target RNAs used for in vitro assays. The target RNA (red) sequence corresponds to positions 1929–1957 of *H. volcanii* 23S rRNA sequence containing U1940 (blue) and U1942 (light green), which are converted to Ψ1940 and Ψ1942. (Original GGG at positions 1932–1934 is changed to CCC in the target to avoid pairing with the CCC at positions 1945–1947.) The PS1 and PS2 sides of the guide are indicated. (*B*) [α-$^{32}$P]UTP-labeled (*U) WT target RNA was incubated with SL2 guide RNA and four recombinant box H/ACA core proteins of *M. jannaschii*. The product was digested with nuclease P1 and analyzed as in Figure 3B. The sequence of the labeled target indicating the positions of the radiolabeled phosphates (dots) is shown on the *side*. A section of pairing of target with Ψ pocket of SL2 is shown *below* the panel. Mole Ψ/mole RNA are indicated *below* the lanes. (*C*) [α-$^{32}$P]ATP- and [α-$^{32}$P]GTP-labeled WT target RNAs (*A and *G, respectively) were separately incubated with SL2 guide RNA and core proteins. The products were digested with RNase T2, resolved by 2D-TLC and phosphorimaged. The presence of Ψp in *A and *G RNAs indicates their origins from U1940 and U1942, respectively. (An unidentified labeled spot of variable intensity, indicated by an * in the panels of *A target was observed with the transcripts made with certain batches of [α-$^{32}$P]ATP. These are not Cp since labeled Cp cannot be produced with these transcripts.) Mole Ψ/mole RNA are shown in the panels. The sequence of the *A and *G targets with the positions of the radiolabeled phosphates (dots) are shown *below* the panels. (*D*) Schematics of Ψ pockets of WT SL2, and mutant 1940, 1942, and 1942s guides. Altered residues are indicated in orange and the number of deleted residues by the number of x's. PS1 and PS2 sides are indicated in WT. (*E*) [α-$^{32}$P]ATP- and [α-$^{32}$P]GTP-labeled (*A and *G, respectively) WT target RNAs were separately incubated with the three mutants of SL2 guide. The products were digested with RNase T2 and separated by 2D-TLC. (Unidentified spots indicated by * are similar to the ones observed in *C*.) Pairings of mutant guides with WT target RNA are shown *below* the panels. Pairings of 1942 and 1942s guides with the target RNA are identical. (*F*) [α-$^{32}$P]UTP-labeled (*U) WT target RNA and its C1942 and C1940 (*upper* panel), and G1942 and G1940 (*lower* panel) mutants were used in separate pseudouridylation reactions with SL2 guide and their nuclease P1 digests were separated by 1D-TLC. Pairings of the mutant targets (mutations in orange) with the Ψ pocket of SL2 are shown on the *side*. (*G*) RNase T2 digests of reaction products of the [α-$^{32}$P]ATP-labeled (*A) C1942 target RNA were separated by 2D-TLC as in *C*. Arrow points to Ψp.

those of C1940 and G1940 did not show any Ψ (Fig. 4F). We confirmed that Ψ in the C1942 target-containing reaction is derived from U1940 by using [α-$^{32}$P]ATP-labeled C1942 target for the reaction and digesting the products with RNase T2 (Fig. 4G). These results suggest that Ψ1940 can be produced even in the absence of U1942, but Ψ1942 formation requires the presence of U (or Ψ) at position 1940.

To further determine the relationship between Ψ1940 and Ψ1942 formation, we conducted a set of two-step reactions as described in Materials and Methods. The amount and source of Ψ produced after the first and second step are shown in Table 1. When the 1940 guide is used first and followed by the SL2 guide, Ψ1940 was hardly increased (0.21–0.25), but the total Ψ production was doubled (0.15–0.30). This increase was due to SL2-guided Ψ1942 production in the substrate that already had Ψ1940. However, when the 1942 guide was used first, SL2 barely increased total Ψ (0.17–0.20), nearly all of which came from U1942 (0.18–0.19). The Ψ1940 production in this case is negligible (0.03), which probably came from the transcripts that escaped U1942 conversion during the first-step reaction. This suggests that U1940 is not (or is negligibly) converted to Ψ, if Ψ1942 is already present in the substrate.

## Only the first U of "UNUN" is modified when it is followed by another U

Changing A1941 to U (A1941U in Fig. 5) creates three consecutive U's at positions 1940–1942. Apparently, all of the observed Ψ in this case comes from U1940, because in the string of U's, the most 5′ unpaired U is modified, which is also the case in SL1-mediated Ψ formation (see U2606 mutant in Fig. 3). This is confirmed by the double mutant AU1941-42UA (Fig. 5). Here, although two unpaired U's are present at positions 1940 and 1941, nearly all the Ψ comes from U1940. This is again confirmed when A1941 is deleted (Δ1941 in Fig. 5), bringing U1940 and (original) U1942 together. Here also nearly all Ψ is coming from U1940. Overall these results suggest that for modification of both of the alternate U's in the "UNUN" sequence, the N between the two U's should not be a U, that is, the sequence should be "UVUN."

## The "UNU" of the target "UNUN" has to be unpaired for SL2-mediated modification of the two alternate U's

Changing A1941 of the target RNA to a G nearly eliminated any Ψ formation (A1941G in Fig. 5). Here, G1941 (the internal N of the "UNUN") can pair with a C of the Ψ pocket of the SL2. However, this is an unstable situation, because this G-C pair is followed by an unpaired U1942 and a G·U pair. On the other hand, both Ψ1940 and Ψ1942 were produced with the double mutant A1941C/G1943A, where both N's of "UNUN" were changed (UAUG to UCUA) such that the 3′ terminal N can pair with the Ψ pocket, leaving only "UNU" (UCU) unpaired (Fig. 5). The corresponding single mutant A1941C, where only A1941 was changed to a C (UAUG to UCUG), also produced both Ψ1940 and Ψ1942 (data not shown), as was the case for A1941C/G1943A, the double mutant. These results suggest that for modification of both U's in the "UNUN," the first three residues (UNU) have to be unpaired and the fourth residue, the 3′ terminal N, may be paired.

## The ACA box is needed for Ψ formation if guide-target pairing differs from the typical unpaired "UN" in the target RNA and the distance between the box and the target U is more than the typical 15 ± 1 bases

The sR-h45 RNA contains the unpaired trinucleotides AUA (part of H box) and ACA (ACA box) at the 3′ end of SL1 and SL2, respectively (see Fig.1A). Cbf5 does bind to SL1 and SL2 guide RNAs and their Ψ pocket variant RNAs

**TABLE 1.** Results of two-step reactions using 29-mer target of SL2 guide

| Radiolabel and enzyme | | Guide used for the first step | Ψ after first step[a] | Ψ after second step | Control (Ψ after one step) |
|---|---|---|---|---|---|
| *U | P1 | 1940 guide | Total (1940) Ψ—0.15 | Total Ψ—0.30 | Total Ψ—0.22 |
| *A | T2 | 1940 guide | Ψ1940—0.21 | Total Ψ—0.25 | Ψ 1940—0.15 |
| *U | P1 | 1942 guide | Total (1942) Ψ—0.17 | Total Ψ—0.20 | Total Ψ—0.28 |
| *G | T2 | 1942 guide | Ψ1942—0.18 | Ψ1942—0.19 | Ψ 1942—0.25 |
| *A | T2 | 1942 guide | Ψ1940—0.00 | Ψ1940—0.03 | Ψ 1940—0.21 |

[a]Total Ψ in nuclease P1 digests after first step is derived only from one U, either U1940 or U1942.
*U, *A, and *G: [α-$^{32}$P]UTP, [α-$^{32}$P]ATP, and [α-$^{32}$P]GTP-labeled target RNA.
P1 and T2, Nuclease P1 and RNase T2.
SL2 guide was used in controls and second step of all reactions.
Numbers in the columns indicate mole Ψ/mole RNA of total Ψ, Ψ1940, or Ψ 1942. Source of Ψ is determined as in Figure 4.

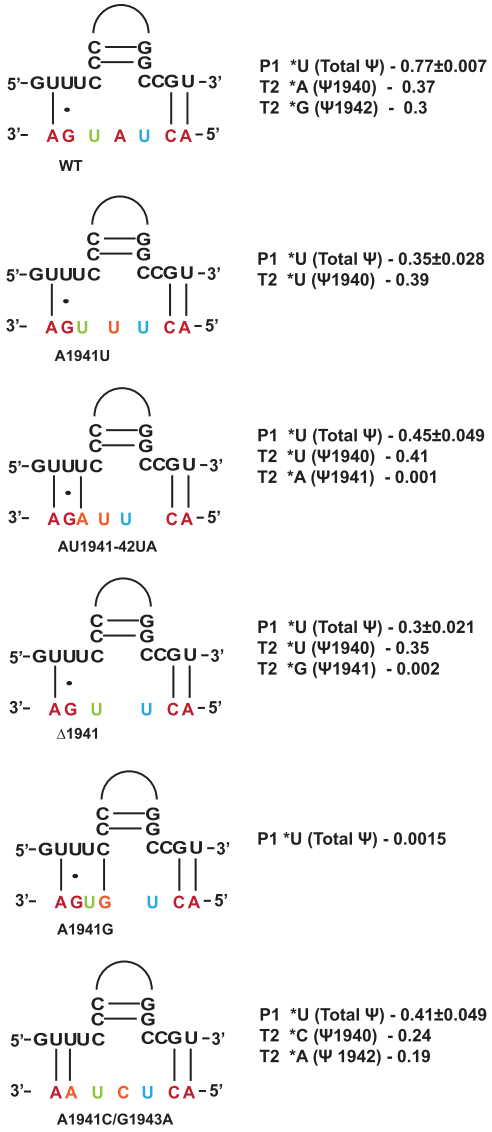


**FIGURE 5.** Deleting or changing A1941 of target "UAUG" affects SL2-mediated Ψ formation. Sections of the pairing of different mutant targets (red) with the Ψ pocket of SL2 are shown here. Original U1940 and U1942 are in blue and light green, respectively, and mutant U replacing A1941 in orange. Total Ψ, Ψ1940, or Ψ1942 (in parentheses) produced by nuclease P1 or RNase T2 digests of [α-$^{32}$P]UTP-, [α-$^{32}$P]ATP-, and [α-$^{32}$P]GTP-labeled (*U, *A, and *G, respectively) substrate are shown on the *side* as mole Ψ/mole RNA in each case. Reactions for the nuclease P1 digestions were done twice, except the one for the A1941G mutant, which showed negligible Ψ production. Reactions for the RNase T2 digestions were done only once, because these determine whether the total Ψ produced was coming from one position (Ψ at one position equal to total Ψ) or two positions.

(Supplemental Fig. S3). However, when the ACA box sequence of these guide RNAs is changed to UGU (schematics in Fig. 6B), Cbf5 no longer binds (Supplemental Fig. S3).

