

Substrate Recognition by RNA 5-Methyluridine Methyltransferases and Pseudouridine Synthases: A Structural Perspective*

Published, JBC Papers in Press, November 3, 2006, DOI 10.1074/jbc.R600034200

Sun Hur, Robert M. Stroud, and Janet Finer-Moore¹

From the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

Through all kingdoms of life RNAs are modified as, or after, they are synthesized. The types and sites of modification are often conserved, implying conservation of function. Many of these modifications are located at key functional regions of the ribosome or other RNAs. The enzymes that carry out the modifications exhibit unique or limited multisite specificity.

Substrate recognition by RNA-modifying enzymes is more challenging than for DNA-modifying enzymes, because the complex tertiary folds of RNA often prevent direct read out of the target sequence. In some cases, there is no consensus target sequence for the multiple substrates of an RNA-modifying enzyme beyond the target base itself. A challenge for biologists is to determine how an RNA-modifying enzyme selects the correct segment from a large folded or partially folded RNA and then recognizes a single target base in the segment without relying exclusively on sequence. Several crystal structures of RNA-modifying enzymes with and without substrate bound have recently been determined. These structures are beginning to “decode” the basis for selectivity of enzymes for the many different states and folds of RNA.

We focus on two of the classes of RNA-modifying enzymes for which mechanisms of substrate recognition are understood to some degree. These are the 5-methyluridine (m^5U)² methyltransferases (MTases) and the pseudouridine synthases (Ψ synthases). These families catalyze some of the most common modifications of the more than 100 different types of posttranscriptional modification seen in RNA. Each modification is evoked at a particular stage in RNA folding suggesting that the determinants of selectivity often involve unique three-dimensional folded structures. In some cases a fragment of the physiological substrate, for example an isolated RNA stem-loop, can act as substrate; in others only folded or partially folded ribosome or full tRNA suffices. Both of

these classes of RNA-modifying enzymes use covalent addition of a nucleophilic side chain (Cys or Asp) to initiate the chemical modification of uridine. Synthetic minimal substrates containing 5-fluorouridine at the target site, when reacted with the proteins, yield covalent protein-substrate analog complexes, providing an effective means to crystallize enzyme-RNA complexes.

RNA m^5U MTases

The m^5U MTases catalyze the S_N2 transfer of the methyl group from the cofactor, AdoMet, to the C-5 of the target uridine (1). The first m^5U MTase to be studied in detail was TrmA (formerly called RUMT), which methylates U54 in the T-arm of most *Escherichia coli* tRNAs. An intriguing problem was how TrmA was able to recognize tRNAs with a different sequence and structure and catalyze the specific methylation at only U54. The minimal RNA structure for substrate activity contains the 7 bases of the T loop and a short base-paired stem of the T-arm (2). The composition of the base-paired stem is unimportant, and most base substitutions in the 7-base loop do not eliminate TrmA activity except for any mutation of the methyl acceptor U54 and the C56G mutation. Therefore the specificity of TrmA does not reside in the sequence but rather in secondary and tertiary structural features of the T-arm.

How does the enzyme-AdoMet binary complex gain access to the base of U54, which is buried deep inside the folded structure of tRNA, to perform the complex steps of catalysis? Stable hydrogen bonds between the D and T loops must be disrupted in order for TrmA to access the T loop. The kinetics of formation of the TrmA-tRNA complex is consistent with a two-step binding mechanism: a rapid association of the enzyme with tRNA followed by a slow step in which the enzyme and tRNA tightly associate into stable complexes that can be isolated on nitrocellulose filters (1). The overall rate of complex formation is significantly faster for tRNAs in which T loop/D loop hydrogen bonds have been eliminated by mutagenesis, implying the slow step involves disruption of T loop/D loop interactions (1).

