AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136
(Some corrections may occur before final publication online and in print)

Annu. Rev. Biochem. 2005. 74:199–217
doi: 10.1146/annurev.biochem.74.082803.133136

First published online as a Review in Advance on February 18, 2005

AN ABUNDANCE OF RNA REGULATORS*

Gisela Storz,[1] Shoshy Altuvia,[2] Karen M. Wassarman[3]
[1]*Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892; email: storz@helix.nih.gov*
[2]*Department of Molecular Genetics and Biotechnology, The Hebrew University Faculty of Medicine, Jerusalem 91120, Israel; email: shoshy@cc.huji.ac.il*
[3]*Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706; email: wassarman@bact.wisc.edu*

Key Words noncoding RNA, antisense RNA, small RNA, microRNA

■ **Abstract** The importance of small, noncoding RNAs that act as regulators of transcription, of RNA modification or stability, and of mRNA translation is becoming increasingly apparent. Here we discuss current knowledge of regulatory RNA function and review how the RNAs have been identified in a variety of organisms. Many of the regulatory RNAs act through base-pairing interactions with target RNAs. The base-pairing RNAs can be grouped into two general classes: Those that are encoded on the opposite strand of their target RNAs such that they contain perfect complementarity with their targets, and those that are encoded at separate locations on the chromosome and have imperfect base-pairing potential with their targets. Other regulatory RNAs act by modifying protein activity, in some cases by mimicking the structures of other RNA or DNA molecules.

CONTENTS

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

INTRODUCTION

There has been an explosion in the identification and characterization of RNAs that have functions other than those of messenger RNAs, transfer RNAs, or ribosomal RNAs. Many of these RNAs are integral components of RNA-protein complexes, in which the RNAs serve a variety of functions. For instance, the U small nuclear RNA (snRNA) constituents of the eukaryotic spliceosome utilize base-pairing interactions to identify splice sites as well as to interact with other spliceosomal RNAs [reviewed in (1)]. The U snRNAs ultimately may serve as the catalytic center within the spliceosome. The RNA present in the signal recognition particle, 4.5S RNA and SRP RNA in bacteria and eukaryotes, respectively, is a major structural component of the ribonucleoprotein complex in which it resides [reviewed in (2)]. The telomerase RNA serves as a template for the synthesis of telomeres on the ends of eukaryotic chromosomes [reviewed in (3)]. RNase P RNA has intrinsic enzymatic function, although it too is found within a ribonucleoprotein in vivo [reviewed in (4)]. Bacterial SsrA/tmRNA has a unique mechanism of action in which it serves as both a tRNA mimic and an mRNA, functioning to release ribosomes trapped on damaged mRNAs [reviewed in (5)]. In addition to these RNA components of critical cellular enzymes, a large number of newly identified RNAs have been found to function as regulators. As illustrated in Figure 1, these regulatory RNAs impact all steps in gene expression pathways. It is these RNAs and how they function that are the focus of this review. In general, regulatory RNAs act via one of two basic mechanisms: base-pairing interactions with other nucleic acids and binding to and modifying the activity of a protein or protein complex.

Studies of the regulatory RNAs have largely occurred in parallel in bacterial and eukaryotic systems and are in the early stages in archaea. Examination of the mechanisms of action of many of the RNAs in bacteria and eukaryotes points to a surprising number of similarities between regulatory RNAs in all types of organisms. Thus, one aim of this review is to compare and contrast what has been learned about the regulatory RNAs found in bacteria and eukaryotes. It is clear that approaches used in one system are applicable to other systems, and more importantly, insights gained regarding the mechanism of action of RNAs in one organism often are relevant to the regulatory RNAs found in other organisms.

Unfortunately, the nomenclature used for regulatory RNAs has not been uniform. Generally, the term small RNA has predominated for the bacterial RNAs, whereas noncoding RNA has predominated for the eukaryotic RNAs, although the terms functional RNA (fRNA) and small nonmessenger RNAs (snmRNA) also are used. Here, we refer to the RNAs as noncoding RNAs (ncRNAs). We also use some of the designations given to specific classes of eukaryotic RNAs, such as small nucleolar RNAs (snoRNAs) and microRNAs (miRNAs) based on their subcellular localization or size. We discuss examples of well-characterized regulatory RNAs; more comprehensive tabulations of regulatory RNA sequences and properties can be found at http://www.sanger.ac.uk/Software/Rfam/ (6), http://biobases.ibch.poznan.pl/ncRNA/ (7), http://jsm-research.imb.uq.edu.au/rnadb, http://bioinfo.md.huji.ac.il/marg/ (8), and http://dir2.nichd.nih.gov/nichd/cbmb/segr/segr.html.

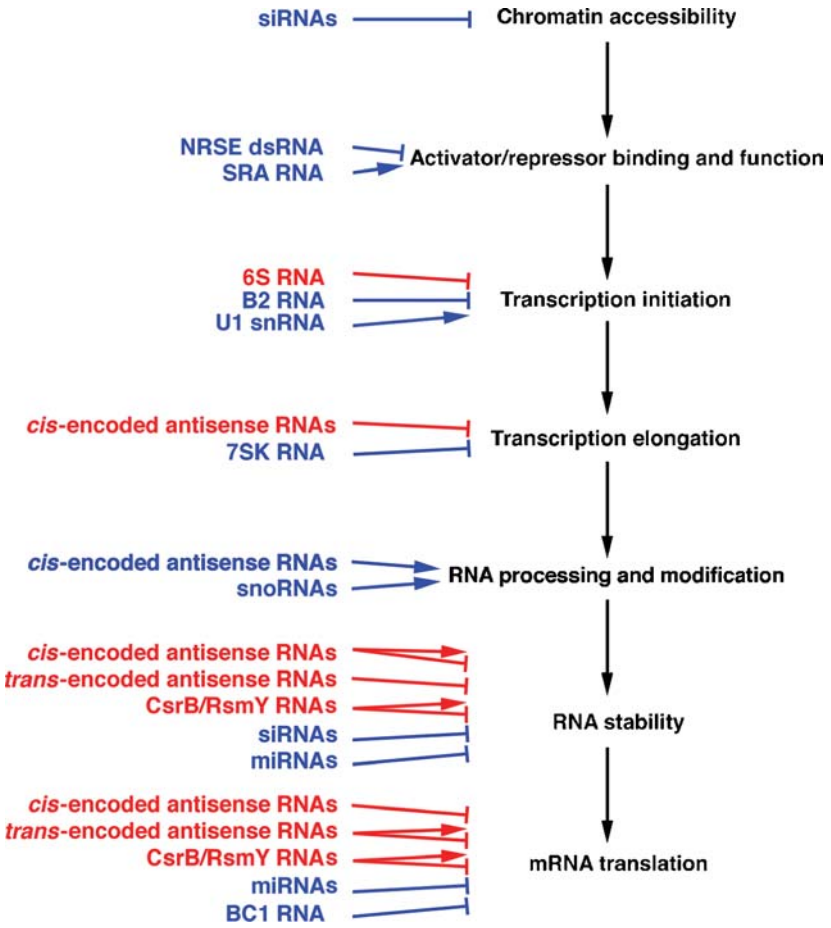

Figure 1 Steps in gene expression at which RNAs have been found to modulate gene expression. Bacterial and plasmid (*red*) and eukaryotic (*blue*) noncoding RNAs (ncRNAs) exert positive (*arrows*) and negative (*bars*) regulation at every step in the expression of a gene. Specific examples of *cis*- and *trans*-encoded antisense RNAs and microRNAs (miRNAs) that modulate transcription elongation, RNA processing or stability, or mRNA translation are described in the text. Abbreviations are siRNA, small interfering RNAs, and snoRNA, small nucleolar RNA.