We individually transformed Δ*sR-h45* with UGU variants of pHSL1 and pHSL2 (mSL1 and mSL2, respectively) and determined Ψ formation in vivo. Ψ2605 was produced with mSL1, but U1940 and U1942 remained unmodified with mSL2 (Fig. 6A), suggesting that the ACA box is not required for the activity of SL1 as a typical single stem–loop guide RNA, but it is required for SL2, which is atypical. The same results were obtained from in vitro experiments (mSL1 and mSL2 in Fig. 6C). A time-course analysis using UGU-containing SL1 (mSL1) showed an initial delay, but eventually the amount of Ψ produced was the same as with the WT SL1 (Supplemental Fig. S4A). Furthermore, Ψ2605 was still produced when the AUA of SL1 was deleted (ΔAUA SL1 in Fig. 6B,C), suggesting that the ACA box is not required for the activity of WT SL1 functioning as a single stem–loop guide. We believe that the ACA box is required for SL2 activity because, as mentioned above, instead of the typical single unpaired "UN," two adjacent unpaired "UN" (Figs. 4A, 6B) are accommodated in the Ψ pocket of this guide RNA.

We prepared a variant of our SL1 guide (14 + U SL1 in Fig. 6B), where C15 at the 3′ end of the PS2 side of SL1 was changed to a U. This can potentially pair with the A on the 3′ side of target U2605. This guide RNA did produce Ψ2605 (14 + U SL1 in Fig. 6C). (This result is somewhat similar to A1941C/G1943A in Figure 5, where both Ψ1940 and Ψ1942 are produced even though the 3′ A of "UCUA" is paired.) However, when AUA of this atypical guide was changed to UGU (14 + U mSL1), Ψ2605 production was negligible (Fig. 6C). These results suggest that the ACA box is required when the N of the typically unpaired "UN" forms a Watson–Crick pair with the PS1 side of the Ψ pocket.

We also changed ACA to UGU in the above-mentioned 1940 and 1942 guides (m1940 and m1942 guides in Fig. 6B), the "perfect" guides for U1940 and U1942 modifications. The m1940 guide produced Ψ in comparable amounts to the original 1940 guide (Fig. 6C). In contrast, the UGU version of the 1942 guide (m1942 guide) produced hardly any Ψ (Fig. 6C).

The distance between AUA of SL1 and its U2605 target during guide-target pairing is typically 15 bases (Fig. 6B), which allows Ψ formation even in the absence of ACA. However, the distances between the ACA of the SL2 guide and its U1940 and 1942 targets are 17 and 19 bases, respectively, which are retained in its 1940 and 1942 guide variants. We deleted two bases at the 3′ end of the PS2 side of the pocket of the UGU containing 1942 guide (m1942s in Fig. 6B) reducing the distance from 19 to 17. This guide shows Ψ formation comparable in amounts to that of the original 1942 guide. Based on our results with the variants of the 1940 and 1942 guides, it seems that ACA is not required for modification when the distance from the 3′ base of the proximal end of the PS1 stem, that is, position of the ACA box, to the target U is up to 17 bases. However, ACA is required when this distance increased to 19 bases.

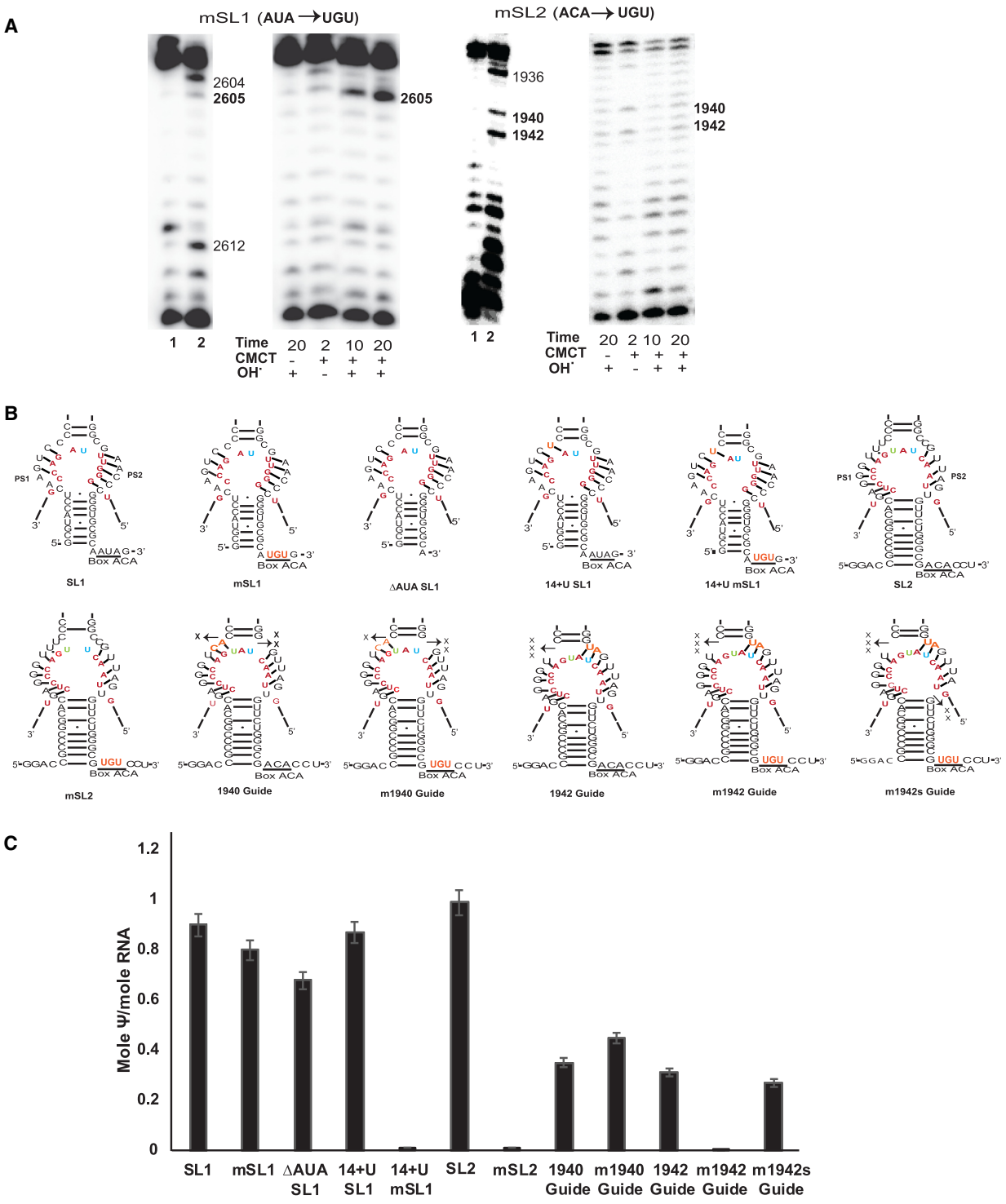


**FIGURE 6.** The ACA box is needed for Ψ synthesis when the target lacks the typical unpaired "UN" in target RNA or spacing is more than the standard "$n+15$." (*A*) The Δ*sR-h45* strain of *H. volcanii* was independently transformed with the mSL1 and mSL2 mutants of pHSL1 and pHSL2, respectively. ACA box sequence of the two stem–loops was changed to UGU in these mutants (see the schematics of mSL1 and mSL2 in *B* below). Total RNA of the transformants was treated as in Figure 2B–E to determine the modification status of U's at positions 1940, 1942, and 2605. Mutants are indicated *above* each half of the figure. The *left* panels in each half are for U-specific reaction and the *right* panels for CMCT-primer extension. (*B*) Schematics of the lower parts of different guide RNAs showing their pairings with the 40-mer or 29-mer WT target (red), as appropriate. Some of the guides are the same as in Figure 4F. Changes (orange) and deletions are indicated as in that figure. U2605 and U1940 of target RNAs are in blue and U1942 is in light green. The PS1 and PS2 sides are indicated in SL1 and SL2. (*C*) Pseudouridylation reactions using [α-$^{32}$P] UTP-labeled WT target RNAs and the guides as shown in *B* were followed by nuclease P1 digestion and analyzed by TLC. Total mole Ψ/mole RNA production in each case was determined from TLC analyses. Values are mean ± SE.

### Changes to the guide RNA that alter the interaction of target RNA and core proteins affect pseudouridylation

Binding of Cbf5 to the ACA box and positioning of its active site constrains the distance between ACA and the target U to ~15 bases. Therefore, we determined the effect of increasing this distance in SL1 from 15 to 18 by adding three bases on the PS2 side of the Ψ pocket (62aAUA in row I, Fig. 7A) or three bp in the P1 (lower) stem (7aCUA-62aAUA in row I, Fig. 7A), or three bases on the PS1 side (7aCUA in row I, Fig. 7A) for comparison. The extension of the P1 stem and PS2 side reduced the Ψ2605 formation to about half of the WT, both in vivo and in vitro. Time course analyses with these mutants showed an initial delay and the amount of Ψ produced did not reach the same levels as the WT SL1, even after 120 min (Supplemental Fig. S4A). On the other hand, an extension of the PS1 side (7aCUA) did not significantly affect Ψ2605 formation either in vivo or in vitro. We did not reduce the spacing to less than 15 with an atypical guide-target pair, believing this would be incompatible with the constraints of the molecular structure of Cbf5. Our results suggest that "$n+15$" spacing, which refers only to the 3′ side of the Ψ pocket, is important for Ψ formation both in vivo and in vitro.

Since L7Ae binds to the sheared G:A pairs in the K-turn of the P2 stem, we changed adjacent GA to CC on both 5′ and 3′ sides of the P2 stem of SL1 (GA22-23CC and GA46-47CC, respectively, in row II, Fig. 7A) to check the importance of these pairs. In both cases, Ψ2605 formation was abolished in vivo and drastically reduced in vitro; the reduction was more pronounced by the change on the 3′ side than on the 5′ side. In vitro results were reflected in the time course studies as well (Supplemental Fig. S4B). Similarly, when the GA pairs of the P2 stem of SL2 were changed to CC (SL2-GA156-157CC and SL2-GA179-180CC in row I, Fig. 7B), in vivo Ψ formation was abolished in both cases. In vitro, although Ψ was nearly absent in the 3′ mutant SL2 (GA179-180CC), small amounts of both Ψ1940 and Ψ1942 were produced in the 5′ mutant of SL2 (GA156-157CC). An unpaired U is usually present on the 3′ end of the three-base loop of the K-turn. Changing this U to C did not affect Ψ2605 formation either in vivo or in vitro (U45C in row II, Fig. 7A). Overall these data suggest that the two sheared G:A pairs in the upper stem of the guide RNA are important for Ψ formation.