If a substrate fragment is accessible but remains rigid as in the folded RNA structure, the target base may not be solvent-accessible. The crystal structure of the ribosomal m^5U MTase RumA bound to RNA and *S*-adenosylhomocysteine illustrates in detail the importance of conformational adaptability of an RNA substrate for recognition by the modification enzyme (3, 4). RumA methylates U1939 in a conserved region of *E. coli* 23 S RNA (5). A 37-nucleotide RNA with the same sequence as a 23 S RNA fragment containing the target uridine is a RumA substrate. The 37-mer, substituted with a 5-fluorouridine at U1939, was co-crystallized with the enzyme and an AdoMet analog, SAH. In the complex, the 5'-end loop adopts a new fold that complements the detailed shape and electrostatics of the protein surface and positions the 5-fluorouridine in the active site, where it is covalently bound to the catalytic Cys (Fig. 1). Another base from the 5'-loop, A1937, also inserts into the active site where it stacks against the adenine ring of the SAH. Not only does

* This minireview will be reprinted in the 2006 Minireview Compendium, which will be available in January, 2007. This review was written with support from National Institutes of Health Grant GM51232 (to R. M. S.) and a Larry Hillblom Foundation fellowship (to S. H.).

¹ To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, University of California, San Francisco, CA 94143. Tel.: 415-502-5426; Fax: 415-476-1902; E-mail: finer@msg.ucsf.edu.

² The abbreviations used are: m^5U , 5-methyluridine; F^5U , 5-fluorouridine; MTase, methyltransferase; Ψ , pseudouridine; SAH, *S*-adenosylhomocysteine; OB, oligonucleotide/oligosaccharide-binding; PUA, pseudouridine synthase and archaeosine-specific transglycosylase; archaeosine, 7-formamido-7-deazaguanosine; RNP, ribonucleoprotein particle; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA.

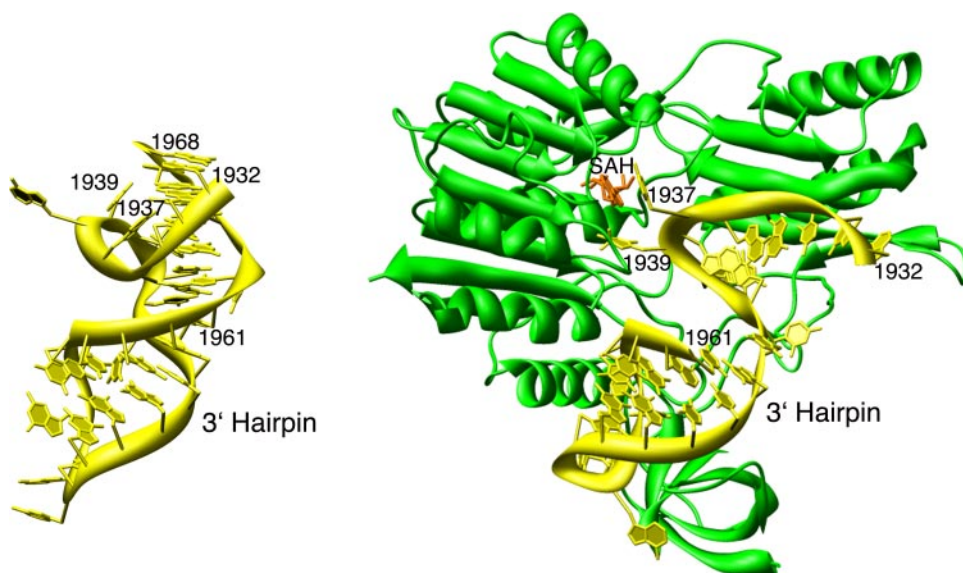


FIGURE 1. **Ribbon drawing of the RumA-SAH-RNA structure (right) (29).** The protein is green, the RNA is yellow, and SAH is orange. The SAH and the RNA bases are shown in stick representation, whereas the ribose moieties are represented by tubes. The ribbon plot on the left shows the conformation that the RumA substrate RNA assumes in the *E. coli* ribosome (from Protein Data Bank accession number 2AW4). The 5'-end loop of the RNA refolds upon binding to RumA, whereas the 3'-hairpin does not change conformation and interacts specifically only with the OB-fold domain. In the complex, nucleotides A1937 and F⁵U1939 (the target) are inserted in the active site and nucleotides 1962–1968 are disordered. This figure was produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by National Institutes of Health Grant P41 RR-01081).

