RNA-guided Nucleotide Modification of Ribosomal and Other RNAs*

Published, JBC Papers in Press, November 12, 2002, DOI 10.1074/jbc.R200023200

Wayne A. Decatur and Maurille J. Fournier†
From the Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, Massachusetts 01003

One of the exciting frontiers in the field of RNA editing is the phenomenon of RNA-guided nucleotide modification. In this type of editing, a nucleotide in a precursor RNA is converted to another form by an RNA-protein complex (RNP)1 (1). The RNPs that mediate these reactions include a guide RNA that provides site specificity through base pairing with the substrate and a set of proteins, one of which catalyzes the modification reaction. The phenomenon was first discovered in the modification of ribosomal RNA (rRNA) in the nucleolus of eukaryotic cells (Fig. 1). Two common alterations are relevant, formation of 2',O-methylated nucleosides (Nm; the guided mechanism was reported in 1996) and conversion of uridine to pseudouridine (Ψ; the guided process was reported in 1997) (2–4). These modifications are mediated by two large, heterogeneous populations of RNPs that are modification type-specific and site-specific. The RNPs contain a small nucleolar RNA (snoRNA) and several associated proteins, and the snoRNA-protein complexes are called snRNP ("snorps"). The snoRNA provides the guide function, and an integral snoRNP protein catalyzes the modification reaction. When discovered, this type of reaction scheme was not only novel but in sharp contrast to the rRNA modification schemes used by Eubacteria, where the synthesis of Nm and Ψ is mediated (thus far) by protein enzymes that do not include an RNA co-factor (5). Guided modification was subsequently discovered to apply to the U6 snRNA (small nuclear RNA in vertebrates and Caenorhabditis elegans) and likely to mRNAs (mammals, trypanosomes) (6–9). Strikingly, from an evolutionary perspective, the new paradigm was also discovered (in 2000) to apply to Archaeal organisms where substrates include tRNA as well as rRNA (10).

Recent advances have revealed guided modification to be more complex and widespread and are almost certainly a harbinger of exciting new developments to come. Key developments include: 1) identification of new guide RNAs that reside in mammalian Cajal bodies (these RNAs are specific for the four snRNAs transcribed by RNA polymerase II (pol II), which are thought to undergo maturation and possibly RNP assembly at this location (11)); 2) evidence that the trypanosome trans-spliced leader is a substrate for guided modification (9); and 3) successful development of the first cell-free Nm modification system, using recombinant archaeal components (12). Taken together, these findings argue that additional modifying RNPs and substrates subjected to RNA-guided modification will be discovered. In this minireview we describe the present state of knowledge about the various RNP-modifying complexes, the processes they mediate, where in the cell these reactions occur, and the range of substrates. Because of limited space the reader is also referred to other recent reviews (1, 13–18).

Occurrence and Effects of Nm and Ψ Modifications

The Nm and Ψ nucleotides appear to be universal among rRNAs and small stable RNAs such as splicing snRNAs, tRNAs, and snoRNAs; however, the abundance and locations of these nucleotides vary phylogenetically (19–21). Although the modifications are believed to be beneficial, evidence is still sparse. Blocking Nm or Ψ modification at a global level in yeast rRNA has strong negative effects on growth rate (22, 23), and the absence of Nm and Ψ in U2 snRNA impairs its assembly into an active spliceosome in Xenopus oocytes (24). New three-dimensional modification maps of rRNA show the modifications are heavily concentrated in regions of the ribosome known or predicted to be functionally important, suggesting that the modifications benefit ribosome activity, either directly or indirectly (25–27). At the level of individual modifications blocking Nm or Ψ synthesis in bacterial or yeast rRNA has, thus far, resulted in neither a detectable effect or only a slight effect on growth rate, implying that many (perhaps most) modifications affect ribosome structure and function in a synergistic way (25, 27, 28). The changes in RNA structure caused by modification could “fine-tune” many events in rRNA folding, ribosome assembly, or ribosome activity as well as trafficking and half-life. The same reasoning applies to modification of the small RNAs (e.g. Ref. 29).