The number of bases between the proximal base pair of the P2 (upper) stem of SL1 and its distal G:A pair are eight on the 5′ side and ten on the 3′ side, resulting in two single base bulges on the 3′ side (see P2a and P2b in the WT SL1 in row II, Fig. 7A). We mutated the P2 stem of SL1 to produce eight Watson–Crick pairs proximal to the two G:A pairs (mP2 in row III, Fig. 7A). This mutant showed reduced activity both in vivo and in vitro. Surprisingly, in its time-course analysis, it showed an initial rate of reaction similar to the WT, but the maximum Ψ produced did not reach the WT levels (Supplemental Fig. S4B). A comparable mutant of SL2 (SL2-mP2 in row I, Fig. 7B) also showed partial in vivo activity, but had a more drastic reduction in vitro compared to SL1. Both Ψ1940 and Ψ1942 were produced in vitro by this SL2-mP2 mutant. When the three-base loop of the K-turn in SL1 was paired by adding three bases on the 5′ (P2a) side of the P2 stem, not much difference was seen from the WT Ψ production, both in vivo and in vitro (23aAUC in row III, Fig. 7A). However, the initial rate of reaction was reduced (Supplemental Fig. S4B). In this 23aAUC mutant, there is an unpaired base on the 5′ side and two unpaired bases on the 3′ side, next to the two G:A pairs. We removed unpaired bases on both sides of these G:A pairs by combining the mP2 and 23aAUC mutants to produce the mP2-23aAUC mutant (row III, Fig. 7A). Here, there is a continuous stem where Watson–Crick pairs are present on both sides of the G:A pairs. Ψ formation was abolished in this case in vivo and it was reduced by half in vitro. The in vitro results were also reflected in the time-course studies (Supplemental Fig. S4B). We also created a mutant where the K-turn was replaced by a K-loop (K-loop in row III, Fig. 7A). In vivo activity was lost but the in vitro activity was comparable to the WT. Time course analysis showed just a reduced initial rate of reaction. A similar mutant of SL2 (SL2-K-loop in row II, Fig. 7B) also did not produce Ψ in vivo but showed some reduction in the in vitro production of both Ψ1940 and Ψ1942. Overall results of the mutations in the P2 stem suggest that for in vivo Ψ production a K-turn at an appropriate distance from the proximal end of the P2 stem is needed. A paired stem instead of the three-base loop of the K-turn is tolerated, if there are some unpaired bases on the other side of the G:A pairs, but a K-loop is not tolerated. However, Ψ formation is not completely abolished under in vitro conditions with similar changes in the P2 stem. These differences between the in vivo and in vitro Ψ production with a K-loop might be due either to differences in the reaction conditions or that *H. volcanii* proteins have slightly different requirements than those of *M. jannaschii*.

We disrupted a few guide-target pairings by mutating certain residues of the Ψ pockets. Elimination of this pairing on either the 5′ or 3′ side of the SL1 pocket (M5Gi for the PS1 side and M3Gi for the PS2 side, row IV, Fig. 7A) eliminated Ψ formation in vivo. In vitro, it was reduced to about half with M5Gi, but was nearly absent with M3Gi. Time course study with M5Gi reflected the reduction in Ψ formation (Supplemental Fig. S4C). Similar results were obtained with the comparable mutants of SL2 (SL2-M5Gi and SL2-M3Gi in row II, Fig. 7B), and both Ψ1940 and Ψ1942 were produced at a reduced level with SL2-M5Gi. Another mutation of the SL1 guide (M3Gi(2) in row IV, Fig. 7A) that eliminated the pairing of UU at the 5′ side (at positions 2603 and 2604) of the target U2605 also eliminated in vivo Ψ2605 formation. Although the unpaired

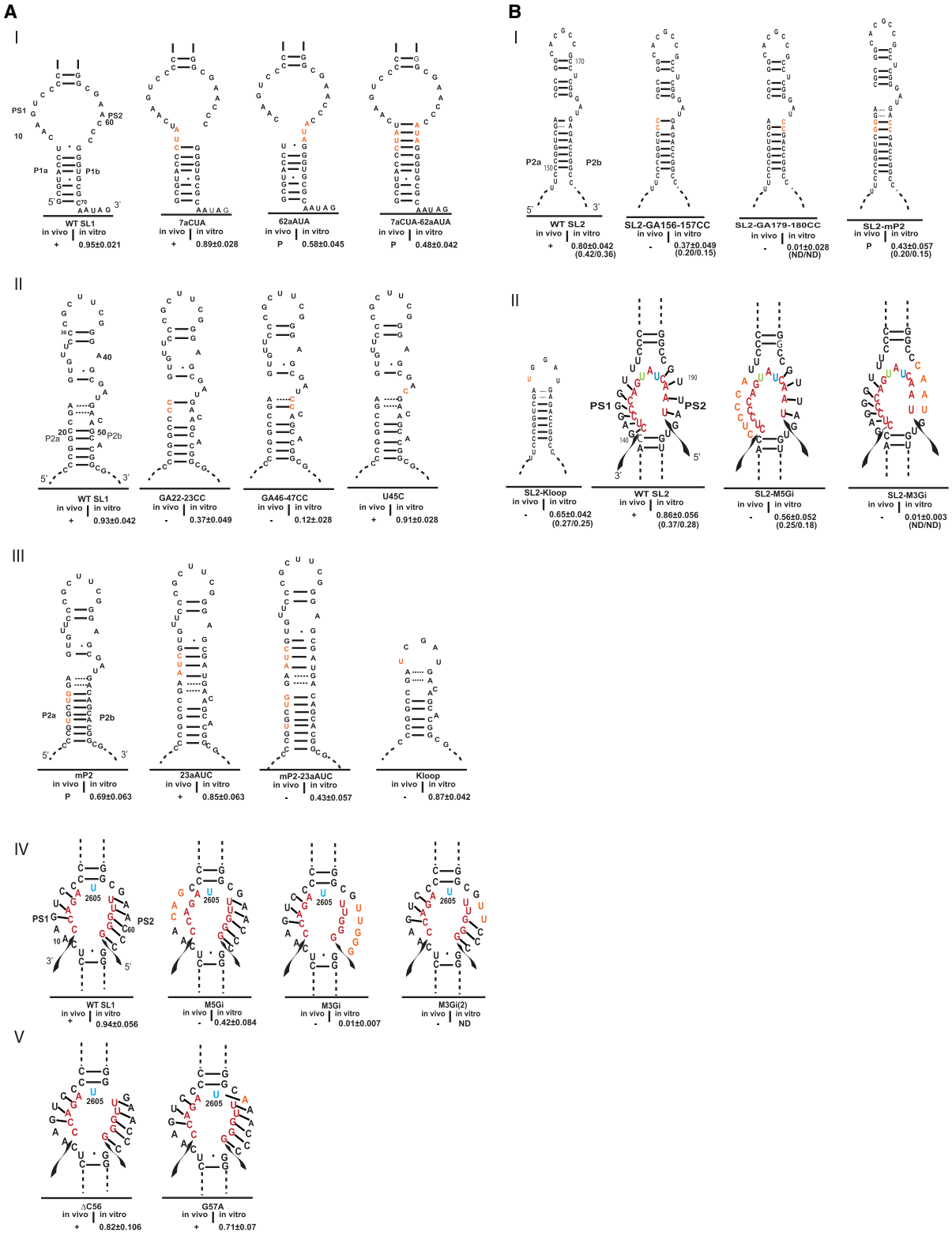


**FIGURE 7.** Several changes in guide RNAs affect Ψ formation. Schematics of the P1 and P2 stems, and Ψ pockets of SL1 (*A*) and SL2 (*B*) guide RNAs showing mutations. Most changes are shown in orange. Guide-target pairings are shown only in Ψ pocket mutants. U1940 and U2605 of target RNAs are in blue and U1942 is in light green. In vivo Ψ formation is indicated *below* the schematics as present (+), absent (–) and partial (P). For in vivo analyses, the Δ*sR-h45* strain of *H. volcanii* was independently transformed with corresponding SL1 and SL2 mutants of pHSL1 and pHSL2, respectively. Total RNA of the transformants was treated as in Figure 2B–E to determine the modification status of U's at positions 1940, 1942, and 2605. Representative results of primer extensions are shown in Supplemental Figure S5. Total mole Ψ/mole RNA production during in vitro reactions using [α-$^{32}$P]UTP-labeled WT target RNAs are also shown *below* the schematics. Values are mean ± SE of two independent reactions. (ND) Not determined. Values in parentheses in *B* indicate the mole Ψ/mole RNA produced at the two sites (Ψ1940/Ψ1942) using [α-$^{32}$P]ATP- and [α-$^{32}$P]GTP-labeled WT target RNAs, as in Figure 4C. Row numbers are indicated as I to V in *A*, and I and II in *B*.

target UUUA in this M3Gi(2) mutant of SL1 is very similar to the UUUG of the A1941U mutant of SL2 (Fig. 5), four bases of the upper part (5′ end) of the PS2 side of the Ψ pocket are unpaired in this case, whereas there are only two such bases in the case of A1941U. Also, this M3Gi(2) guide forms 3 bp on both the PS1 and PS2 sides of the pocket, instead of the normal 3 and 5 bp, respectively. Deletion of a C at the 5′ end of the PS2 side of the pocket (ΔC56 in row V, Fig. 7A) potentially created an additional C-G pair at the proximal end of the P2 stem (pair not indicated in the figure). This mutation did not affect Ψ2605 formation either in vivo or in vitro. In another variant, a G near the 5′ end of the PS2 side of the pocket was changed to an A (G57A in row V, Fig. 7A) where target U2605 could potentially form an A-U pair with the guide. This mutant produced Ψ2605 in vivo and there was only a slight reduction in the in vitro reaction, which was also observed in our time-course study (Supplemental Fig. S4C). Overall these results suggest that both sides of the Ψ pocket need to pair with the target RNA for in vivo Ψ production and only three pairs on each side of the pocket are not sufficient for this. Furthermore, although target U is unpaired under typical guide-target interaction conditions, it can be modified even if it has the potential to form a base pair with the guide RNA.

## Certain mutants of Cbf5 affect modification more under atypical guide-target interactions than under typical ones

In a previous in vivo study, we observed that certain mutations of *H. volcanii* Cbf5 either eliminated or reduced the modification of one or more of the three U's, U1940, U1942, and U2605 (Majumder et al. 2016). We reasoned that partial conversion of U to Ψ in vivo indicated a slower rate of Ψ formation. *H. volcanii* Cbf5 mutants from that study that showed partial modification or that eliminated the modification at one but not all sites, are listed in Table 2 along with their activities. Figure 8 shows the modeled structures of *H. volcanii* and *M. jannaschii* Cbf5 showing positions of most of the residues listed in Table 2. Here we prepared recombinant *M. jannaschii* Cbf5 proteins with mutations corresponding to those of *H. volcanii* Cbf5 in Table 2. The amount of in vitro Ψ formation by these Cbf5 mutants along with WT recombinant *M. jannaschii* L7Ae, Gar1 and Nop10 proteins, and different guide RNAs and corresponding targets are also shown in Table 2. Two Ala-substituted mutants, L80A and P82A did not show any activity with either SL1 or SL2 guides. The L80 and P82 residues are on the two sides of the catalytic D81, and are conserved in nearly all Cbf5 proteins (Majumder et al. 2016). Therefore, these may be essential for catalysis, at least in vitro. Also, there might be improper folding of these recombinant proteins. The other five Cbf5 mutants did produce Ψ using the SL1 guide, but the amount of Ψ was considerably reduced, between one fourth and half of the WT Cbf5. All these five mutants produced a negligible amount of Ψ (~2% or less) with either the SL2 or 1942 guides, which are not typical guide RNAs. However, results were different with the 1940 guide and 1942s guide, which are somewhat similar to typical guide RNAs (Fig. 4F). The P140A, P139A/L141A and L149A mutants, which are in the thumb loop of the Cbf5, produced Ψ almost equal to that of WT Cbf5 when used with the 1940 guide. Mutation of Y109 (Y109A), which stacks close to the target U, produced about half as much Ψ as the WT, using the 1940 guide. However, Y178A showed negligible activity with this guide. The conformation of Y178 differs between substrate-free and substrate-bound RNP. (Roles of different residues of *M. jannaschii* Cbf5 residues mentioned here are based on the published roles of corresponding residues of