A1937 assist in cofactor binding, but it also enhances catalysis by positioning the methyl group for transfer to the target base (4). Two other bases from the RNA loop are inserted into the spaces vacated by the F⁵U1939 and A1937, and the new fold is further stabilized by intra-RNA and RNA-protein hydrogen bonds. Protein-RNA interactions are enhanced by conformational changes in RumA, although these changes are far less dramatic than the refolding of the RNA substrate. The requirements for an RNA substrate to be able to “refold” into the unusual conformation and form specific hydrogen bonds with the protein are powerful constraints that help explain the high specificity of RumA.

Interactions between RNA-modifying enzymes and RNA substrate distal to the active site can enhance the chemical rate (k_{cat}) as well as the substrate binding affinity. The 3'-segment of the 37-mer co-crystallized with RumA is a hairpin that makes few contacts with the 5'-end loop and is distant from the target base. The hairpin binds to a small RNA-binding domain in RumA with an OB-fold, which is separated from the catalytic domain by a flexible linker. The 12-mer substrate RNA, in which the hairpin is absent, has a 30-fold lower k_{cat} and 110-fold lower catalytic efficiency (k_{cat}/K_m) than the 37-mer. The hairpin binding to the OB-fold domain apparently helps stabilize productive alignment of the target base and active site residues.

RNA Ψ Synthases

Ψ , the C-C glycosyl isomer of uridine, is the most common modification of RNA, and it occurs in most stable RNAs including tRNA, rRNA, tmRNA, and snRNAs (6, 7). In prokaryotes, pseudouridylation is mediated by a set of enzymes that recognize one or a few sites. Eukaryotes and archaea have analogous

site-specific Ψ synthases but also use a set of ribonucleoprotein particles (RNPs) for pseudouridylation. Guide snoRNAs in these RNPs determine substrate specificity. Although the reaction catalyzed by each of these enzymes is the same, the substrate specificity varies from simple stem-loop structures to larger and more complex three-dimensional RNA molecules.

Based on sequence alignments, Ψ synthases have been grouped into five families, each named after its first representative: TruA, TruB, TruD, RsuA, and RluA (8–10). Since the first crystal structure of a Ψ synthase was solved in 2000 (11), structures of Ψ synthases from all five families have been reported. The structures reveal that, despite limited sequence similarity across the families, the enzymes have a conserved core architecture and active site, consistent with their evolution from a common ancestor.

The core architecture is unique to Ψ synthases. It features two antiparallel β -sheets that are joined to each other by hydrogen bonds between two contiguous parallel strands, one from each antiparallel sheet, to form a continuous, bifurcated ~ 8 -stranded β -sheet that is decorated by helices and loops. The center of the extended sheet, where the antiparallel sheets join, forms the floor of the active site cleft. The walls of the cleft are formed by a conserved loop-helix structure on one side and by a long loop containing the invariant catalytic Asp on the other.

Apart from the conserved core the enzymes are structurally very diverse (Fig. 2). Inserts unique to a particular Ψ synthase, some quite long, cluster on either side of the active site cleft, suggesting they have a role in RNA recognition. Most Ψ synthases have a small RNA-binding domain, such as a PUA domain, attached to the N or C terminus of their catalytic domain (12). Others form homodimers or complexes with other proteins with some specific function.