BASE-PAIRING RNAs

Thus far, the most common mechanism by which ncRNAs regulate gene expression is by base pairing with target transcripts. The RNAs that act by base pairing can be grouped into two broad classes; *cis*-encoded RNAs that are encoded at the same genetic location, but on the opposite strand to the RNAs they act upon and therefore

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

contain perfect complementarity to their target, and *trans*-encoded RNAs that are encoded at a chromosomal location distinct from the RNAs they act upon and generally do not exhibit perfect base-pairing potential with their targets.

cis-Encoded Antisense RNAs

cis-encoded antisense RNAs were first discovered in bacterial plasmids, where the RNAs modulate the expression of genes involved in replication and stable plasmid inheritance. Subsequently, *cis*-encoded antisense RNAs also were found to be associated with some transposons and bacteriophages. There are fewer examples of chromosomal *cis*-encoded antisense RNAs with known functions, although expression studies suggest antisense transcripts can be detected for a significant percentage of protein-coding genes in bacterial and eukaryotic cells, raising the possibility that some of these transcripts may function as regulatory RNAs.

PLASMID-ENCODED ANTISENSE RNAs The plasmid-encoded ncRNAs have served as the paradigm for studying the functions of *cis*-encoded antisense RNAs [reviewed in (9)]. For most of the plasmid examples, the antisense RNAs base pair with and regulate the transcription, stability, or translation of mRNA-encoding proteins critical for replication or stable plasmid inheritance. In these cases, the antisense RNAs are constitutively synthesized but are metabolically unstable. As a consequence, changes in plasmid concentrations are reflected in changes in the levels of the antisense RNAs. For example, for the plasmid pT181, an increased copy number results in increased levels of the RNAI (∼85 nt) and RNAII (∼150 nt) antisense RNAs. These RNAs base pair with and stabilize a structure associated with transcription termination upstream of the coding sequence for RepC, a protein required for replication initiation (10). When plasmid copy numbers decrease, RNAI and RNAII levels decrease, allowing transcription readthrough, leading to increased RepC levels and renewed replication. Other *cis*-encoded antisense RNAs control so-called plasmid addiction systems, which ensure that plasmid-containing cells survive, whereas cells that do not contain plasmid are killed. A prototype of these toxin-antitoxin systems is the *hok-sok* genes of plasmid R1 [reviewed in (11)]. The *hok* (host killing) gene encodes a small protein that damages the bacterial membrane leading to cell death. In plasmid-containing cells, *hok* mRNA translation is repressed by the Sok RNA (suppressor of killer). When the plasmid is lost, the differential stability of the Sok and *hok* RNAs determines that the Sok RNA levels decrease faster than the *hok* mRNA levels, leading to Hok protein expression and cell death.

The nature of the base pairing that occurs between the plasmid-encoded antisense RNAs and their targets has been explored in detail. Although there are extensive regions of complementarity that can give rise to long uninterrupted stretches of base pairing, it has been found that full base pairing generally is not required for regulation by these RNAs (12). Instead, the most critical interactions are in the short single-stranded regions where the first base pairs bring the antisense RNA and

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

target together in what is termed the "kissing complex." Usually, the base-pairing interaction is then extended to form a more stable complex called the "extended kissing complex," which exerts the regulatory effects.

CHOROMOSOMALLY ENCODED ANTISENSE RNAs There are fewer known examples of distinct *cis*-encoded antisense RNAs expressed from bacterial genomes. Six regions of the *Escherichia coli* chromosome show homology to the *hok-sok* loci found on plasmids, and transcripts have been detected for a subset of these regions (13). However, it is not known whether the Sok-like RNAs are required to repress expression of toxic proteins encoded on the opposing strand or whether they have independent functions. Four long-directed-repeat (LDR) sequences in *E. coli* also express both an mRNA (*ldr*) encoding a toxic peptide and a *cis*-encoded antisense RNA (*rdl*) (14). In addition, there is an example of two ncRNAs encoded on opposite strands on the *E. coli* genome (275 nt RyeA/SraC and 100 nt RyeB) (15, 16). Although the functions of these two RNAs are not known, they show differential expression patterns. Although many of the bacterial *cis*-encoded antisense RNAs are likely to fit the toxin-antitoxin paradigm in which the ncRNA is required to maintain very low expression of the target mRNA, one RNA in *E. coli* (105 nt GadY/1S183) has been shown to increase the expression of the *cis*-encoded mRNA (*gadX*) by stabilization of the transcript (17). In this case, the 3′ end of the GadY ncRNA overlaps the 3′ end of the *gadX* mRNA. This overlap region is required for GadY dependent accumulation of *gadX* mRNA, leading to the model that GadY base pairing with the 3′-untranslated region (3′ UTR) of the *gadX* mRNA confers increased stability (17). Whether other *cis*-encoded RNAs fall into this category remains to be determined.

As will be described below, antisense transcripts have been detected for a sizable fraction of protein-coding genes in eukaryotic cells. However, with the exception of the siRNAs, only a very limited number of these antisense transcripts have been shown to exert a regulatory effect on the *cis*-encoded mRNA. One example is the Rev-ErbAα RNA, an antisense transcript detected in B lymphocytes. This antisense transcript overlaps one of two antagonistic sites in the mRNA encoding the thyroid hormone receptor, and expression of the Rev-ErbAα RNA correlates with a change in the ratio between the two splice forms of the thyroid hormone receptor transcript (18, 19).

siRNAS small interfering RNAs (siRNAs) comprise the one subclass of eukaryotic *cis*-encoded antisense RNAs, which has received extensive attention [reviewed in (20, 20a)]. These 21– ∼25 nucleotide (nt) RNA fragments usually are derived from double-stranded RNA of exogenous origin and are thought to be a defense against foreign RNA. Double-stranded RNA that has entered the cell is first cleaved into random 21–25 nt double-stranded fragments by a complex of proteins containing Dicer, an RNase III-type endonuclease specific for double-stranded RNA. Another protein complex, RNA-induced silencing complex (RISC), then separates the two RNA fragments and facilitates base pairing with a complementary RNA. This target RNA is subsequently cleaved within the complementary

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

sequence, most likely by Argonaute, a conserved component of the RISC complex. Although siRNAs are usually derived from foreign RNAs, there are examples in which sense and antisense transcripts derived from endogenous sequences, particularly repeated sequences, can "silence" the expression of the sense transcript by targeting the sense transcript for degradation or by modulating the structure of the chromosome encoding the RNAs [reviewed in (21, 21a)]. The details by which siRNAs modulate chromatin structure remain to be delineated but also require Dicer- and Argonaute-containing complexes and may involve DNA-RNA or RNA-RNA base pairing given that silencing is highly sequence specific.

trans-Encoded Antisense RNAs

In contrast to the relatively few chromosomal *cis*-encoded antisense RNAs known to have function, many chromosomal *trans*-encoded antisense RNAs have been found to exert regulation in bacteria and eukaryotes. Whether this discrepancy reflects the mechanistic requirements for antisense regulation or is simply a consequence of the RNAs that have been fortuitously characterized is not clear.