**TABLE 2.** In vivo and in vitro activities of Cbf5 mutants

| HvCbf5 mutants | HvCbf5 in vivo results | | | MjCbf5 mutants | In vitro (Ψ 2605) | In vitro (total Ψ) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Ψ 2605 | Ψ 1940 | Ψ 1942 | | SL1 guide | SL2 guide | 1940 guide | 1942 guide | 1942s guide |
| WT | + | + | + | WT | 0.85 ± 0.05 | 0.83 ± 0.02 | 0.35 ± 0.02 | 0.28 ± 0.045 | 0.3 ± 0.022 |
| P111A | P | P | P | P140A | 0.45 ± 0.05 | 0.02 ± 0.02 | 0.38 ± 0.056 | 0.001 ± 0.02 | 0.12 ± 0.054 |
| P110A/R112A | + | + | – | P139A/L141A | 0.45 ± 0.04 | 0.015 ± 0.024 | 0.32 ± 0.057 | 0.02 ± 0.001 | 0.21 ± 0.02 |
| L120A | + | P | + | L149A | 0.31 ± 0.001 | 0.002 ± 0.01 | 0.4 ± 0.034 | 0.01 ± 0.034 | 0.002 ± 0.024 |
| Y81A | P | P | P | Y109A | 0.32 ± 0.056 | 0.001 ± 0.002 | 0.22 ± 0.025 | 0.023 ± 0.02 | 0.001 ± 0.015 |
| Y149A | P | P | P | Y178A | 0.22 ± 0.01 | 0.023 ± 0.034 | 0.017 ± 0.04 | 0.006 ± 0.02 | 0.005 ± 0.028 |
| P54A | P | P | P | P82A | Not active | Not active | ND | ND | ND |
| L52A | P | P | – | L80A | Not active | Not active | ND | ND | ND |

Data of *H. volcanii* Cbf5 (HvCbf5) are from Majumder et al. (2016).
In vitro Ψ formation by *M. jannaschii* Cbf5 (MjCbf5) as described in Materials and Methods.
Structures of guides are shown in Figures 3 and 4.
P, Partial; ND, not determined.

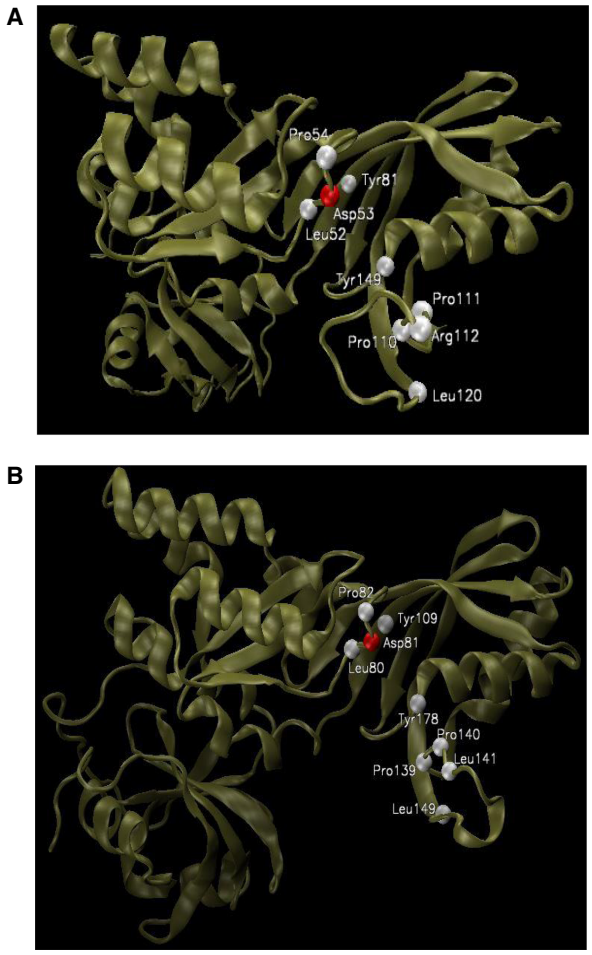


**FIGURE 8.** Homology models of Cbf5 of *H. volcanii* and *M. jannaschii*. (*A*) *H. volcanii* model is based on I-TASSER predictions of *Pyrococcus furiosus* Cbf5 structure extracted from the Cbf5-Nop10-Gar1 crystal structure (PDB 2RFK) (Liang et al. 2007a). Spheres indicate the α-carbons of catalytic Asp53 (red) and Leu52, Pro54, Tyr81, Pro110, Pro111, Arg112, Leu120, and Tyr149 (white). (*B*) *M. jannaschii* model is based on I-TASSER predictions of modified *M. jannaschii* Cbf5 structure extracted from the *M. jannaschii* Cbf5-Nop10 complex (PDB 2APO) (Hamma et al. 2005). Residues at positions corresponding to those in *A* are Asp81 (red), and Leu80, Pro82, Tyr109, Pro139, Pro140, Leu141 (instead of Arg), Leu149, and Tyr178 (white). *H. volcanii* Cbf5 figure (*A*) is based on Majumder et al. (2016).

*Pyrococcus* Cbf5.) With the 1942s guide, only P140A and P139/L141A showed some activity but at a reduced level, while the other three mutants had negligible activity. In general, the mutations that reduced the in vivo activity of *H. volcanii* Cbf5, also showed reduction or elimination of the in vitro activity of *M. jannaschii* Cbf5.

### The ANA sequence of the H box in double stem–loop sR-h45 guide RNA is needed for Ψ2605 formation, even under typical conditions

Both SL1 and SL2, the single stem–loop guides, can independently modify their target U's inside the cell in the absence of the other stem–loop of sR-h45 RNA (Fig. 2). Furthermore, the ACA box of SL1 is not needed for in vivo Ψ2605 formation under typical conditions (Fig. 6A). However, this ACA box of SL1 (AUA in this case) is normally part of the H box between the two stem–loops of sR-h45 RNA. To determine whether the ACA box in the tail region of the single stem–loop RNA and ANA sequence in the H box of the double stem–loop RNA function similarly, we changed AUA of the H box of sR-h45 to UGU (h45-mSL1) and determined its activity in vivo (Fig. 9A). Surprisingly, this construct nearly abolished Ψ2605 formation. U2605 mostly remained unmodified. However, Ψ1940 and Ψ1942 production was not affected. This suggests that, unlike the ACA box of single stem–loop guide, ANA sequence of the H box in double stem–loop guide is needed for Ψ formation even under typical conditions.

Changing the ACA box sequence of sR-h45 to UGU did not produce Ψ1940 and Ψ1942, but the Ψ2605 production was not affected (data not shown). This was expected because ACA to UGU change in single stem–loop SL2 also did not produce Ψ1940 and Ψ1942 (Fig. 6A).

### Disruption of sheared G:A pairs of one stem–loop of a double stem–loop sR-h45 guide does not affect the activity of the other stem–loop

We independently disrupted sheared G:A pairs of the two stem–loops of the sR-h45 RNA and determined their in vivo activity for the three Ψ formation (Fig. 9B). Disruption of G:A pairs in the 5′ stem–loop (h45-SL1-GA22-23CC) abolished Ψ2605 formation without affecting Ψ1940 and Ψ1942 formation. The reverse was the case when G:A pairs of the 3′ stem–loop (h45-SL2-GA156-157CC) were disrupted. This suggests that the role of G:A pairs in one stem–loop is independent of the G:A pairs in the other stem–loop of sR-h45 RNA.

## DISCUSSION

Extensive structural and modification studies have established certain common requirements for a typical H/ACA RNA-guided pseudouridylation. These requirements are mostly similar for Archaea and eukaryotes. The guide RNA contains an internal loop (Ψ pocket) between two stable stems and a 3′ tail that has an "ANA" sequence, both in the ACA box and H box. The Cbf5 enzyme anchors to the two adenines of ANA and the lower stem of the guide RNA. The Ψ pocket pairs with the target RNA and target U and a base to its 3′ side ("UN") remain unpaired during guide-target pairing. The binding of Cbf5 to the lower stem of the guide RNA and to the helix formed between the bases 5′ to the target U and the 3′ side of the Ψ pocket constrains the distance between the ACA box and the target U to approximately 15 bases, that is, the "$n + 15$" spacing rule (Duan et al. 2009). Archaeal H/ACA guide RNAs, in

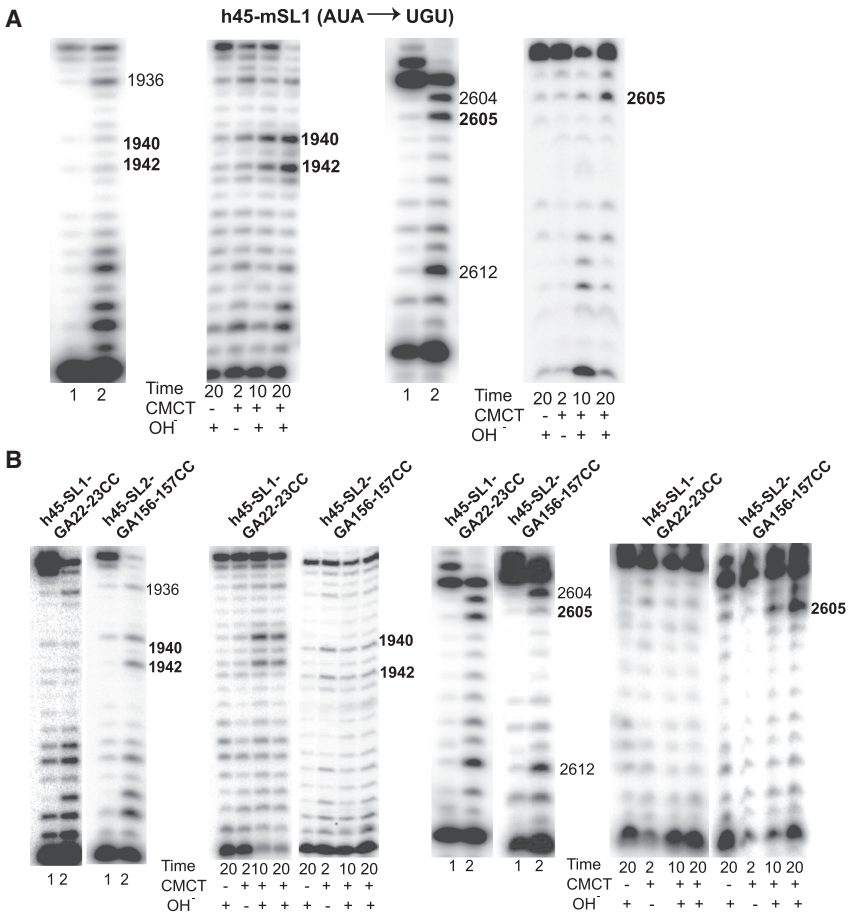


**FIGURE 9.** ANA of H box is needed even for typical pseudouridylation and disruption of sheared G:A pairs of a stem–loop in sR-h45 affects its activity, but not of the other stem–loop. The Δ*sR-h45* strain of *H. volcanii* was transformed with mutants of full-length sR-h45 RNA. Total RNA of the transformants was treated as in Figure 2B–E to determine the modification status of U's at positions 1940, 1942, and 2605. Mutants are indicated *above* the panels. Two-lane panels are for U-specific reaction and four-lane panels are for CMCT-primer extension. (*A*) h45-mSL1 mutant where AUA of the H box was changed to UGU. (*B*) Mutants h45-SL1-GA22-23CC and h45-SL2-GA156-157CC where GA at positions 22–23 in the 5′ stem–loop and GA at positions 156–157 in the 3′ stem–loop, respectively, were changed to CC.

addition contain a K-turn or K-loop in the upper stem. However, there are several instances of H/ACA RNA-guided Ψ formation, both in Archaea and eukaryotes that depart from this standard model.