TruB—TruB is responsible for the universally conserved $\Psi 55$ in the T stem-loops of tRNAs (13). TruB accepts as substrates isolated RNA stem-loops with the same sequences as portions of tRNA T stem-loops (14). Co-crystal structures of TruB with isolated stem-loops beautifully illustrate how the often elaborate inserts seen in Ψ synthases (Fig. 2) can participate in RNA binding and determine substrate specificity (15–17). TruB has a PUA domain at its C terminus and a 29-amino acid “thumb” domain inserted in the core architecture adjacent to the active site cleft, both of which make extensive interactions with RNA. Upon substrate binding, the thumb domain, which is disordered in the apoenzyme structure, adopts a helical conformation. At the same time three of the bases in the stem-loop,

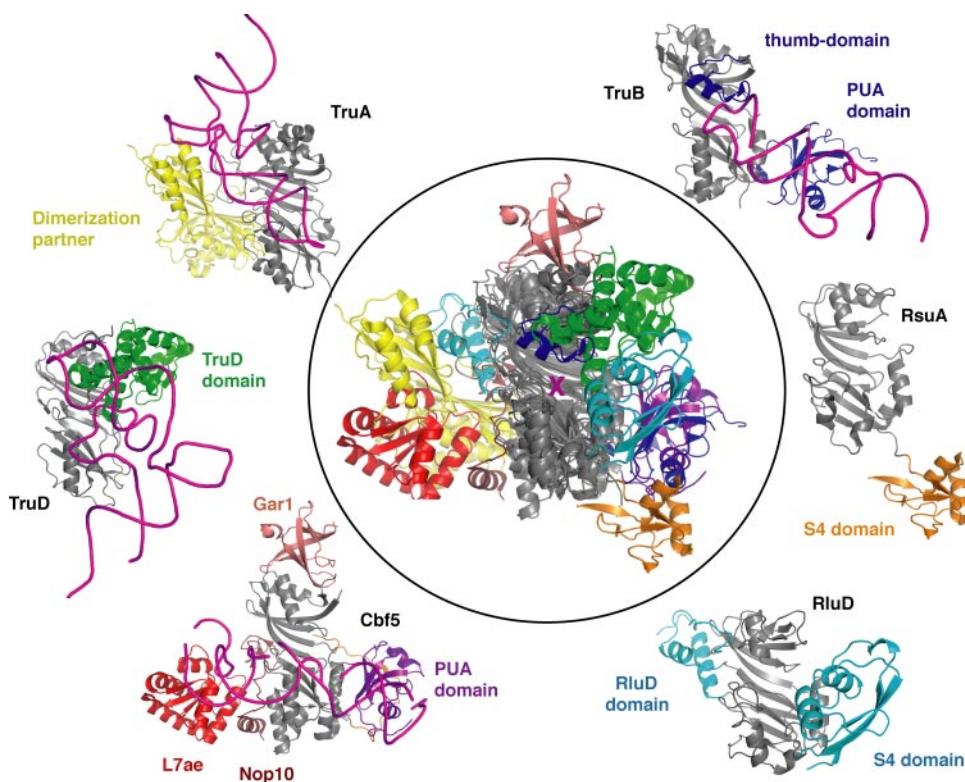


FIGURE 2. Inside the circle is shown the structural alignment of the following representatives from each of the five Ψ synthase families (Protein Data Bank accession codes are given in parentheses after the protein name): TruA (1DJ0), TruB (1R3E), TruD (1SB7), RsaA (1KSK), RluD (2IST), and Cbf5-L7a3-Nop10-Gar1 (2HVY). The conserved catalytic cores are colored gray with the invariant catalytic Asp marked with an "X", and the appended/inserted domains or accessory proteins are colored differently. Around the circle, each protein is shown separately in the same orientation as in the superposition. RNAs (magenta ribbons) bound to TruB and Cbf5 are taken from the crystal structure of the respective substrate-RNA complex, whereas those bound to TruA and TruD are models. The RNA bound to Cbf5 is the guide snoRNA, not the substrate RNA. This figure was produced with PyMOL, from Delano Scientific, San Carlos, CA.

including U55, flip into the active site. The thumb domain clasps the RNA and makes stabilizing interactions with nucleotides bound at the active site. In contrast to TrmA, which recognizes the overall shape rather than specific sequence of the T loop, TruB requires the consensus sequence U54-U55-X56-X57-A58 (where X is any nucleotide) for function but tolerates one-base changes to loop size (14). The thumb domain-RNA interactions are proposed to play a role in interrogation of the consensus sequence (16). The "thumb" domain insert is unique to Ψ 55 synthases in the TruB family.