Guided Nm and Ψ Modifications in the Eukaryotic Nucleolus

Since their discovery, scores of guide snoRNAs have been identified for rRNA, primarily in yeast and human cells, and it seems likely that most Nm and Ψ modifications in cytoplasmic rRNA are formed by snoRNPs (1). In support of this view, candidate guide snoRNAs have been identified in yeast for 51 of 55 known Nm sites and 30 of 44 Ψ sites (30, 31). However, it is possible that some modifications are created by enzymes without an RNA cofactor. The snoRNPs are thought to act early in rRNA synthesis as the modification level of uncleaved, primary transcripts is high (14, 15, 32–34). This situation is consistent with modification occurring co-transcriptionally or in large pre-ribosome complexes that evolve into the individual subunits (15) (Fig. 1A). In addition to mediating modification reactions, a few snoRNPs are required for processing (cleavage) of pre-rRNA (17). The corresponding snoRNAs interact directly with rRNA, but the actual functions of the snoRNPs in processing are not known in most cases. Additional roles for snoRNAs also seem possible. For example, in the context of guided modifica-

* This minireview will be reprinted in the 2003 Minireview Compendium, which will be available in January, 2004. This work was supported by National Institutes of Health Grant GM19351.
† To whom correspondence should be addressed: Dept. of Biochemistry and Molecular Biology, Lederle Graduate Research Tower, University of Massachusetts, Amherst, MA 01003. Tel.: 413-545-0353; Fax: 413-545-3291; E-mail: 4nier@biochem.umass.edu.
‡ W. A. Decatur and M. J. Fournier, unpublished data.
a, a snoRNP could affect other aspects of RNA synthesis or function in addition to modification (see below).

Each general class of guide snoRNAs and snoRNPs is both modification type-specific and site-specific (Fig. 1, B and C). Each class of snoRNA contains one or two targeting motifs that act independently at sites in the same or different pre-rRNA(s), and each class of snoRNP contains a different set of four common core proteins. During modification, the guide sequence selects (by complementary base pairing) a target sequence in the substrate RNA, and modification occurs within this target sequence at a characteristic distance from a short “box” element in the interacting snoRNA.

The Nm guide snoRNAs contain one or two pairs of small, distinguishing sequence elements called boxes C and D, and C’ and D’. These elements define the C/D family of snoRNAs, which also includes a few species involved in rRNA processing (16, 17). Boxes C/D occur near the 5′ and 3′ ends of the RNA and boxes C’/D’ are located internally. The methylation guide sequence is located upstream of the box D/D’ element and consists of 10–21 nucleotides (1, 35).

The canonical C/D boxes are required for snoRNA stability and proper end formation (note that vertebrate snoRNAs are typically derived from introns of protein genes (16)). Remarkably, the C/D elements are also necessary and sufficient for localizing snoRNAs to the nucleolus, which occurs by way of the Cajal bodies in vertebrate cells and in at least some cases functionally related nucleolar bodies in yeast (11, 18, 36) (see below). These last findings indicate that snoRNA synthesis and localization are coupled (16).

The methylating snoRNPs have four common core proteins. The total number of proteins in a particle is not known nor is it known if all modifying snoRNPs are identical except for the RNA component. The core proteins in yeast (and humans) are: Snu13p (15.5 kDa), Nop56p (hNop56p), Nop58p (hNop58p), and Nop1p (fibrillarin). Snu13p binds to a characteristic stem-asymmetric loop-stem structure that includes the canonical C/D elements in the loop portion (Fig. 1B). Interestingly, Snu13p is also part of the U4 snRNP where it binds a similar, common RNA fold called the K-turn (37, 38). Nop56p and Nop58p are related to each other and both interact with snoRNA (16, 39).

Nop1p (fibrillarin) is generally accepted to be the 2′-O-methyltransferase. In support of this contention, point mutations in methylase-like elements have been shown to block ribose methylation globally in yeast cells. In addition, a crystal structure of an archaeal ortholog that contains the methylase signature elements showed that most of the protein has a three-dimensional structure like that of many known S-adenosylmethionine-dependent methylases (40). In a process yet to be defined but assuredly interesting, the snoRNP methylase acts on a ribose of a nucleotide that is initially base paired with the snoRNA guide sequence. The target site is 5 nucleotides upstream of box D/D’, located 4–5 nucleotides within the region of complementarity (Fig. 1B). Key mechanistic questions to be resolved include whether base pairing actually occurs over the full length of the guide sequence (9 base pairs are required for methylation to occur (41)) and if accessibility to the target nucleotide involves the action of other snoRNP or non-snoRNP proteins such as a helicase. Each type of box element is required for methylation: boxes D and D’ because they are spatial determinants and the C/D and C′/D′ pairs because they affect protein binding (1, 37, 39).