The sR-h45 RNA of *H. volcanii* and corresponding H/ACA RNAs of other haloarchaea contain two stem–loops that are separated by a large spacer that can potentially form a pseudoknot structure (Fig. 1A; Blaby et al. 2011). Other archaeal H/ACA RNAs, like in eukaryotes, have a short spacer. In this study, we show that SL1, the 5′ stem–loop of sR-h45, which is a typical H/ACA guide, produces Ψ2605 of 23S rRNA in an unpaired "UN" target configuration. On the other hand, SL2, the 3′ stem–loop, produces both Ψ1940 and Ψ1942 of 23S rRNA in an atypical manner, because the two Ψ's are produced in an unpaired "UNUN" configuration. The modifications that correspond to Ψ1940 and Ψ1942 of *H. volcanii* are conserved from bacteria to humans, and in eukaryotes are also guided by H/ACA RNAs (human Ψ3741 and Ψ3743 by U19, yeast Ψ2258 and Ψ2260 by snR191). (The numbering of human and yeast residues are as in https://people.biochem.umass.edu/fournierlab/3dmodmap/eqhmarlsu.php.) However, unlike *H. volcanii*, these two modifications are guided in eukaryotes by two separate stem–loops of one guide RNA (Ganot et al. 1997a; Bortolin and Kiss 1998; Badis et al. 2003; Liang et al. 2007b). Both SL1 and SL2 of sR-h45 can efficiently function independent of each other, as well as together (Fig. 2), which is different from the eukaryotic conditions where both stem–loops of an H/ACA RNA are required for accumulation of the RNA and its function (Bortolin et al. 1999). The sR-h45 can function even in the absence of the linker (Fig. 2) suggesting no role for the large spacer, at least under laboratory conditions. We used mutants of the two stem–loops of the sR-h45 RNA of *H. volcanii* and their targets, as well as mutants of *M. jannaschii* Cbf5 protein to determine the conditions for some atypical H/ACA RNA-guided Ψ formations. We also determined the importance of the H box and sheared G:A pairs for the in vivo activity of the double stem–loop sR-h45 RNA.

## Alternate U's of the unpaired "UNUN" of the target RNA bound to the Ψ pocket can be modified sequentially in the presence of the normal thumb loop structure of the Cbf5 protein

Our results indicate that in vitro U1942 modification requires the presence of a modified or unmodified U at position 1940. However, an unmodified U1940 is not converted to Ψ1940 after conversion of U1942 to Ψ1942. Theoretically this means that after SL2-guided modification, the substrate can either have one (Ψ1940 or Ψ1942) or two (Ψ1940 and Ψ1942) modifications, and in the latter case, the modifications occurred sequentially, that is, Ψ1940 formed first, followed by Ψ1942 formation, even if the SL2 RNP is capable of modifying U1942, independent of U1940 modification. During this sequential

modification, the SL2 RNP may bind to the substrate once and release it after performing both modifications. Alternatively, the RNP can release the substrate after U1940 modification, and then the same or another RNP can bind the substrate to modify U1942.

We believe that in vivo, SL2-guided modification occurs sequentially and these two modifications occur when the substrate remains bound to RNP. Our reasons are as follows: (i) As mentioned above, Ψ1940 cannot be produced if Ψ1942 is already present in the substrate. (ii) U's corresponding to U1940 and U1942 of 23S rRNA of *H. volcanii* are constitutively modified in all three domains of life (Bortolin and Kiss 1998). If U1942 is modified first then there is a possibility of partial modification of U1940 (or less than that of U1942), which is difficult to conjecture, if both Ψ1940 and Ψ1942 are constitutive modifications. Our CMCT reactions, although not quantitative, showed equally intense bands for Ψ1940 and Ψ1942 (see Figs. 2C, 9A; Supplemental Fig. S5D). (iii) Modifications of pre-rRNA occur cotranscriptionally (Osheim et al. 2004; Koš and Tollervey 2010; Woolford and Baserga 2013), suggesting U1940 would be modified before U1942. (iv) Configuration of U1940 and U1942 during SL2-mediated modification is such that unpaired target "UAUG" is at the base of the P2 stem (top of the Ψ pocket) and bases on its two sides pair with the two sides of the Ψ pocket. This is different from the corresponding modifications by U19 snoRNA in humans and snR191 in yeast, where the two stem–loops of the guide RNA independently modify the two residues. There, both modifications occur in "UN" configurations and guide-target pairings for the two modifications are different. It is unlikely that in the sR-h45-mediated modification in *H. volcanii*, the substrate would be released after U1940 modification and then the same or a different RNP would modify U1942 by forming exactly the same guide-target pairing. (v) Mutations of certain residues in the thumb loop of Cbf5 abolish in vitro Ψ production by SL2, but only reduce production by SL1 (Table 2). This loop normally clamps the target RNA in the correct place and is involved in substrate release (Duan et al. 2009; Liang et al. 2009; Watkins and Bohnsack 2012). This loop contacts the substrate but not the guide RNA. Apparently, interactions of thumb loop with the substrate during SL2-mediated modifications of U1940 and U1942 in the "UNUN" configuration are different from SL1-mediated modification of U2605 in the "UN" configuration. The sequential nature of SL2-guided-modifications could be one of the several reasons for this difference. (vi) Our previous in vivo study (Majumder et al. 2016) showed that certain mutations of Cbf5 produce Ψ2605 and Ψ1940, but not Ψ1942 (P110A/R112A and L52A in Table 2), suggesting that in vivo these mutants can catalyze the first (Ψ1940) but not the second (Ψ1942) reaction. (P110 and R112 are in the thumb loop of Cbf5 and L52 precedes catalytic D53.) These mutants may not be able to replace U1942 in the catalytic site after production of Ψ1940, if normally the substrate is released only after modification of both U1940 and U1942. Alternatively, these mutants are not able to bind Ψ1940-containing substrate, if normally the substrate is released after Ψ1940 formation and binds again for Ψ1942 formation. In either case, the data here would suggest that U1942 is modified after U1940.

## The ACA box of the guide RNA, although not essential for typical pseudouridylation, is required for atypical reactions

Mutation of the ACA box in all of our constructs eliminated Cbf5 binding to guide RNA under our EMSA conditions (Supplemental Fig. S3), which agrees with previous reports (Baker et al. 2005; Charpentier et al. 2005) and with the established RNP structure (Li and Ye 2006). However, in a yeast system, the ACA box of the guide RNA is not needed for the complete RNP formation when all four core proteins are used together (Caton et al. 2018).

Changing the ACA sequence to UGU did not abolish Ψ2605 formation by single stem–loop SL1, either in vivo or in vitro, and even elimination of ACA produced Ψ2605 in vitro (mSL1 and ΔAUA SL1 in Fig. 6). A similar ACA to UGU change of our "perfect" 1940 guide construct, a derivative of SL2, which is like a typical guide RNA, also did not abolish Ψ1940 formation (1940 and m1940 guides in Fig. 6). It appears that the transient binding of Cbf5 to the lower stem of the guide RNA, even in the absence of the ACA box is sufficient to produce Ψ under typical conditions, because SL1 and our "perfect" 1940 guide RNA produce Ψ2605 and Ψ1940, respectively, under these conditions. This nonessential nature of the ACA box for typical pseudouridylation by a single stem–loop guide agrees with Pab91 sRNA-guided Ψ formation, one of the early reports of H/ACA RNA-guided Ψ formation (Charpentier et al. 2005).

Another early report that used Pf9 guide RNA differs in showing that the ACA box is required for the H/ACA RNA-guided Ψ formation (Baker et al. 2005). We believe the ACA box was required in that case because it is an atypical reaction. Only the target U of "UN" in this case is unpaired (Liang et al. 2007a). The 3′ N (a G) of the "UN" of the target RNA can form a pair with a base (a C) at the proximal end of the upper stem. We confirmed a similar need for the ACA box by mutating a base of the SL1 RNA to pair with the N of the "UN" of the target RNA. This mutant guide produces Ψ, but its ACA to UGU version does not (14 + U SL1 vs. 14 + U mSL1 in Fig. 6). These results suggest that the N on the 3′ side of the target U need not be unpaired as long as the ACA box is present in the guide RNA. This is observed even when the most 3′ N of the double U targets in "UNUN" conformation is paired (see A1941C/G1943 in Fig. 5).

The ACA to UGU mutant of SL2 did not produce any Ψ (mSL2 in Fig. 6), because, as mentioned above, SL2 produces Ψ in an atypical fashion. The two unpaired target U's here are in a "UNUN" configuration. Our "perfect" 1942 guide RNA, which has only one U as the target in an unpaired "UN" configuration, but has a spacing of 19 bases between ACA box and target U, can direct Ψ1942 formation. However, its ACA to UGU version does not produce Ψ (1942 and m1942 guides in Fig. 6). On the other hand, when we used a 17-base version of this ACA to UGU mutant, Ψ was produced (m1942s guide in Fig. 6).

To summarize, these results suggest that the ACA box is required when the N of the "UN" is paired, when the "*n* + 15" spacing is more than 17 bases, or when two U's of "UNUN" are the targets for modification. The ACA box was present whenever Ψ was produced in our experiments in all atypical cases. We believe that the ACA box functions as an anchor to stabilize the binding of Cbf5 to the guide RNA and this stability is essential for reactions under atypical conditions. A less stable binding of the Cbf5 to the single stem–loop guide RNA, in the absence of the ACA box can still produce Ψ under typical conditions.

### Except in the case of unpaired "UVUN," only the single most 5′ unpaired U of the target is converted to Ψ and a paired target U may become transiently unpaired for the reaction

Both U's of the normal unpaired UAUG sequence in the target RNA of SL2 were modified, even when this sequence was changed to UCUA (Fig. 5) or UCUG (data not shown), but not when it was changed to UGUG (Fig. 5), that is, the first N of the "UNUN" should be unpaired for modification. However, when UAUG was changed to UUUG, UUAG, or UUG, only one U, which was the most 5′, was modified in each case (Fig. 5). This suggested a couple of things. First, that modifications of the two alternate U's by the same guide require the four-base "UNUN" sequence to be "UVUN." Second, that in a string of unpaired U's, only the most 5′ U is modified. These results agree with a yeast in vivo study in which it was shown that the base of the upper stem (top of the Ψ pocket) of the stem–loop is flexible to accommodate four unpaired bases and that the most 5′ U in a string of unpaired U's is modified (De Zoysa et al. 2018).