In three independently solved TruB-RNA crystal structures the stem-loop bound at the active site is stacked end-to-end with either the stem-loop from a neighboring complex (15, 17) or with an RNA duplex formed by extra copies of the co-crystallized RNA fragment (16). The stacked RNAs mimic a single extended RNA stem-loop bound at the active site. In all three of the crystal structures the PUA domain has moved as a rigid body toward the active site and makes nonspecific interactions with the extended stem. When whole tRNA is superimposed on the TruB complex by aligning its T stem-loop with the active site-bound RNA fragment, the tRNA acceptor stem-loop aligns with the second, "stacked" RNA fragment. These three independent observations of PUA-RNA interactions indicate the strong tendency of the PUA domain to bind RNA and, according to the TruB-tRNA model, to recognize the acceptor stem-

loops in the intact tRNA. A PUA domain in an archaeal tRNA-modifying enzyme, the archaeosine-specific transglycosylase, also interacts with the acceptor stem-loop of tRNA, despite having low sequence homology to the TruB PUA domain (18). On the other hand, in the eukaryotic Ψ 55 synthase Pus4, a domain with an unknown fold replaces the PUA domain, suggesting that the PUA domain is a generic RNA binding motif that can be interchanged with a variety of other such motifs (19).

TruD—Although TruD is evolutionarily distant from all other Ψ synthases, its structural and functional similarities to TruB provide clues to its substrate recognition mechanism. TruD modifies U15 in the D stem-loop of tRNA. Like TruB, TruD must dissociate the D and T stem-loops that are joined by hydrogen bonds in the folded tRNA and modify a single site with high precision. TruD has a \sim 140-amino acid insertion in the C-terminal half of the protein on the RNA binding face of the protein. The position of this TruD-specific insert, called the TRUD domain, and its highly positive electrostatic nature are reminiscent

of the thumb domain in TruB. Comparison of multiple structures of apo-TruD reveals a hinge motion of \sim 18° that alters the relative positions of the TRUD domain and the conserved core, suggesting the TRUD domain may clamp down on the RNA substrate much like the thumb domain of TruB (20–22). The structure of a TruD-tRNA complex has been modeled by assuming the tRNA adopts the same conformation as it does when bound to a transglycosylase that modifies position 15 in the tRNA D loop of many archaea. In order for the D loop to access the active site of the transglycosylase the tRNA assumes a conformation radically different from the canonical L-shaped tRNA (18). When TruD is docked to tRNA in this alternate conformation, the TRUD domain interacts with the anticodon stem-loop as well as the D stem-loop, whereas in TruB-tRNA models the smaller (29 amino acids) thumb domain interacts only with the T stem-loop (20). The TRUD domain interactions with the anticodon stem-loop may provide one mechanism by which TruD distinguishes between the D and T stem-loops of tRNA.

TruA—TruA modifies uridines at positions 38–40 of the anticodon stem-loop of tRNAs; thus it is an example of a Ψ synthase with regional as opposed to site specificity. TruA is the only Ψ synthase in *E. coli* that functions as a homodimer. Unlike RumA or TruB, TruA requires a whole tRNA for modification; an isolated anticodon stem-loop is not a substrate. A

model of a TruA-tRNA complex that is consistent with these data has been constructed by docking whole tRNA to an electropositive surface surrounding the active site cleft of the apo-TruA crystal structure such that the target sites of the tRNA are placed near the active site and the body of tRNA contacts the protein (11) (Fig. 2). This binding mode has recently been confirmed by our crystal structure.³ In this binding mode, the tRNA body is bound by the second subunit of the TruA homodimer. This interaction may initiate the TruA-tRNA association by orienting the tRNA such that the target sites are localized near the active site but with enough degrees of freedom that any one of them can productively bind to the catalytic Asp.