HFQ-BINDING RNAs IN BACTERIA The functions of more than a dozen *trans*-encoded base-pairing RNA regulators have been characterized in *E. coli* [reviewed in (22)]. These RNAs have been shown to destabilize mRNAs and to either repress or activate translation and are defined by the fact that they bind to a common protein, Hfq. Examples of RNAs that block translation by base pairing with mRNA sequences adjacent to or overlapping the ribosome binding sites are MicC RNA (109 nt) (23) and MicF RNA (93 nt) [reviewed in (24)], which repress the translation of the OmpC and OmpF outer membrane porins, respectively. Because MicC and MicF are reciprocally expressed under a variety of environmental conditions, the RNAs contribute to the reciprocal expression of the two porin proteins. Another example of an RNA that represses translation is Spot42 RNA (109 nt); its expression is modulated in response to different carbon sources. This RNA base pairs with sequences internal to the *galETKM* mRNA, resulting in differential expression of different genes within the operon (25). Two RNAs, DsrA RNA (85 nt) and RprA RNA (105 nt), have been shown to promote translation of the *rpoS* mRNA by preventing the formation of an inhibitory secondary structure that normally occludes the ribosome binding site within the long *rpoS* transcript (26–29). Examples of ncRNAs that affect mRNA stability are the RyhB/SraI RNA (90 nt), which is induced under conditions of low iron (30) and the SgrS/RyaA RNA (∼200 nt), which is induced by elevated phosphosugar levels (31). Upon iron starvation, RyhB RNA base pairs with and promotes the degradation of transcripts encoding iron-containing enzymes, thus allowing alternative utilization of the limited iron. Elevated levels of SgrS RNA are associated with decreased levels of the mRNA encoding the major glucose transporter (*ptsG*). The functions of these regulatory RNAs were elucidated using knowledge of the conditions under which the RNAs were expressed, defining phenotypes associated with overexpression of the RNAs, and testing predictions of possible base-pairing targets.

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

Although *trans*-encoded antisense RNAs were first described in *E. coli*, several similar RNAs have recently been discovered in other bacteria. Among these are two ncRNAs (PrrF1 and PrrF2); their expression is induced by low iron in *Pseudomonas aeruginosa*. Although these RNAs are induced under the same conditions as the *E. coli* RyhB RNA and also target mRNAs encoding iron-containing proteins (32), there is not recognizable primary sequence similarity to RyhB, indicating that the RyhB and PrrF RNAs either evolved separately or have diverged significantly. It is intriguing that two ncRNAs in *P. aeruginosa* apparently carry out the same function as a single RNA in *E. coli*. In a similar vein, four homologous RNAs (Qrr1, Qrr2, Qrr3 and Qrr4) in *Vibrio harveyi* and *Vibrio cholerae* are required to destabilize the mRNA encoding a key regulator of the quorum-sensing response in which the bacteria monitor their population density (33). The reason for multiple homologs is not known, although the presence of multiple RNAs should allow for more nuanced regulation.

Regulation by all of the *trans*-encoded antisense RNAs characterized thus far in *E. coli* requires the RNA chaperone protein Hfq. This protein is a homolog of the Sm and Sm-like proteins that form the core of splicing and mRNA degradation complexes in eukaryotic and archaeal cells (34–36). Like the Sm and Sm-like proteins, Hfq binds AU-rich sequences and forms a homo-hexameric ring with dimensions similar to the heptameric and hexameric rings formed by the eukaryotic and archaeal proteins. Hfq binding to the bacterial RNAs has been shown to result in various outcomes. For several RNAs, Hfq binding promotes base pairing between the ncRNA and its mRNA target. In some cases, Hfq binding leads to changes in ncRNA or mRNA accessibility to RNases, indicating that Hfq induces structural changes in these RNAs (34, 36, 37). In addition, Hfq binding can protect against digestion by the ribonuclease RNase E (38–40). The mechanisms by which the archael and eukaryotic Sm and Sm-like proteins contribute to splicing and mRNA degradation are not fully understood; however, it is conceivable that these proteins also promote RNA-RNA interactions and RNA structural changes similar to bacterial Hfq.

Several features of the bacterial *trans*-encoded antisense RNAs are worth noting. All of these RNAs are expressed as independent transcription units, and all are induced by specific environmental conditions, suggesting that one function of bacterial antisense RNA is to integrate different regulatory networks to allow optimal survival under unfavorable growth conditions. Many of the *trans*-encoded RNAs also have multiple targets, and some targets are regulated by multiple ncRNAs. Given that more than one third of all *E. coli* ncRNAs are bound by Hfq and are suspected to act by base pairing (40), *trans*-encoded antisense RNAs appear to be an abundant class of regulatory molecules in bacteria.

EUKARYOTIC miRNAs The *trans*-encoded miRNAs found in worms, flies, plants, and vertebrates base pair with mRNAs and modulate mRNA stability and translation, analogous to many of the bacterial *trans*-encoded antisense RNAs. These ∼22 nt miRNAs are cleaved from longer stem-loop RNAs by the same enzyme

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

complexes that generate the siRNAs [reviewed in (41–43)]. However, a notable difference between siRNAs and miRNAs is that miRNAs have strand specificity and defined ends, whereas siRNAs can be generated anywhere along the length of a double-stranded RNA. The first two miRNAs to be discovered, *Caenorhabditis elegans lin-4* and *let-7*, have been shown to base pair with complementary sequences at the 3′ ends of the target mRNAs, *lin-14* and *lin-41*, respectively (44–46). The base-pairing interactions block productive translation of the target in a manner that remains to be mechanistically determined. For several plant miRNAs and at least one mammalian miRNA, base pairing with the target mRNAs has been found to lead to degradation of the targets (47–49). Unlike the bacterial RNAs, which base pair with one or two complementary sequences in their target mRNAs, *lin-4*, *let-7*, and other miRNAs can base pair with multiple repeats of the complementary sequence found in the target mRNAs. The *lin-4* and *let-7* RNAs regulate the timing of *C. elegans* larval development, and many of the hundreds of other recently discovered miRNAs also appear to have regulatory functions in developmental pathways.