The Ψ guide snoRNAs have characteristic small sequence elements referred to as boxes H and ACA and are members of a larger family of H/ACA snoRNAs (1, 27, 42) (Fig. 1C). Like the C/D elements, the H and ACA boxes and neighboring duplexes are required for processing of snoRNA precursors, protein binding, and localization (16, 18). Most H/ACA snoRNAs are guide
The snoRNAs that guide mammals but not yeast), guides telomere formation (17, 18). However, as with the C/D snoRNA family, a few partic-
tances in a loop portion of the duplex structures and a dis-
plex-tail–duplex–hinge–stranded hinge and 3′ tail segments, respectively. Substrate
targeting involves base pairing through two short guide se-
lements conserved among known Y synthases, and point mu-
tions in two such elements in yeast Cbf5p disrupt Y formation in rRNA in a global manner (23). Natural mutations in
human dyskerin have been linked to a premature aging disease in humans (dyskeratosis congenita), but interference with te-
lerase function may be the basis of the disease rather than
defective Y synthesis (17). Interestingly, the Nhp2p protein in
the H/ACA snoRNPs and the Snu13p protein in the C/D
snoRNPs are related to each other (in yeast, 38% identity and
61% similarity), suggesting the proteins have related functions.
The bi-partite nature of the H/ACA snoRNAs is reflected in
electron micrograph images of two yeast snoRNPs, where two
oblong lobes are joined at one end to form a V-like structure estimated at 15 nm long and 12 nm wide (43). In the context of
a growing ribosome, these snoRNPs are roughly 10% the mass
of the ribosome, which would favor their acting before a com-
compact preribosomal RNP is formed (Fig. 1A).
Results from recent structure studies of a bacterial Y syn-
that provide valuable insight into how this type of modifying
enzyme interacts with its substrate (44). The results should
apply to guided Y formation too, as the bacterial enzyme,
Escherichia coli TruB, and Cbf5p are homologs, and Cbf5p has
the motifs that occur in the active site region of TruB. The
crystal structure of a TruB enzyme complexed with a tRNA
fragment reveals that the enzyme causes the target uracil to
flip out and the folded tRNA structure to be destabilized. These
distortions could provide access for a conserved aspartate to
attack and initiate the reaction in which the glycosidic bond is
broken and the uracil base is rotated and reattached (44, 45).

New Types of Guide RNAs, Cellular Locations, and Substrates in Eukaryotes

In a surprising development, new candidate human guide
RNAs have been discovered to reside exclusively in Cajal (coiled)
汽車 bodies (11, 46). The small Cajal body-specific RNAs, called sca-
RNAs, have the hallmark features of the archetypical snoRNAs,
except that some have an unusual arrangement and content of
guide motifs. Among the current set of scaRNAs are: species with
both Nm and Y modification motifs and species that resemble the
classically defined guide snoRNAs. Guide sequences have been
identified thus far in the scaRNAs for known modifications in pol
II-transcribed snRNAs (U1, U2, U4, and U5). Newly synthesized snoRNAs that participate in modification
(and processing) of pre-rRNA are also known to localize to Cajal
bodies before entering the nucleolar complex (11, 47–49). These
results suggest that the nuclear specific guide RNAs (and
possibly additional RNAs) might also undergo modification in
Cajal bodies (CBs). The new dichotomy in localization of

snoRNA-like guide RNAs indicates diversity in the determi-
nants responsible for retaining RNA in the Cajal bodies. Selection
could involve substrate binding or interaction with CB
proteins that are specific for the nascent scaRNPs.

Adding to the excitement, new results with yeast have iden-
tified novel structures called nucleolar bodies (NBs), which
have overlapping functions with Cajal bodies (36, 50) (Fig. 2).
Newly synthesized U3 and U14 snoRNAs, and C/D and H/ACA
snoRNPs proteins have been detected in the NBs. Strong evidence
that both snoRNAs and snRNAs undergo maturation events in nucleolar bodies comes from the discovery that an
enzyme, Tgs1p, that hypermethylates the 5′-cap structures of a subset of snoRNAs and the pol II snRNAs appears to be localized
to this structure (51). Consistent with parallel functions of CBs and NBs, the vertebrate ortholog of Tgs1p occurs in the
Cajal bodies, suggesting that cap formation of snoRNAs takes
place in this nuclear compartment (50). (Note that most verte-
brate snoRNAs undergo 5′-end processing and lack caps.) The
parallel is not complete, however, as hypermethylation of pol II
snRNAs is known to occur in the cytoplasm and Tgs1 protein
localizes to the cytoplasm as well as the CBs (50).

It seems possible that a variety of mRNAs might be sub-
strates for guided modification, based on two fascinating recent
reports. In the first, an analysis of mouse and human brain
small RNAs identified a candidate Nm snoRNA with a guide
sequence that is complementary (18 nt) to mRNA for a sero-
tonin receptor (8). The mRNA transcripts undergo alternative
splicing and adenosine to inosine editing at four sites to yield
proteins with different signal transduction potentials. Guided
methylation is predicted to occur at one of the editing sites, all
of which are close to each other (within a 13-nt span) and
proximal to an alternative splice site (10 nt from the closest
editing site). Clearly, modification at such an editing site could
have several important effects on subsequent expression and
function of the protein products. Intriguingly, this putative
guide RNA and several others are encoded at imprinted gene
loci, suggesting they may have a role in imprinting (8, 52).
Also striking is the evidence for T-Granulocyte split
leader (SL RNA) undergoing guided Y modification (9). Here, a
candidate guide RNA was identified (in both the nucleolus
and nucleoplasm) that is predicted to target a site known to
undergo modification, and mutating the region of complementar-
ity in the SL RNA blocked this modification. Because the SL
RNA is spliced to all mRNAs, a single modification could affect
the synthesis and activity of many proteins. These two situa-
tions with mRNAs raise the specter that RNA-guided modi-
fication may play important roles in altering the structure and
function of mRNAs (and proteins) as well as stable RNAs.