RsuA and RluA Family

The RsuA and RluA families are the two most closely related of the five families based on sequence similarities. Full-length crystal structures of RsuA (23) and of the RluA family member RluD⁴ have been solved. Each has an N-terminal domain that is connected to the catalytic domain by a flexible linker. These domains have the same fold as the RNA-binding domain of ribosomal protein S4, and hence are referred to as S4 domains (12). Superpositions of RsuA and RluD onto TruB show that the S4 domains can interact extensively with an RNA helix that extrudes away from the binding pocket like the extended stems in the TruB-RNA complexes. However, the S4 domain does not appear to be absolutely required for the function of enzymes in the RluA family as RluA and TruC from this family do not have the domain.

At least in prokaryotes, RsuA family members are highly specific, usually modifying a single site on rRNA. RluA family members, in contrast, typically show broader, regional specificity. Neither sequence differences nor the apoenzyme structures provide obvious explanations for the differences in selectivity, and no crystal structures of substrate-bound enzymes have been reported. Therefore, the substrate recognition mechanisms for the RsuA and RluA family are still poorly understood.

H/ACA RNPs: Use of snoRNAs for Substrate Recognition

Eukaryotes and archaea utilize a different strategy than prokaryotes for selecting rRNA sites for the Ψ modification (24). A TruB homolog, Cbf5, forms a heterotetramer with three accessory proteins. The heterotetramer then binds to one of a family of guide snoRNAs that are characterized by two conserved sequence motifs, called the box H (sequence ANANNA) and the box ACA motifs. The assembled H/ACA snoRNP then recognizes the target uridine through transient base pair interactions between the snoRNA and sequences in the rRNA on either side of the target site (25). Thus, ~50–100 target sites in rRNA can be efficiently modified by a single protein complex simply by using a different H/ACA snoRNA for each target site.

Crystal structures of an archaeal H/ACA snoRNP reveal the close structural homology between Cbf5 and TruB (26–28). Cbf5 has a PUA domain, which interacts with the ACA motif at the 3'-end of the guide snoRNA and a flexible loop, corre-

sponding to the thumb domain of TruB, which could bind and stabilize substrate RNA. The PUA-ACA interactions precisely align the guide sequences of the snoRNA with the active site of Cbf5 (27). The three accessory proteins in the archaeal RNP (L7a3, Nop10, and Gar1), which are essential for function, have been proposed to play roles in supporting the snoRNA, stabilizing the RNP, and possibly guiding the target RNA to the active site (26–28). Although the structures of the RNP do not contain substrate RNA, their parallels to the TruB structures have suggested detailed mechanisms for snoRNA-guided substrate binding (27). Because the RNP can be assembled with mixtures of eukaryotic and archaeal components and there is a high degree of sequence identity between corresponding components from the two kingdoms, the proposed mechanisms for RNA recognition can likely be generalized to eukaryotes (26).

Conclusions

RNA-modifying enzymes face the challenging task of binding target nucleotides deeply buried in the folded or partially folded RNA. Structural studies of m⁵U MTases and Ψ synthases have revealed some common strategies for substrate recognition. These include electrostatic attraction and shape complementarity, stabilization of new substrate RNA folds, and use of flexible protein loops to precisely position the target nucleotide in the active site. The wide range of substrates and different degrees of substrate specificity for the enzymes in these families implies each enzyme's substrate binding mechanism will have unique features that we can begin to understand through the structures of their substrate complexes.