Representative base-pairing interactions between the bacterial *trans*-encoded antisense RNAs and eukaryotic miRNAs and their targets are shown in Figure 2. The rules as to what constitutes productive base pairing are just being elucidated. Systematic analysis of the base pairing between synthetic miRNAs and target sequences indicates that the ability of the miRNA to repress translation is strongly dependent on the first 8 nts at the 5′ end of the miRNA (50). A few algorithms designed to identify targets for miRNAs have been developed (51–54). Many of the predicted targets need to be tested, and it is clear that full representation of miRNA targets has not been achieved. The consequences of base pairing also remain to be understood. In general, the perfect or near-perfect base pairing associated with siRNAs and some miRNAs leads to RNA degradation, whereas the imperfect base pairing associated with other miRNAs leads to repression of translation. However, it is not known what other factors contribute to the outcome of ncRNA-target-RNA pairing. It also is possible that base pairing between some *trans*-encoded RNAs and their targets can affect both mRNA stability and translation. Another problem that deserves further attention is the question of how *trans*-encoded RNAs are turned over once the environmental or developmental signal leading to their induction is removed. In the case of the *E. coli* RyhB RNA, the degradation of regulatory RNA and its mRNA targets has been shown to be coupled (55).

EUKARYOTIC snoRNAs The majority of snoRNAs, which target RNAs for methylation or pseudouridylation via base pairing, modify rRNAs and are expressed constitutively [reviewed in (56)]. However, a subset of the snoRNAs have been found to be expressed in a tissue-specific manner. For example, several snoRNAs encoded in the chromosomal region affected in the Prader-Willi/Angelman syndrome are only expressed from the paternal chromosome and are highly abundant in brain tissue (57–59). Because these snoRNAs do not show complementarity to

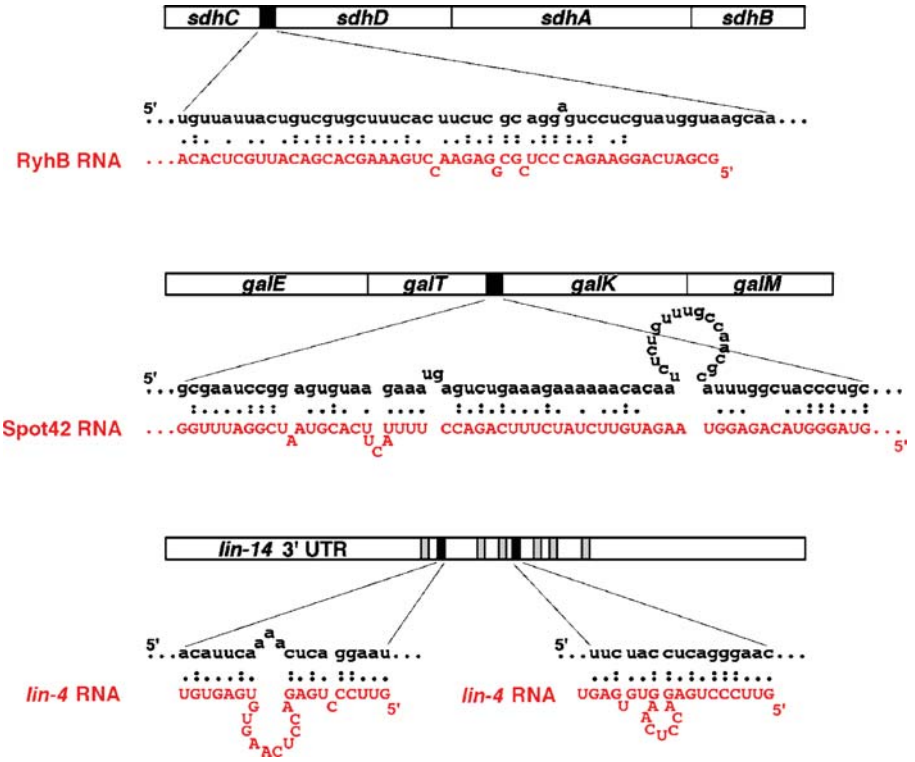

Figure 2 Similarities between *trans*-encoded base-pairing RNAs in bacteria and eukaryotes. The ncRNA sequences are given in red, capital letters, and the sequences of the target mRNA are given in black, lower case letters. Base pairing between the *E. coli* RyhB RNA and *sdhCDAB* mRNA leads to *sdhCDAB* mRNA degradation. Base pairing between the *E. coli* Spot42 RNA and *galETKM* mRNA leads to repression of *galK* translation. The *C. elegans lin-4* miRNA is able to base pair with seven different regions, shaded in gray or black, within the 3′ UTR of the *lin-14* mRNA. Base pairing between the *lin-4* miRNA and *lin-14* mRNA also leads to decreased translation. Regions of potential base pairing are shown; it is not clear whether complete base pairing is needed for regulation.

rRNAs, it is thought that they may regulate the expression of specific mRNAs by directing their methylation or pseudouridylation.

RNAs THAT MODIFY PROTEIN ACTIVITY

Not all regulatory RNAs act by base pairing. In recent years several regulatory RNAs that bind to proteins and modify the activities of these proteins have been characterized. Although far fewer of this type of ncRNA are known at this time,

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

it is likely that others will be found among the multitude of newly discovered ncRNAs.

RNAs that Modulate Transcription

A number of different bacterial and eukaryotic RNAs have been shown to bind to and modulate the activities of proteins that impact transcription. Interestingly, several of these regulatory RNAs are very abundant and thus were among the first ncRNAs to be discovered; yet their cellular roles have only been uncovered recently. Although the precise mechanisms of action have not been elucidated for these RNAs, it has been postulated that at least a subset of these RNAs act by mimicking nucleic acid interactions normally made by the target protein.

BACTERIAL 6S RNA The *E. coli* 6S RNA (184 nt) was among the first ncRNAs to be detected and sequenced, in part owing to its high abundance. However, the function of the RNA long remained unknown because of the lack of obvious defects in deletion strains during exponential growth. The RNA was known to be in an RNA-protein complex (60), and a breakthrough in understanding the 6S RNA function came from coimmunoprecipitation and UV cross-linking experiments that showed 6S RNA directly binds to the housekeeping form of RNA polymerase (σ^{70}-RNA polymerase) (61). 6S RNA levels are elevated in response to decreased nutrients as cells enter stationary phase, and 6S RNA is required for optimal long-term survival (62). 6S RNA interactions with RNA polymerase repress σ^{70}-dependent transcription during the stationary phase, although, interestingly, only a subset of σ^{70} promoters is affected. The predicted 6S RNA secondary structure is largely double-stranded with a central single-stranded bulge that is required for 6S RNA function. This structure is reminiscent of the open conformation of the promoter DNA formed during transcription initiation. One model for 6S RNA activity is that it binds RNA polymerase similarly to DNA, thus acting as a DNA mimic and a competitive inhibitor for transcription.