FIG. 2. Guided modification appears to occur in vertebrate
Cajal bodies. New snoRNAs enter the CBs before localizing to the
nucleolus (NO); however, a subset is retained there. The small Cajal
body-specific RNAs (scaRNAs) are complementary to pol II-derived
snRNAs, which are also found in CBs (11). Modification of yeast sno-
RNAs and snRNAs may occur in yeast NBs, which share properties
with Cajal bodies (50).
Archaeal rRNA and tRNA Are Also Modified by snoRNP-like Complexes

The Archaeal kingdom contains orthologous components of the eukaryotic Nm and Ψ snoRNPs (10). Scores of candidate Nm guide RNAs for rRNA and tRNA have been identified, and recently, the first putative Ψ guide RNAs have also been reported (four RNA species) (53). Although in vivo verification of guide function is not yet possible, modifications are known to occur at many predicted Nm sites in rRNA and tRNA and at six predicted Ψ sites in tRNA. The snoRNA-like guide RNAs are called sRNAs and the modifying complexes are sRNPs. Relative to the eukaryotic counterparts, the archaeal machinery is somewhat simpler. Although the Nm guide sRNAs can have one or two targeting motifs, the sizes are generally at the lower end of the eukaryotic size range and they have shorter guide elements. Three candidate Ψ sRNAs identified thus far are unusually short and contain only one of the two targeting domains of the archetypical guide snoRNAs; these RNAs closely resemble Ψ guide RNAs in Trypanosomes (9), which are early branching eukaryotes. Provocatively, a candidate Nm guide RNA has been discovered in an intron of an archaeal pre-tRNA that is specific for two sites methylated in the same tRNA; this arrangement suggests that RNA-guided modification may be able to act in cis (10, 54). For the proteins, the set of C/D core proteins is simpler too, with three proteins rather than four. Accounting for this latter difference, archaeal C/D sRNPs contain a single ortholog of the eukaryotic Nop56p and Nop58p proteins (aNop56). Like its eukaryotic counterpart, the archaeal variant of Snu13p/15.5 kDa, ribosomal protein Al7a, is a dual purpose protein, which binds to kink turns in both rRNA and sRNA (55). All four orthologs of the eukaryotic Ψ snoRNPs have been identified in the archaea (56).

In a particularly important breakthrough, RNA-guided methylation has been demonstrated in vitro, with a simple, reconstituted archaeal sRNP and a short rRNA fragment as substrate (12). Accurate, site-specific methylation of the rRNA fragment was achieved with an sRNP complex formed by incubating a cognate sRNA (expressed in vitro) with the three C/D core proteins. This demonstration shows that an sRNP complex containing only the core proteins and guide RNA is sufficient to catalyze site-specific methylation; the results also strengthen the belief that Nop1p is the Ψ methylase. This advance should spur attempts to establish homologous eukaryotic systems and in vitro systems for Ψ modification.

Perspective

It is reasonable to expect that the list of RNAs subjected to RNA-guided modification will increase, in particular for eukaryotic nascent RNAs that move through the Cajal bodies or nucleolus. In addition to U1, U2, U4, and U5 snRNAs, Cajal bodies also contain U11 and U12 splicing snRNAs and U7 snRNA (involved in 3′-end formation of histone mRNAs). Similarly, the growing number of RNAs that appear to have a nucleolar phase is also growing and now includes the U6 snRNA, tRNA, RNase P RNA, signal recognition particle RNA, telomerase RNA, mRNA, and perhaps HIV transcripts (57, 58). Moreover, the many new, small non-coding RNAs discovered recently are also candidates for guided modifications (59). The results of the recent past suggest that the field of RNA-guided modification will remain an exciting frontier of the RNA world for many years.

REFERENCES

1. Kiss, T. (2001) EMBO J. 20, 3617–3622
2. Kiss-Laszlo, Z., Henry, Y., Bachelier, J. P., Caizergues-Ferrer, M., and Kiss, T. (1998) Cell 85, 1077–1088
3. Ni, J., Tien, A. L., and Fournier, M. J. (1997) Cell 89, 565–573