Acknowledgment—We thank Pat Greene for critically reading this manuscript.

REFERENCES

- Kealey, J. T., Gu, X., and Santi, D. V. (1994) *Biochimie (Paris)* **76**, 1133–1142
- Gu, X., Ivanetich, K. M., and Santi, D. V. (1996) *Biochemistry* **35**, 11652–11659
- Lee, T. T., Agarwalla, S., and Stroud, R. M. (2004) *Structure* **12**, 397–407
- Lee, T. T., Agarwalla, S., and Stroud, R. M. (2005) *Cell* **120**, 599–611
- Agarwalla, S., Kealey, J. T., Santi, D. V., and Stroud, R. M. (2002) *J. Biol. Chem.* **277**, 8835–8840
- Samuelsson, T., and Olsson, T. (1990) *J. Biol. Chem.* **265**, 8782–8787
- Patton, J. R. (1993) *Biochem. J.* **290**, 595–600
- Gustafsson, C., Reid, R., Greene, P. J., and Santi, D. V. (1996) *Nucleic Acids Res.* **24**, 3756–3762
- Kaya, Y., and Ofengand, J. (2003) *RNA (N. Y.)* **9**, 711–721
- Koonin, E. V. (1996) *Nucleic Acids Res.* **24**, 2411–2415
- Foster, P. G., Huang, L., Santi, D. V., and Stroud, R. M. (2000) *Nat. Struct. Biol.* **7**, 23–27
- Aravind, L., and Koonin, E. V. (1999) *J. Mol. Evol.* **48**, 291–302
- Watanabe, Y., and Gray, M. W. (2000) *Nucleic Acids Res.* **28**, 2342–2352
- Gu, X., Yu, M., Ivanetich, K. M., and Santi, D. V. (1998) *Biochemistry* **37**, 339–343
- Hoang, C., and Ferre-D'Amare, A. R. (2001) *Cell* **107**, 929–939
- Pan, H., Agarwalla, S., Moustakas, D. T., Finer-Moore, J., and Stroud, R. M. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12648–12653
- Phannachet, K., and Huang, R. H. (2004) *Nucleic Acids Res.* **32**, 1422–1429
- Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003) *Cell* **113**, 383–394
- Becker, H. F., Motorin, Y., Planta, R. J., and Grosjean, H. (1997) *Nucleic*

³ S. Hur and R. M. Stroud, unpublished data.

⁴ P. G. Foster, J. Finer-Moore, and R. M. Stroud, unpublished data.

- Acids Res.* **25**, 4493–4499
20. Ericsson, U. B., Nordlund, P., and Hallberg, B. M. (2004) *FEBS Lett.* **565**, 59–64
 21. Hoang, C., and Ferre-D'Amare, A. R. (2004) *RNA (N. Y.)* **10**, 1026–1033
 22. Kaya, Y., Del Campo, M., Ofengand, J., and Malhotra, A. (2004) *J. Biol. Chem.* **279**, 18107–18110
 23. Sivaraman, J., Sauve, V., Larocque, R., Stura, E. A., Schrag, J. D., Cygler, M., and Matte, A. (2002) *Nat. Struct. Biol.* **9**, 353–358
 24. Ofengand, J. (2002) *FEBS Lett.* **514**, 17–25
 25. Meier, U. T. (2005) *Chromosoma (Berl.)* **114**, 1–14
 26. Hamma, T., Reichow, S. L., Varani, G., and Ferre-D'Amare, A. R. (2005) *Nat. Struct. Mol. Biol.* **12**, 1101–1107
 27. Li, L., and Ye, K. (2006) *Nature* **443**, 302–307
 28. Rashid, R., Liang, B., Baker, D. L., Youssef, O. A., He, Y., Phipps, K., Terns, R. M., Terns, M. P., and Li, H. (2006) *Mol. Cell* **21**, 249–260
 29. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J. Comput. Chem.* **25**, 1605–1612