MOUSE B2 RNA The B2 RNA (178 nt) is expressed by RNA polymerase III from short interspersed repetitive elements (SINES) in the mouse genome. Expression of this RNA is increased up to 100-fold in response to environmental stresses such as heat shock. The knowledge that RNA polymerase III activity is required for heat shock-induced inhibition of RNA polymerase II prompted experiments to examine the effects of B2 RNA on RNA polymerase II-directed transcription (63). Coimmunoprecipitation and binding experiments provided evidence that B2 binds to RNA polymerase II upon heat shock, and in vivo and in vitro transcription experiments revealed that B2 RNA inhibits RNA polymerase II by preventing the formation of active preinitiation complexes (63, 64). The RNA polymerase II-dependent promoters of heat shock genes are not inhibited, indicating promoter specificity. As for the 6S RNA, the basis of this specificity is not known.

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

MAMMALIAN 7SK RNA The mammalian 7SK RNA (330 nt) also was among the first ncRNAs to be identified. Similar to the bacterial 6S RNA, 7SK RNA was known to be contained within a ribonucleoprotein complex, but its cellular role was unclear (65). 7SK function was revealed by the discovery that the RNA bound to and inhibited the function of the transcription elongation factor P-TEFb, which consists of a CDK9-cyclin T1 heterodimer (66, 67). Later studies showed that 7SK actually binds to a protein denoted HEXIM1/MAQ1, which brings 7SK to P-TEFb, thus inhibiting the activity of the elongation factor (68–71). A variety of stress conditions lead to 7SK RNA release from the complex, resulting in P-TEFb activation, although the mechanism controlling the RNA release has not been elucidated. Interestingly, the HEXIM1/MAQ1 domain required for 7SK RNA binding is highly homologous to the RNA-binding motif of the HIV-1 TAR protein (71).

HUMAN U1 snRNA Given the intimate link between mRNA transcription and processing in eukaryotic cells, it is perhaps not surprising that the U1 small nuclear RNA (snRNA), a core component of the splicesome, also binds to the general transcription factor TFIIH (72). Assays of transcription in reconstituted systems showed that U1 snRNA stimulates transcription initiation, but the details of this stimulatory effect need to be elucidated. Other components of the splicesome have been suggested to stimulate transcription elongation (73).

HUMAN SRA RNA The steroid receptor RNA activator (SRA) (700–850 nt) was first identified in a screen for cofactors of the steroid hormone receptors (74). Additional experiments then showed that the SRA RNA acts as a coactivator to stimulate the transcription of steroid receptor-dependent genes. Mutational studies suggest that a variety of domains are required for the coactivation function of SRA RNA (75), but the manner in which this RNA interacts with the steroid hormone receptor and modifies its activity is not understood.

NEURONAL NRSE dsRNA Another RNA that modulates transcription is the NRSE dsRNA, a small, double-stranded RNA. This RNA stood out in a screen of 20 to 40 nt RNAs from adult hippocampal neural stem cells because it contained a match to the 21 nt sequence bound by the NRSF/REST protein, a negative transcriptional regulator that restricts neuronal gene expression to neurons (76). Northern analysis revealed that 20 nt RNAs corresponding to both strands of this sequence were expressed. Mobility shift experiments then showed that the NRSF/REST protein binds to the double-stranded RNA. In addition, the NRSE dsRNA was found to antagonize the effects of the NRSF/REST repressor in overexpression experiments. Given the sequence similarity to the NRSF/REST DNA-binding site, an attractive hypothesis is that the NRSE dsRNA competes for NRSF/REST binding to the promoter sequences. However, chromatin immunoprecipitation experiments indicate that the NRSF/REST protein is stably bound to DNA in both the presence and absence of the NRSE dsRNA, indicating that other models need to be considered.

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

RNAs that Modulate mRNA Stability and Translation

A few bacterial and eukaryotic RNAs also have been found to bind to and modify the activities of proteins that regulate mRNA stability and translation. Again some of these regulatory RNAs appear to be acting by mimicking the structures of other nucleic acids, in this case other RNAs.

BACTERIAL CSRB/RSMY FAMILY OF RNAs A family of homologous RNA-binding proteins, including CsrA of *E. coli* and RsmA of *Pseudomonas* species, have been shown to play a role in regulating a variety of processes in the bacterial cell, including glycogen biosynthesis, flagellar motility, and biofilm formation [reviewed in (77)]. These proteins bind to the 5′ regions of the target mRNAs, blocking translation initiation and stimulating mRNA decay or, in other cases, stimulating translation and blocking mRNA decay (78). A family of RNAs, of which the CsrB (360 nt) and CsrC RNAs (270 nt) of *E. coli* and RsmY (118 nt) and RsmZ RNAs (127 nt) of *Pseudomonas fluorescens* are representative, blocks the actions of the CsrA and RsmA proteins (79, 80). Each RNA contains multiple sequences that are similar to the mRNA sequences bound by the CsrA and RsmA proteins, and it has been found that ∼18 CsrA molecules are in a ribonucleoprotein complex with the *E. coli* CsrB RNA (81). Thus, current models postulate that the CsrB/RsmY family of regulatory RNAs modulates mRNA stability and translation by acting as RNA mimics, sequestering multiple copies of the CsrA and RsmA proteins and blocking their functions.

DENDRITIC BC1 RNA The FMRP protein is an RNA-binding protein that is highly expressed in brain. Mutations associated with the absence of the FMRP protein or altered FMRP lead to the fragile X syndrome, the most frequent cause of inherited mental retardation. Similar to the bacterial CsrA and RsmA proteins, FMRP appears to repress the translation of target mRNAs. In addition to binding to the target mRNAs, FMRP has been found to bind the BC1 RNA (∼150 nt) (82), an RNA transcribed by RNA polymerase III in specific neuronal cells (83). Unlike the inhibitory effect exerted by the CsrB/RsmY family of RNAs, BC1 appears to promote the interaction between FMRP and its target mRNAs, possibly via base-pairing interactions.

SYSTEMATIC APPROACHES FOR IDENTIFYING ncRNAs

Until five years ago, the genes encoding regulatory RNAs were largely overlooked, primarily because they were hard to detect de novo. New ncRNA genes generally have not been identifiable during genome sequence analysis owing to their lack of defined sequence features, and thus they are not included in genome annotations. Even recently identified ncRNA genes are often not annotated because of the inability to identify the 5′ and 3′ ends by sequence gazing, unlike protein-coding genes for which start and stop codons are more easily identified. RNA genes also

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

are missed in genetic studies as a consequence of their small size and because they are resistant to frameshift and nonsense mutations, which only apply to protein-coding genes. In addition, RNAs are often missed in biochemical experiments designed to assay proteins. Thus, it is of value to discuss the methods that have recently been developed to identify ncRNAs and their genes. It should be noted that many features of ncRNAs, such as the ability to form secondary structures, also can be found in many regulatory 5′ and 3′ untranslated regions (UTRs) of mRNAs. Therefore, all predictions of ncRNA genes need to be confirmed by direct detection of these transcripts, for instance by Northern analysis.