The most 5′ unpaired U is modified, albeit with a lower efficiency, even when the two unpaired bases are not in "UN" format. This was observed when a normally unpaired UA sequence is changed to CU (C2605U2606 target in Fig. 3D–F), that is, "UN" changed to "NU." Here, because of the absence of the U at the normal position, a U artificially placed one base to the 3′ side of normal position, was modified.

The target U, even if it is paired, is observed to be modified both in vivo and in vitro (G57A in Fig.7A). Probably this new U-A pair at the end of the helix between the 3′ side of the Ψ pocket and target RNA allows transient unpairing ("breathing") of this U, and conversion to Ψ. Furthermore, in the absence of a U at either of the unpaired "UN" positions, a paired U on the 5′ side of the normal position can be modified. This was observed in the modification of paired U2604 when the unpaired UA sequence was changed to CA or AA (C2605 and A2605 in Fig. 3B,C), which may also be due to "breathing."

### The pairing of the target RNA with the 3′ side of the Ψ pocket is more important than the 5′ side

We observed some Ψ formation in vitro, when pairing of the target RNA is only retained with the 3′ side of the Ψ pocket, but not when it is retained only on the 5′ side (M5Gi vs. M3Gi in row IV, Fig 7A, and SL2-M5Gi vs. SL2-M3Gi in row II, Fig. 7B). The greater importance of 3′ pairing than 5′ pairing was also observed by others (Muller et al. 2008). However, the pairing of target RNA with both sides of the Ψ pocket is required in vivo, probably to provide additional stability. Complete disruption of 3′ side pairing in a yeast in vivo study also abolished Ψ formation (De Zoysa et al. 2018). The importance of pairing on the 3′ side agrees with the "*n* + 15" spacing between the ACA box and the target U, being dependent on the stacked lower stem of the H/ACA RNA and the helix formed between the 3′ side of the Ψ pocket and target RNA (Duan et al. 2009). This is not surprising because the 3′ side, but not the 5′ side of the Ψ pocket is observed to interact with Cbf5 in the crystal structure of the substrate-bound box H/ACA RNP (Duan et al. 2009). That the 3′ side of the Ψ pocket is more important than the 5′ side is also noted when their lengths are increased. An increase of three bases on the 5′ side has very little effect on Ψ formation, both in vivo and in vitro, but is reduced when the bases are added on the 3′ side (7aCUA vs. 62aAUA in row I, Fig. 7A).

### Roles of the ACA box of single stem–loop RNA guide and ANA sequence of the H box of double stem–loop guide are different even for typical pseudouridylation

In eukaryotes, presence of both stem–loops of a double stem–loop H/ACA RNA is required for either stem–loop to function in vivo (Bortolin et al. 1999). However, in *H. volcanii*, the two stem–loops of sR-h45 can function independently as well as together in vivo (Fig. 2), but the role of the AUA sequence (of the H box) is different in the two cases. Sequence of the H box of sR-h45 is AUAGCU and it is preceded by one unpaired residue. This does not differ from the consensus H box sequence ANANNA of eukaryotes, where the last A is less conserved and the box is separated from the P1 stem by one or two residues (Ganot et al.

1997b; Bortolin et al. 1999). Although, AUA of the H box of sR-h45 is not needed for typical pseudouridylation of U2605 when it functions as an ACA box of the single stem–loop (SL1) guide (Fig. 6), it is needed when it is part of the H box of the double stem–loop (sR-h45) guide (Fig. 9A). Furthermore, mutation of the H box of sR-h45 while blocking Ψ2605 formation by the 5′ stem–loop does not affect Ψ1940 and Ψ1942 formation by the 3′ stem–loop in vivo (Fig. 9A). This is similar to a yeast in vitro study, where mutation of H box affected Ψ formation in the 5′ substrate but not in the 3′ substrate (Caton et al. 2018). However, it is different from the yeast in vivo study, where mutation of the H box affects the activities of both stem–loops (Bortolin et al. 1999).

As suggested above for the role of the ACA box of the single stem–loop-guided atypical pseudouridylation, AUA of the H box also serves as an anchor to stabilize the binding of Cbf5 to the guide RNA and this stability is essential even for the typical pseudouridylation of U2605 by the double stem–loop guide. This interaction of Cbf5 with H box may be more important than that with the ACA box, because human Cbf5 (dyskerin) preferentially associates with H box (Kishore et al. 2013). It has been suggested that since the two stem–loops of eukaryotic H/ACA RNA are always located in close proximity, a direct or an adaptor protein-mediated interaction occurs between the two sets of proteins that are bound to the two stem–loops of the RNA, and this may be the reason for the presence of two stem–loops in most eukaryotic H/ACA RNAs even when they contain only one functional Ψ pocket (Bortolin et al. 1999). Probably the two RNPs assembled on the two stem–loops of sR-h45 also interact similarly and this association may be needed for the functioning of this double stem–loop guide as is the case in eukaryotes. The Cbf5, bound to the H box (and 5′ stem–loop), would then be the protein that directly or indirectly interacts with the proteins bound to the 3′ stem–loop. Cbf5 associated with the 5′ stem–loop would not bind strongly (or not bind correctly) when AUA of the H box is mutated to UGU, thus eliminating the Ψ2605 forming activity of the 5′ stem–loop even under typical conditions. The pseudoknot in the spacer of the sR-h45 may be a mechanism to bring the two stem–loops into close proximity needed for the association of the two RNPs assembled on these stem–loops. Removal of this spacer naturally brings the two stem–loops into close proximity and therefore, has no effect on the activity of the two stem–loops (Fig. 2).

Overall, our study suggests that H/ACA RNP-mediated Ψ's can be produced by certain structural variants of the guide RNA and can occur under several atypical guide-target interactions. Moreover, the thumb loop and catalytic site of Cbf5 can function even when the target U is slightly out of position or not in the correct configuration. Sometimes, Ψ is produced under these conditions only in vitro and not in vivo. Some of these conditions may be applicable to Archaea and not to eukaryotes. Several box H/ACA RNAs of archaeon *Pyrobaculum* have been discovered that lack the P1 stem, ACA box and even the PS1 side of the Ψ pocket (Bernick et al. 2012). Most likely, these guide RNAs produce Ψ efficiently under some atypical conditions as observed by us. Our study along with recent guide-target base-pairing studies (De Zoysa et al. 2018; Kelly et al. 2019) will help in determining the conditions for certain atypical H/ACA RNA-guided pseudouridylations.

In our study, several atypically produced Ψ's were reduced in an amount relative to standard conditions, suggesting that certain H/ACA RNA-guided Ψ's that are produced nonconstitutively in vivo, may be the products of atypical reactions under stress or other regulated gene-expression conditions. The possibility of unintended targets in targeted pseudouridylations for promoting nonsense suppression during translation has been suggested (Karijolich and Yu 2011), which may occur due to atypical reactions. Furthermore, several human H/ACA RNAs are called orphan, because so far, no target RNAs are known for them. It is possible that (some of) these orphan H/ACA RNAs produce Ψ by interacting with their targets in an atypical manner and searches using the rules for typical guide-target interactions cannot find these targets.

## MATERIALS AND METHODS

Standard molecular biology procedures (Green and Sambrook 2012) were used unless specifically described. Oligonucleotides used in this work are listed in Supplemental Table S1.

### Strains, media, and transformation procedures

*H. volcanii* H26 (Blaby et al. 2011) and Δ*sR-h45* (received from J. Soppa, Goethe University, Germany) were used as the WT and *sR-h45* gene-deleted strains. *H. volcanii* cells were routinely grown at 42°C–44°C in Hv-YPC medium, as described in the HaloHandbook (https://haloarchaea.com/wp-content/uploads/2018/10/Halohandbook_2009_v7.3mds.pdf) or in a medium described previously (Gupta 1984), supplemented with mevinolin (10 µg/mL) as required. Transformation of *H. volcanii* was done as described before (Blaby et al. 2011). LB (Fisher) or LB agar (Fisher) was used to routinely grow *E. coli* at 37°C, and supplemented with ampicillin (100 µg/mL), IPTG (0.2 mM), and X-Gal (40 µg/mL) as needed.

### Cloning and expression of WT and mutants of *H. volcanii sR-h45* for in vivo analysis

*E. coli* vector pKS$^+$MetPro and *E. coli-H. volcanii* shuttle vector pDS2 were used for cloning the *sR-h45* gene. Construction of these vectors is described in the Methods section of the Supplemental Material and schematically shown in Supplemental Figure S6. pKS$^+$Metpro is a derivative of pBlueScript KS$^+$ (Stratagene) with an insert (sequence in Supplemental Fig. S7) that contains the

promoter and terminator regions of the *H. volcanii* intron-containing elongator tRNA-Met gene (Datta et al. 1989), but not the structural gene. The two regions are separated by a *Sma*I site. pDS2 is derived from the shuttle vector pWL102 (Lam and Doolittle 1989) and contains the selectable markers ampicillin for *E. coli* and mevinolin for *H. volcanii*.

The *sR-h45* gene (HVO_2651s, genomic position 2499846–2500055) was PCR amplified using *H. volcanii* genomic DNA and HVHA-F2 and HVHA-R primers. The product was digested with *Pvu*II and *Dra*I and ligated into a *Sma*I-digested pKS$^+$Metpro vector. The direction of the ligated gene was checked by sequencing. The *Kpn*I-*Xba*I fragment (~430 bp) from this clone was re-cloned into pDS2 to prepare pHsR-h45 that can express WT *sR-h45* RNA. Mutants of the sR-h45 gene were created in the pKS$^+$Metpro clones by site-directed mutagenesis and then transferred to pDS2 to express mutant RNAs. The correct sequence of each mutant insert was confirmed both in pKS$^+$Metpro and pDS2 clones. pDS2 clones were used to transform the Δ*sR-h45* strain of *H. volcanii*.

## Determination of the presence or absence of Ψ at specific positions in RNA

The presence of Ψ or its absence (presence of unmodified U) at a specific position in RNA was determined by primer extension following 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate (CMCT) treatment (Ofengand et al. 2001a; Motorin et al. 2007) and U-specific sequencing (Peattie 1979; Gupta 1984), respectively, as described previously (Blaby et al. 2011). The CMCT reaction basically involves treatment of RNA with CMCT, ethanol precipitation, alkali treatment, and ethanol precipitation again, followed by primer extension. CMCT forms adducts with Ψ, U, and G. Alkali removes all CMCT groups except those attached to $N_3$ of Ψ (Bakin and Ofengand 1993). This method is not suitable for absolute quantitation of Ψ, because partial reaction conditions are used for CMCT reactions. Furthermore, sensitivity of reverse transcriptase-mediated primer extensions varies due to secondary structure and different modified residues present in RNA. The U-specific reactions are also partial. Here uridines in RNA are hydrazinolysed using aqueous hydrazine followed by strand scission at the position of damaged bases using acidified aniline. Both CMCT and U-specific reactions were done on the same batch of RNA. Reactions for the effects of each mutant RNA were done at least twice on the RNAs prepared from independent cell cultures harvested at different times. Results of repeat reactions were similar. We consider Ψ present at a position in the RNA if a dark band is observed at that position in the gel after CMCT but not after a U-specific reaction and absent when the reverse is the case. We scored partial Ψ production when dark bands were observed at a particular position after both CMCT and U-specific reactions, and we did not quantitate in vivo Ψ production. The presence of mutant RNA in every *H. volcanii* strain that did not show Ψ at the appropriate position was checked by northern blot analyses (see Supplemental Fig. S8).