Computational Approaches

Computational approaches have been quite successful in finding families of ncRNAs with well-defined sequence elements or characteristics, such as the C/D box family of snoRNAs (84, 85). Few regulatory RNA families contain such defined elements. Even so, a number of computational approaches have successfully predicted the presence of additional ncRNA genes. Approaches based on straight sequence conservation between related species in nonprotein-coding regions of the genome (intergenic regions), alone and in combination with algorithms based on other criteria listed below, have been very successful in predicting ncRNA genes in *E. coli* (15, 86), *C. elegans* (87), and *Arabidopsis thaliana* (51). Other searches based on predictions of RNA structure conservation in intergenic regions led to the identification of regulatory RNAs in *E. coli* (88), *Saccharomyces cerevisiae* (89), *C. elegans* (87), and *A. thaliana* (51). The presence of binding sites for specific DNA-binding proteins as well as promoter and terminator sequences in the intergenic regions was another criteria used to predict possible RNA genes in *E. coli* (90, 86), *P. aeruginosa* (32), *V. harveyi*, *V. cholerae* (33), and *S. cerevisiae* (91). The detection of GC-rich regions in the AT-rich genomes of *Methanococcus jannaschii* and *Pyrococcus furiosus* led to the identification of ncRNAs (92). A few *S. cerevisiae* ncRNAs also were detected in unusually long intergenic regions (>2 kb) (91). Finally, others have extracted features of known ncRNAs using a machine learning approach to search for other ncRNAs in *E. coli* (93). However, these last predicted ncRNA genes have not been experimentally tested for expression.

The computational approaches have led to the identification of many ncRNAs in bacteria, yeast, *C. elegans*, and *A. thaliana*, but they are limited because they have focused on the intergenic regions and thus have missed ncRNAs encoded within protein-coding regions of the genome, such as expected for *cis*-encoded antisense RNAs. Most of the computational approaches also rely heavily on sequence conservation and therefore overlook ncRNAs that are species specific or are less well conserved.

Direct Detection

The first ncRNAs were identified directly by size fractionation of total RNA labeled in vivo (94–96). This approach was successful in the detection of very abundant

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

RNAs. More recently, a number of groups have used direct cloning after size selection. Brosius (97) and Hüttenhofer (16) and colleagues have employed a strategy they have termed "RNomics" to identify ncRNAs in *E. coli* (16), *Archaeoglobus fulgidus* (97), *A. thaliana* (98), *Drosophila melanogaster* (99), and mouse (100). In this approach, total RNA is size fractionated, and all RNAs between 50 and 500 nts are cloned. Hybridization was used to identify the clones of known RNAs. Clones showing hydridization signals below a specific threshold were considered novel and then sequenced. Another direct cloning approach, based on the isolation of RNAs of a smaller size, was used in the first large scale screens for miRNAs in *D. melanogaster*, *C. elegans*, *A. thaliana*, and different human and mouse tissues (101–105).

Several ncRNAs also have been identified on the basis of their association with RNA-binding proteins. In early screens, RNAs that coimmunoprecipitate with Sm proteins were sequenced directly (106). More recently, RNAs that coimmunoprecipitate with *E. coli* Hfq (40), as well as with Lhp1p, the *S. cerevisiae* La protein (107), were detected on microarrays.

Finally, microarray expression studies as well as the analysis of libraries of mRNAs and expressed sequence tags (EST) have allowed the detection of antisense transcripts and of transcripts outside the known transcription units, some of which are likely to be ncRNAs [reviewed in (108)]. For example, Selinger et al. (109) reported expression from the antisense strand of more than half of the predicted *E. coli* ORFs, although it is possible that a subset of the detected signals are due to cross-hybridization. Analyses of human sequence databases have led to predictions of greater than 1600 antisense RNAs (110–112).

Despite the many screens that have been carried out in recent years, the total number of ncRNAs is not known for any organism. The number of predicted ncRNAs for any given organism varies widely, ranging from a few 10s to 100s in bacteria and from several 100s to 1000s in mammalian cells.

CONCLUDING REMARKS

Although there has been significant progress in elucidating the functions of base-pairing RNAs and in identifying protein targets of other ncRNAs, there are still many ncRNAs for which the cellular roles are not clear. Notable among these RNAs are the ∼100,000 nt Air RNA and 16,500 nt Xist, associated with autosomal gene imprinting and X-chromosome inactivation [reviewed in (113)], as well as the ∼70–110 nt Y RNAs, which have been shown to be associated with the Ro RNA-binding protein in a variety of vertebrates as well as the radiation- and drought-resistant bacterium *Deinococcus radiodurans* (114, 115).

A relevant question is whether all of the RNAs that have been detected have function. Antisense RNAs and RNAs processed from the 5′ or 3′ ends of transcripts or intronic sequences may exist as distinct entities in the cell yet still not have intrinsic functions. Mattick (116) has proposed that ncRNAs, derived from

the many intergenic and intronic regions found in the genomes of higher organism, represent a fundamental advance in the genetic operating system of these organisms; however, this hypothesis ignores the prevalence of ncRNAs in bacterial cells and the fact that only a very small fraction of the intronic and intergenic RNAs in eukaryotic cells have been shown to be functional. It is possible that the process of transcription itself has a regulatory effect and that some of the detected RNAs are just a secondary consequence of this mode of regulation. Transcription from an upstream promoter could block RNA polymerase binding to a downstream promoter. This has been proposed for the ~550 nt SRG-1 transcript expressed upstream of the *S. cerevisiae SER3* gene, which encodes a phosphoglycerate dehydrogenase (117). When the full length *SRG-1* RNA or heterologous derivatives of the RNA are expressed in *cis* to the *SER3* gene, expression of *SER3* is repressed, leading to the suggestion that transcription across the *SER3* promoter interferes with the binding of activators resulting in repression of the *SER3* promoter. Transcription of an antisense RNA also has in some, but not all cases, resulted in decreased expression of the opposing genes (118).

Other open questions are whether some RNAs act by more than one mechanism and whether any RNAs with intrinsic functions also encode proteins. A recent report suggests that the SRA RNA encodes a peptide (119). Until we have a more complete understanding of the unique properties of functional ncRNAs, there is a need to be cautious about interpreting whether an RNA is a ncRNA, mRNA, or solely a by-product of transcription.

Finally, it is worth considering the advantages of RNA regulators over protein regulators. Many of the characterized RNA regulators are expressed and function during specific developmental stages or under stress conditions. Noncoding RNA regulators may be particularly advantageous under these conditions when resources are limited, given the lower input of energy and shorter time required to synthesize a short ncRNA compared to a protein. In addition, many ncRNAs act at a posttranscriptional level, which also would ensure a fast response to a developmental cue or environmental signal. It also is conceivable there are evolutionary advantages to RNA regulators; *cis*-encoded antisense RNAs coevolve with their target RNAs, whereas the limited base pairing needed for the interactions between *trans*-encoded RNAs and their targets may allow for flexibility in evolving RNA regulators of new target genes.