## Cloning and purification of *M. jannaschii* proteins

We prepared and used recombinant *M. jannaschii* L7Ae, Cbf5, Nop10, and Gar1 proteins as described previously (Tran et al. 2003; Gurha et al. 2007). *M. jannaschii* proteins were used instead of *H. volcanii* proteins for in vitro pseudouridylation assays, because recombinant *H. volcanii* box C/D RNP proteins showed no activity in our previous 2′-*O*-methylation assays (Joardar et al. 2011). Most halophilic enzymes are inactivated when the $Na^+$ or $K^+$ concentration in the solution decreases to less than 2 M (Madern et al. 2000). Moreover, halophilic proteins are mostly negatively charged and so do not easily associate with RNA in low salt concentrations. The mutants of Cbf5 were created by site-directed mutagenesis and were confirmed by DNA sequencing. Concentrations of all recombinant proteins were determined by the Coomassie Protein Assay Reagent. Coomassie blue-stained gels of the recombinant proteins are shown in Supplemental Figure S9. *M. jannaschii* Cbf5 was also used for RNA-binding studies using gel-shift assays as described in the Methods section of Supplemental Material.

## Preparation of guide and target RNAs

PCR-amplified DNAs using appropriate plasmids as templates were used for in vitro transcription to prepare guide RNAs. Forward primers for the PCR contained T7 RNA polymerase promoter sequences. Target RNAs were produced using PCR with the T7P oligonucleotide and another primer that contained its complementary sequence toward its 3′ end. In vitro transcriptions were carried out as described before (Gurha et al. 2007; Gurha and Gupta 2008). $^{32}$P-labeled transcripts were prepared using [α-$^{32}$P]UTP, ATP, CTP, or GTP.

## Pseudouridylation by recombinant proteins and thin layer chromatography

A typical 25 µL pseudouridylation reaction contained 0.01 µM radiolabeled target RNA, and 0.1 µM of each of the four proteins and 0.1 µM of guide RNA in a buffer (20 mM Na-HEPES, pH 7.0, 150 mM NaCl, 0.75 mM DTT, 1.5 mM EDTA, 10% glycerol and 1 µg yeast tRNA). The amount of target RNA varied in multiple turnover reactions as shown in Supplemental Figure S4. After incubation at 68°C for 1 h, the reactions were stopped by adding 175 µL stop solution (0.2 mM EDTA, 0.1% SDS), followed by phenol/chloroform extraction, ethanol precipitation and digestion with Nuclease P1 or RNase T2. The digests were resolved by TLC, and radioactivity was revealed and quantitated as described before (Deogharia et al. 2019). All assays were repeated twice, except where stated. The TLCs shown in figures are representative. Results of repeat reactions were similar. To reduce variability and to get comparable results, as much as possible, the same batch of recombinant proteins were used for a particular set of experiments. Mole Ψ produced/mole RNA in nuclease P1 digests was determined by the formula: (radioactivity in pΨ spot × the number of U's in the RNA)/(sum of the radioactivity in pU and pΨ spots). Mole Ψ/mole RNA in RNase T2 digests was determined as: (radioactivity in Ψp spot × the number of N's in the RNA preceding the labeled nucleotide)/(sum of all the radiolabeled Np spots).

A set of two-step reactions was also done. In these the radiolabeled 29-mer RNA (target for the SL2 guide) was first incubated with the 1940 guide or 1942 guide and four core proteins for 1 h at 68°C. Following phenol/chloroform extraction and ethanol precipitation, a fraction of the product was digested with

appropriate enzyme and analyzed by TLC. The remaining fraction was purified by denaturing PAGE and incubated again with the SL2 guide and core proteins for 1 h at 68°C. Again, following phenol/chloroform extraction and ethanol precipitation, the product was digested with an appropriate enzyme and analyzed by TLC. Controls for these were one-step reactions using the same radiolabeled target RNA and SL2 guide RNA, but the incubation was for 2 h at 68°C. This set of experiments was done only once.

## Homology modeling

The homology models were made based on predictions using the I-Tasser protein structure suite (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), as done previously (Majumder et al. 2016). For the *H. volcanii* Cbf5 homology model, the *Pyrococcus furiosus* Cbf5 from the Cbf5-Nop10-Gar1 crystal structure (PDB 2RFK) was utilized. For the *M. jannaschii* Cbf5 model, the crystal structure on file, from the *M. jannaschii* Cbf5-Nop10 Complex (PDB 2APO), was not usable because it lacked residues 139–149, which were relevant to our work. Therefore, a homology model was created, based on the same structure but including these missing residues. The final refined structures for *H. volcanii* and *M. jannaschii* Cbf5 were saved as PDB files. These were visualized in the Visual Molecular Dynamics (VMD) suite (Humphrey et al. 1996). The VMD figures with labeled residues are represented in Figure 8.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

We thank Julia Babski from the laboratory of Joerg Soppa (Goethe University, Frankfurt, Germany) for providing the Δ*sR-h45* strain of *H. volcanii*, Srinivas Malliahgari for preparing plasmid pKS$^+$Metpro, Dan Shao for preparing plasmid pDS2, Mike Bosmeny for help in preparing Figure 8, and David Clark (Southern Illinois University) for critical reading of the manuscript. This work was supported by National Institutes of Health grant number GM055945 to R.G.

*Received October 22, 2019; accepted December 30, 2019.*

## REFERENCES

Badis G, Fromont-Racine M, Jacquier A. 2003. A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA* **9:** 771–779. doi:10.1261/rna.5240503

Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM, Terns MP. 2005. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19:** 1238–1248. doi:10.1101/gad.1309605

Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyltransferase center: analysis by the application of a new sequencing technique. *Biochemistry* **32:** 9754–9762. doi:10.1021/bi00088a030

Balakin A, Smith L, Fournier M. 1996. The RNA World of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86:** 823–834. doi:10.1016/S0092-8674(00)80156-7

Baudin-Baillieu A, Fabret C, Liang XH, Piekna-Przybylska D, Fournier MJ, Rousset JP. 2009. Nucleotide modifications in three functionally important regions of the *Saccharomyces cerevisiae* ribosome affect translation accuracy. *Nucleic Acids Res* **37:** 7665–7677. doi:10.1093/nar/gkp816

Bernick DL, Dennis PP, Hochsmann M, Lowe TM. 2012. Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. *RNA* **18:** 402–411. doi:10.1261/rna.031385.111

Blaby IK, Majumder M, Chatterjee K, Jana S, Grosjean H, de Crecy-Lagard V, Gupta R. 2011. Pseudouridine formation in archaeal RNAs: the case of *Haloferax volcanii*. *RNA* **17:** 1367–1380. doi:10.1261/rna.2712811

Bortolin ML, Kiss T. 1998. Human U19 intron-encoded snoRNA is processed from a long primary transcript that possesses little potential for protein coding. *RNA* **4:** 445–454.

Bortolin ML, Ganot P, Kiss T. 1999. Elements essential for accumulation and function of small nucleolar RNAs directing site-specific pseudouridylation of ribosomal RNAs. *EMBO J* **18:** 457–469. doi:10.1093/emboj/18.2.457

Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF. 2011. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* **39:** D195–D201. doi:10.1093/nar/gkq1028

Caton EA, Kelly EK, Kamalampeta R, Kothe U. 2018. Efficient RNA pseudouridylation by eukaryotic H/ACA ribonucleoproteins requires high affinity binding and correct positioning of guide RNA. *Nucleic Acids Res* **46:** 905–916. doi:10.1093/nar/gkx1167

Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49:** 341–351. doi:10.1080/152165400410182

Charpentier B, Muller S, Branlant C. 2005. Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33:** 3133–3144. doi:10.1093/nar/gki630

Datta PK, Hawkins LK, Gupta R. 1989. Presence of an intron in elongator methionine-tRNA of *Halobacterium volcanii*. *Can J Microbiol* **35:** 189–194. doi:10.1139/m89-029

Davis DR. 1995. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23:** 5020–5026. doi:10.1093/nar/23.24.5020

Decatur WA, Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends Biochem Sci* **27:** 344–351. doi:10.1016/S0968-0004(02)02109-6

Decatur WA, Fournier MJ. 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem* **278:** 695–698. doi:10.1074/jbc.R200023200

Dennis PP, Omer A. 2005. Small non-coding RNAs in Archaea. *Curr Opin Microbiol* **8:** 685–694. doi:10.1016/j.mib.2005.10.013

Deogharia M, Mukhopadhyay S, Joardar A, Gupta R. 2019. The human ortholog of archaeal Pus10 produces pseudouridine 54 in select tRNAs where its recognition sequence contains a modified residue. *RNA* **25:** 336–351. doi:10.1261/rna.068114.118

De Zoysa MD, Wu G, Katz R, Yu YT. 2018. Guide-substrate base-pairing requirement for box H/ACA RNA-guided RNA pseudouridylation. *RNA* **24:** 1106–1117. doi:10.1261/rna.066837.118

Duan J, Li L, Lu J, Wang W, Ye K. 2009. Structural mechanism of substrate RNA recruitment in H/ACA RNA-guided pseudouridine synthase. *Mol Cell* **34:** 427–439. doi:10.1016/j.molcel.2009.05.005

Fourmann JB, Tillault AS, Blaud M, Leclerc F, Branlant C, Charpentier B. 2013. Comparative study of two box H/ACA ribonucleoprotein pseudouridine-synthases: relation between

conformational dynamics of the guide RNA, enzyme assembly and activity. *PLoS One* **8:** e70313. doi:10.1371/journal.pone.0070313

Ganot P, Bortolin ML, Kiss T. 1997a. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89:** 799–809. doi:10.1016/S0092-8674(00)80263-9

Ganot P, Caizergues-Ferrer M, Kiss T. 1997b. The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* **11:** 941–956. doi:10.1101/gad.11.7.941

Ge J, Yu YT. 2013. RNA pseudouridylation: new insights into an old modification. *Trends Biochem Sci* **38:** 210–218. doi:10.1016/j.tibs.2013.01.002

Green MR, Sambrook J. 2012. *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Grosjean H. 2009. Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. In *DNA and RNA modifying enzymes: structure, mechanism, function and evolution* (ed. Grosjean H), pp. 1–18. Landes Bioscience, Austin, TX.

Grosjean H, Gaspin C, Marck C, Decatur WA, de Crécy-Lagard V. 2008. RNomics and Modomics in the halophilic archaea *Haloferax volcanii*: identification of RNA modification genes. *BMC Genomics* **9:** 470. doi:10.1186/1471-2164-9-470

Gupta R. 1984. *Halobacterium volcanii* tRNAs. Identification of 41 tRNAs covering all amino acids, and the sequences of 33 class I tRNAs. *J Biol Chem* **259:** 9461–9471.