Given the many open questions together with the possibility that new activities may yet be discovered for ncRNAs, studies of these RNA regulators promise to be an exciting area of research for many years to come.

ACKNOWLEDGMENTS

We thank S. Gottesman, N. Majdalani, D. Rio, and members of the Storz lab for comments regarding the review. We also apologize to those whose work was not cited because of space limitations. GS and SA acknowledge support from the United States-Israel Binational Science Foundation.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Burge CB, Tuschl TH, Sharp PA. 1999. In *The RNA World*, ed. RF Gesteland, TR Cech, JF Atkins, pp. 525–60. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 2nd ed.
2. Koch HG, Moser M, Muller M. 2003. *Rev. Physiol. Biochem. Pharmacol.* 146:55–94
3. Chen JL, Greider CW. 2004. *Trends Biochem. Sci.* 29:183–92
4. Hsieh J, Andrews AJ, Fierke CA. 2004. *Biopolymers* 73:79–89
5. Withey JH, Friedman DI. 2003. *Annu. Rev. Microbiol.* 57:101–23
6. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. *Nucleic Acids Res.* 31:439–41
7. Szymanski M, Erdmann VA, Barciszewski J. 2003. *Nucleic Acids Res.* 31: 429–31
8. Hershberg R, Altuvia S, Margalit H. 2003. *Nucleic Acids Res.* 31:1813–20
9. Wagner EG, Altuvia S, Romby P. 2002. *Adv. Genet.* 46:361–98
10. Brantl S, Wagner EG. 2000. *Mol. Microbiol.* 35:1469–82
11. Gerdes K, Gultyaev AP, Franch T, Pedersen K, Mikkelsen ND. 1997. *Annu. Rev. Genet.* 31:1–31
12. Malmgren C, Wagner EG, Ehresmann C, Ehresmann B, Romby P. 1997. *J. Biol. Chem.* 272:12508–12
13. Pedersen K, Gerdes K. 1999. *Mol. Microbiol.* 32:1090–102
14. Kawano M, Oshima T, Kasai H, Mori H. 2002. *Mol. Microbiol.* 45:333–49
15. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. *Genes Dev.* 15:1637–51
16. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jäger JG, et al. 2003. *Nucleic Acids Res.* 31:6435–43
17. Opdyke JA, Kang J-G, Storz G. 2004. *J. Bacteriol.* 186:6698–705
18. Hastings ML, Ingle HA, Lazar MA, Munroe SH. 2000. *J. Biol. Chem.* 275: 11507–13
19. Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH. 1997. *Nucleic Acids Res.* 25:4296–300
20. Montgomery MK. 2004. *Methods Mol. Biol.* 265:3–21

20a. Meister G, Tuschl T. 2004. *Nature* 431: 343–49

21. Elgin SCR, Grewal SIS. 2003. *Curr. Biol.* 13:R895–98

21a. Lippman Z, Martienssen R. 2004. *Nature* 431:364–70

22. Gottesman S. 2004. *Annu. Rev. Microbiol.* 58:303–28
23. Chen S, Zhang A, Blyn LB, Storz G. 2004. *J. Bacteriol.* 186:6689–97
24. Delihas N, Forst S. 2001. *J. Mol. Biol.* 313:1–12
25. Møller T, Franch T, Udesen C, Gerdes K, Valentin-Hansen P. 2002. *Genes Dev.* 16:1696–706
26. Lease RA, Cusick ME, Belfort M. 1998. *Proc. Natl. Acad. Sci. USA* 95:12456–61
27. Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S. 1998. *Proc. Natl. Acad. Sci. USA* 95:12462–67
28. Majdalani N, Chen S, Murrow J, John KS, Gottesman S. 2001. *Mol. Microbiol.* 39:1382–94
29. Majdalani N, Hernandez D, Gottesman S. 2002. *Mol. Microbiol.* 46:813–26
30. Massé E, Gottesman S. 2002. *Proc. Natl. Acad. Sci. USA* 99:4620–25
31. Vanderpool CK, Gottesman S. 2004. *Mol. Microbiol.* 54:1076–89
32. Wilderman PJ, Sowa NA, FitzGerald DJ, FitzGerald PC, Gottesman S, et al. 2004. *Proc. Natl. Acad. Sci. USA* 101:9792–97

33. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. 2004. *Cell* 118:69–82
34. Møller T, Franch T, Højrup P, Keene DR, Bächinger HP, et al. 2002. *Mol. Cell* 9:23–30
35. Schumacher MA, Pearson RF, Møller T, Valentin-Hansen P, Brennan RG. 2002. *EMBO J.* 21:3546–56
36. Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G. 2002. *Mol. Cell* 9:11–22
37. Geissmann TA, Touati D. 2004. *EMBO J.* 23:396–405
38. Folichon M, Arluison V, Pellegrini O, Huntzinger E, Régnier P, Hajnsdorf E. 2003. *Nucleic Acids Res.* 31:7302–10
39. Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Bläsi U. 2003. *RNA* 9: 1308–14
40. Zhang A, Wassarman KM, Rosenow C, Tjaden BC, Storz G, Gottesman S. 2003. *Mol. Microbiol.* 50:1111–24
41. Bartel DP. 2004. *Cell* 116:281–97
42. Nelson P, Kiriakidou M, Sharma A, Maniataki E, Mourelatos Z. 2003. *Trends Biochem. Sci.* 28:534–40
43. He L, Hannon GJ. 2004. *Nat. Rev. Genet.* 5:522–31
44. Lee RC, Feinbaum RL, Ambros V. 1993. *Cell* 75:843–54
45. Wightman B, Ha I, Ruvkun G. 1993. *Cell* 75:855–62
46. Vella MC, Choi EY, Lin SY, Reinert K, Slack FJ. 2004. *Genes Dev.* 18:132–37
47. Hutvagner G, Zamore PD. 2002. *Science* 297:2056–60
48. Llave C, Xie ZX, Kasschau KD, Carrington JC. 2002. *Science* 297:2053–56
49. Yekta S, Shih IH, Bartel DP. 2004. *Science* 304:594–96
50. Doench JG, Sharp PA. 2004. *Genes Dev.* 18:504–11
51. Jones-Rhoades MW, Bartel DP. 2004. *Mol. Cell* 14:787–99
52. Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB. 2003. *Cell* 115: 787–98
53. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. 2002. *Cell* 110:513–20
54. Stark A, Brennecke J, Russell RB, Cohen SM. 2003. *PLoS Biol.* 1:E60
55. Massé E, Escorcia FE, Gottesman S. 2003. *Genes Dev.* 17:2374–83
56. Kiss T. 2002. *Cell* 109:145–48
57. Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, et al. 2000. *Proc. Natl. Acad. Sci. USA* 97:14311–16
58. de los Santos T, Schweizer J, Rees CA, Francke U. 2000. *Am. J. Hum. Genet.* 67:1067–82
59. Meguro M, Mitsuya K, Nomura N, Kohda M, Kashiwagi A, et al. 2001. *Hum. Mol. Genet.* 10:383–94
60. Lee SY, Bailey SC, Apirion D. 1978. *J. Bacteriol.* 133:1015–23
61. Wassarman KM, Storz G. 2000. *Cell* 101:613–23
62. Trotochaud AE, Wassarman KM. 2004. *J. Bacteriol.* 186:4978–85
63. Allen TA, Von Kaenel S, Goodrich JA, Kugel JF. 2004. *Nat. Struct. Biol.* 11:816–21
64. Espinoza CA, Allen TA, Hieb AR, Kugel JF, Goodrich JA. 2004. *Nat. Struct. Biol.* 11:822–29
65. Ullu E, Esposito V, Melli M. 1982. *J. Mol. Biol.* 161:195–201
66. Nguyen VT, Kiss T, Michels AA, Bensaude O. 2001. *Nature* 414:322–25
67. Yang ZY, Zhu QW, Luo KX, Zhou Q. 2001. *Nature* 414:317–22
68. Michels AA, Nguyen VT, Fraldi A, Labas V, Edwards M, et al. 2003. *Mol. Cell. Biol.* 23:4859–69
69. Michels AA, Fraldi A, Li Q, Adamson TE, Bonnet F, et al. 2004. *EMBO J.* 23:2608–19
70. Yik JHN, Chen RC, Nishimura R, Jennings JL, Link AJ, Zhou Q. 2003. *Mol. Cell* 12:971–82
71. Yik JHN, Chen RC, Pezda AC, Samford CS, Zhou Q. 2004. *Mol. Cell. Biol.* 24:5094–105

11 Feb 2005 11:40 AR AR261-BI74-08.tex XMLPublish^{SM}(2004/02/24) P1: JRX
AR REVIEWS IN ADVANCE10.1146/annurev.biochem.74.082803.133136

72. Kwek KY, Murphy S, Furger A, Thomas B, O'Gorman W, et al. 2002. *Nat. Struct. Biol.* 9:800–5
73. Fong YW, Zhou Q. 2001. *Nature* 414: 929–33
74. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong JM, et al. 1999. *Cell* 97:17–27
75. Lanz RB, Razani B, Goldberg AD, O'Malley BW. 2002. *Proc. Natl. Acad. Sci. USA* 99:16081–86
76. Kuwabara T, Hsieh J, Nakashima K, Taira K, Gage FH. 2004. *Cell* 116:779–93
77. Romeo T. 1998. *Mol. Microbiol.* 29: 1321–30
78. Dubey AK, Baker CS, Suzuki K, Jones AD, Pandit P, et al. 2003. *J. Bacteriol.* 185:4450–60
79. Valverde C, Heeb S, Keel C, Haas D. 2003. *Mol. Microbiol.* 50:1361–79
80. Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, et al. 2003. *Mol. Microbiol.* 48:657–70
81. Liu MY, Gui G, Wei B, Preston JF 3rd, Oakford L, et al. 1997. *J. Biol. Chem.* 272:17502–10
82. Zalfa F, Giorgi M, Primerano B, Moro A, Di Penta A, et al. 2003. *Cell* 112:317–27
83. Tiedge H, Fremeau RT Jr, Weinstock PH, Arancio O, Brosius J. 1991. *Proc. Natl. Acad. Sci. USA* 88:2093–97
84. Lowe TM, Eddy SR. 1999. *Science* 283:1168–71
85. Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP. 2000. *Science* 288:517–22
86. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EGH, et al. 2001. *Curr. Biol.* 11:941–50
87. Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, et al. 2003. *Mol. Cell* 11:1253–63
88. Rivas E, Klein RJ, Jones TA, Eddy SR. 2001. *Curr. Biol.* 11:1369–73
89. McCutcheon JP, Eddy SR. 2003. *Nucleic Acids Res.* 31:4119–28
90. Chen S, Lesnik EA, Hall AT, Sampath R, Griffey RH, et al. 2002. *BioSystems* 65:157–77
91. Olivas WM, Muhlrad D, Parker R. 1997. *Nucleic Acids Res.* 25:4619–25
92. Klein RJ, Misulovin Z, Eddy SR. 2002. *Proc. Natl. Acad. Sci. USA* 99:7542–47
93. Carter RJ, Dubchak I, Holbrook SR. 2001. *Nucleic Acids Res.* 29:3928–38
94. Goldstein J, Harewood K. 1969. *J. Mol. Biol.* 39:383–87
95. Hindley J. 1969. *J. Mol. Biol.* 30:125–36
96. Ikemura T, Dahlberg JE. 1973. *J. Biol. Chem.* 248:5024–32
97. Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, et al. 2002. *Proc. Natl. Acad. Sci. USA* 99:7536–41
98. Marker C, Zemann A, Terhorst T, Kiefmann M, Kastenmayer JP, et al. 2002. *Curr. Biol.* 12:2002–13
99. Yuan G, Klambt C, Bachellerie JP, Brosius J, Hüttenhofer A. 2003. *Nucleic Acids Res.* 31:2495–507
100. Hüttenhofer A, Kiefmann M, Meier-Ewert S, O'Brien J, Lehrach H, et al. 2001. *EMBO J.* 20:2943–53
101. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. *Science* 294:853–58
102. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T. 2002. *Curr. Biol.* 12:735–39
103. Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. *Science* 294:858–62
104. Lee RC, Ambros V. 2001. *Science* 294: 862–64
105. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. 2002. *Genes Dev.* 16:1616–26
106. Montzka KA, Steitz JA. 1988. *Proc. Natl. Acad. Sci. USA* 85:8885–89
107. Inada M, Guthrie C. 2004. *Proc. Natl. Acad. Sci. USA* 101:434–39
108. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. 2004. *Trends Biochem. Sci.* 29:88–94
109. Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, et al. 2000. *Nat. Biotechnol.* 18:1262–68
110. Lehner B, Williams G, Campbell RD,

Sanderson CM. 2002. *Trends Genet.* 18: 63–65
111. Shendure J, Church GM. 2002. *Genome Biol.* 3:1–14
112. Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, et al. 2003. *Nat. Biotechnol.* 21:379–86
113. Andersen AA, Panning B. 2003. *Curr. Opin. Cell Biol.* 15:281–89
114. Chen XG, Quinn AM, Wolin SL. 2000. *Genes Dev.* 14:777–82
115. Van Horn DJ, Eisenberg D, O'Brien CA, Wolin SL. 1995. *RNA* 1:293–303
116. Mattick JS. 2004. *Nat. Rev. Genet.* 5:316–23
117. Martens JA, Laprade L, Winston F. 2004. *Nature* 429:571–74
118. Farrell CM, Lukens LN. 1995. *J. Biol. Chem.* 270:3400–8
119. Chooniedass-Kothari S, Emberley E, Hamedani MK, Troup S, Wang X, et al. 2004. *FEBS Lett.* 566:43–47