Gurha P, Gupta R. 2008. Archaeal Pus10 proteins can produce both pseudouridine 54 and 55 in tRNA. *RNA* **14:** 2521–2527. doi:10.1261/rna.1276508

Gurha P, Joardar A, Chaurasia P, Gupta R. 2007. Differential roles of archaeal box H/ACA proteins in guide RNA-dependent and independent pseudouridine formation. *RNA Biol* **4:** 101–109. doi:10.4161/rna.4.2.5177

Hamma T, Ferré-D'Amaré AR. 2004. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure (Camb)* **12:** 893–903. doi:10.1016/j.str.2004.03.015

Hamma T, Ferré-D'Amaré AR. 2006. Pseudouridine synthases. *Chem Biol* **13:** 1125–1135. doi:10.1016/j.chembiol.2006.09.009

Hamma T, Ferré-D'Amaré AR. 2010. The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification. *J Biol Chem* **285:** 805–809. doi:10.1074/jbc.R109.076893

Hamma T, Reichow SL, Varani G, Ferré-D'Amaré AR. 2005. The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12:** 1101–1107. doi:10.1038/nsmb1036

Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M. 2004a. Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs. *RNA* **10:** 1704–1712. doi:10.1261/rna.7770604

Henras AK, Dez C, Henry Y. 2004b. RNA structure and function in C/D and H/ACA s(no)RNPs. *Curr Opin Struct Biol* **14:** 335–343. doi:10.1016/j.sbi.2004.05.006

Hopper AK, Phizicky EM. 2003. tRNA transfers to the limelight. *Genes Dev* **17:** 162–180. doi:10.1101/gad.1049103

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph* **14:** 33–38. doi:10.1016/0263-7855(96)00018-5

Jin H, Loria JP, Moore PB. 2007. Solution structure of an rRNA substrate bound to the pseudouridylation pocket of a box H/ACA snoRNA. *Mol Cell* **26:** 205–215. doi:10.1016/j.molcel.2007.03.014

Joardar A, Malliahgari SR, Skariah G, Gupta R. 2011. 2′-*O*-methylation of the wobble residue of elongator pre-tRNA$^{Met}$ in *Haloferax volcanii* is guided by a box C/D RNA containing unique features. *RNA Biol* **8:** 782–791. doi:10.4161/rna.8.5.16015

Karijolich J, Yu YT. 2010. Spliceosomal snRNA modifications and their function. *RNA Biol* **7:** 192–204. doi:10.4161/rna.7.2.11207

Karijolich J, Yu YT. 2011. Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474:** 395–398. doi:10.1038/nature10165

Kelly EK, Czekay DP, Kothe U. 2019. Base-pairing interactions between substrate RNA and H/ACA guide RNA modulate the kinetics of pseudouridylation, but not the affinity of substrate binding by H/ACA small nucleolar ribonucleoproteins. *RNA* **25:** 1393–1404. doi:10.1261/rna.071043.119

Kirpekar F, Hansen LH, Rasmussen A, Poehlsgaard J, Vester B. 2005. The archaeon *Haloarcula marismortui* has few modifications in the central parts of its 23S ribosomal RNA. *J Mol Biol* **348:** 563–573. doi:10.1016/j.jmb.2005.03.009

Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M. 2013. Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* **14:** R45. doi:10.1186/gb-2013-14-5-r45

Kiss T, Fayet-Lebaron E, Jády BE. 2010. Box H/ACA small ribonucleoproteins. *Mol Cell* **37:** 597–606. doi:10.1016/j.molcel.2010.01.032

Koš M, Tollervey D. 2010. Yeast pre-rRNA processing and modification occur cotranscriptionally. *Mol Cell* **37:** 809–820. doi:10.1016/j.molcel.2010.02.024

Lam WL, Doolittle WF. 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. *Proc Natl Acad Sci* **86:** 5478–5482. doi:10.1073/pnas.86.14.5478

Lecointe F, Namy O, Hatin I, Simos G, Rousset JP, Grosjean H. 2002. Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases *in vivo* recoding efficiency. *J Biol Chem* **277:** 30445–30453. doi:10.1074/jbc.M203456200

Li H. 2008. Unveiling substrate RNA binding to H/ACA RNPs: one side fits all. *Curr Opin Struct Biol* **18:** 78–85. doi:10.1016/j.sbi.2007.11.004

Li L, Ye K. 2006. Crystal structure of an H/ACA box ribonucleoprotein particle. *Nature* **443:** 302–307. doi:10.1038/nature05151

Li S, Duan J, Li D, Yang B, Dong M, Ye K. 2011. Reconstitution and structural analysis of the yeast box H/ACA RNA-guided pseudouridine synthase. *Genes Dev* **25:** 2409–2421. doi:10.1101/gad.175299.111

Liang B, Xue S, Terns RM, Terns MP, Li H. 2007a. Substrate RNA positioning in the archaeal H/ACA ribonucleoprotein complex. *Nat Struct Mol Biol* **14:** 1189–1195. doi:10.1038/nsmb1336

Liang XH, Liu Q, Fournier MJ. 2007b. rRNA modifications in an intersubunit bridge of the ribosome strongly affect both ribosome biogenesis and activity. *Mol Cell* **28:** 965–977. doi:10.1016/j.molcel.2007.10.012

Liang B, Zhou J, Kahen E, Terns RM, Terns MP, Li H. 2009. Structure of a functional ribonucleoprotein pseudouridine synthase bound to a substrate RNA. *Nat Struct Mol Biol* **16:** 740–746. doi:10.1038/nsmb.1624

Madern D, Ebel C, Zaccai G. 2000. Halophilic adaptation of enzymes. *Extremophiles* **4:** 91–98. doi:10.1007/s007920050142

Majumder M, Bosmeny MS, Gupta R. 2016. Structure-function relationships of archaeal Cbf5 during in vivo RNA-guided pseudouridylation. *RNA* **22:** 1604–1619. doi:10.1261/rna.057547.116

Manival X, Charron C, Fourmann J, Godard F, Charpentier B, Branlant C. 2006. Crystal structure determination and site-directed mutagenesis of the *Pyrococcus abyssi* aCBF5-aNOP10 complex reveal crucial roles of the C-terminal domains of both proteins in H/ACA sRNP activity. *Nucleic Acids Res* **34:** 826–839. doi:10.1093/nar/gkj482

Matera AG, Terns RM, Terns MP. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8:** 209–220. doi:10.1038/nrm2124

Motorin Y, Muller S, Behm-Ansmant I, Branlant C. 2007. Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol* **425:** 21–53. doi:10.1016/S0076-6879(07)25002-5

Mueller EG, Ferré-D'Amaré AR. 2009. Pseudouridine formation, the most common transglycosylation in RNA. In *DNA and RNA modification enzymes: structure, mechanism, function and evolution* (ed. Grosjean H), pp. 363–376. Landes Bioscience, Austin, TX.

Muller S, Leclerc F, Behm-Ansmant I, Fourmann JB, Charpentier B, Branlant C. 2008. Combined in silico and experimental identification of the *Pyrococcus abyssi* H/ACA sRNAs and their target sites in ribosomal RNAs. *Nucleic Acids Res* **36:** 2459–2475. doi:10.1093/nar/gkn077

Namy O, Lecointe F, Grosjean H, Rousset J-P. 2005. Translational recoding and RNA modifications. In *Fine-tuning of RNA function by modification and editing, Topics in Current Genetics* (ed. Grosjean H), Vol. 12, pp. 309–340. Springer-Verlag, Berlin, Heidelberg.

Ni J, Tien AL, Fournier MJ. 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89:** 565–573. doi:10.1016/S0092-8674(00)80238-X

Ofengand J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514:** 17–25. doi:10.1016/S0014-5793(02)02305-0

Ofengand J, Del Campo M, Kaya Y. 2001a. Mapping pseudouridines in RNA molecules. *Methods* **25:** 365–373. doi:10.1006/meth.2001.1249

Ofengand J, Malhotra A, Remme J, Gutgsell NS, Del Campo M, Jean-Charles S, Peil L, Kaya Y. 2001b. Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harb Symp Quant Biol* **66:** 147–159. doi:10.1101/sqb.2001.66.147

Osheim YN, French SL, Keck KM, Champion EA, Spasov K, Dragon F, Baserga SJ, Beyer AL. 2004. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. *Mol Cell* **16:** 943–954. doi:10.1016/j.molcel.2004.11.031

Peattie DA. 1979. Direct chemical method for sequencing RNA. *Proc Natl Acad Sci* **76:** 1760–1764. doi:10.1073/pnas.76.4.1760

Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP, Li H. 2006. Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21:** 249–260. doi:10.1016/j.molcel.2005.11.017

Rintala-Dempsey AC, Kothe U. 2017. Eukaryotic stand-alone pseudouridine synthases – RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol* **14:** 1185–1196. doi:10.1080/15476286.2016.1276150

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Huttenhofer A. 2003. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31:** 869–877. doi:10.1093/nar/gkg175

Spenkuch F, Motorin Y, Helm M. 2014. Pseudouridine: still mysterious, but never a fake (uridine)!. *RNA Biol* **11:** 1540–1554. doi:10.4161/15476286.2014.992278

Straub J, Brenneis M, Jellen-Ritter A, Heyer R, Soppa J, Marchfelder A. 2009. Small RNAs in haloarchaea: identification, differential expression and biological function. *RNA Biol* **6:** 281–292. doi:10.4161/rna.6.3.8357

Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Huttenhofer A. 2002. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc Natl Acad Sci* **99:** 7536–7541. doi:10.1073/pnas.112047299

Tran EJ, Zhang X, Maxwell ES. 2003. Efficient RNA 2′-*O*-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C′/D′ RNPs. *EMBO J* **22:** 3930–3940. doi:10.1093/emboj/cdg368

Watkins NJ, Bohnsack MT. 2012. The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3:** 397–414. doi:10.1002/wrna.117

Woolford JL J, Baserga SJ. 2013. Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*. *Genetics* **195:** 643–681. doi:10.1534/genetics.113.153197

Wu H, Feigon J. 2007. H/ACA small nucleolar RNA pseudouridylation pockets bind substrate RNA to form three-way junctions that position the target U for modification. *Proc Natl Acad Sci* **104:** 6655–6660. doi:10.1073/pnas.0701534104

Wu G, Xiao M, Yang C, Yu YT. 2011. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J* **30:** 79–89. doi:10.1038/emboj.2010.316

Xiao M, Yang C, Schattner P, Yu YT. 2009. Functionality and substrate specificity of human box H/ACA guide RNAs. *RNA* **15:** 176–186. doi:10.1261/rna.1361509

Ye K. 2007. H/ACA guide RNAs, proteins and complexes. *Curr Opin Struct Biol* **17:** 287–292. doi:10.1016/j.sbi.2007.05.012

Yu YT, Meier UT. 2014. RNA-guided isomerization of uridine to pseudouridine—pseudouridylation. *RNA Biol* **11:** 1483–1494. doi:10.4161/15476286.2014.972855

Zhou J, Liang B, Li H. 2011. Structural and functional evidence of high specificity of Cbf5 for ACA trinucleotide. *RNA* **17:** 244–250. doi:10.1261/rna.2415811