Cell

Leading Edge
Minireview

The Long and Short of MicroRNA

Luke A. Yates,[1] Chris J. Norbury,[2] and Robert J.C. Gilbert[1,*]
[1]Division of Structural Biology, Wellcome Trust Centre for Human Genetics
[2]Sir William Dunn School of Pathology
University of Oxford, Oxford OX3 7BN, UK
*Correspondence: gilbert@strubi.ox.ac.uk
http://dx.doi.org/10.1016/j.cell.2013.04.003

MicroRNAs (miRNAs) are versatile regulators of gene expression in higher eukaryotes. In order to silence many different mRNAs in a precise manner, miRNA stability and efficacy is controlled by highly developed regulatory pathways and fine-tuning mechanisms both affecting miRNA processing and altering mature miRNA target specificity.

Introduction

The discovery in 1993 of a small endogenous regulatory RNA molecule in *C. elegans* paved the way for description of a large family of short (~22 nt) single-stranded ribonucleic acids termed microRNAs (miRNAs). These molecules are critical posttranscriptional regulators of gene expression in complex life. It is not surprising, therefore, that miRNAs are themselves tightly regulated to allow the shaping of gene expression in a temporally restrained and tissue-specific manner instrumental for properly structured organismal development and growth.

In this review, we focus on the regulatory processes in the cytoplasm controlling miRNA biosynthesis and target specificity through subtly tuned modifications in length at different stages in maturation.

Overview of Canonical miRNA Biogenesis

The generation of miRNAs is a multistage process (Figure 1). Briefly, the mature ~22 nt miRNA sequence is embedded in one strand of an ~33 bp double-stranded stem characteristic of hairpin structures in primary miRNA (pri-miRNA) transcripts produced by RNA polymerase II or III. The miRNA must therefore be excised during its biogenesis to elicit gene silencing; two endoribonucleolytic enzymes are responsible for this excision (Figure 1A). The nuclear endoribonuclease Drosha, in complex with the dsRNA-binding protein DGCR8 (DiGeorge critical region 8), is responsible for the first endonucleolytic reaction. DGCR8 (also known as Pasha) functions as a molecular ruler that positions the Drosha cut site 11 bp from the base of the hairpin stem. This releases an ~70 nt stem-loop precursor miRNA (pre-miRNA) that possesses a 3′ overhang. The pre-miRNA is exported to the cytoplasm via Exportin-5 (Exp5) in complex with Ran-GTP. Once exported, the pre-miRNA is processed by a second endoribonucleolytic reaction, catalyzed by Dicer, yielding an ~22 nt RNA duplex with protruding 3′ overhangs at both ends (Figure 1A). The duplex is loaded onto an Argonaute protein where one strand, complementary to the target mRNA (guide strand), is selected and subsequently forms the miRNA effector as part of a miRISC (miRNA-induced silencing complex), while the remaining strand (the "passenger strand") is released and degraded. Typically, miRNA-binding sites of animal mRNAs reside in their 3′ untranslated regions (UTRs) where recognition occurs via base pairing of the seed sequence (nucleotide positions 2 to 8) of a miRNA. The overall degree of miRNA:mRNA complementarity is considered a determinant of mRNA regulation, with miRNA-mediated gene silencing occurring through translation inhibition, mRNA deadenylation, and decay. However, the exact order of these events is the subject of ongoing debate.

Dicer as a Regulatory Node in the Cytoplasm

The two endoribonucleolytic cleavage steps represent the most obvious points at which functional miRNA production could be regulated. In the nucleus, Drosha-DGCR8-mediated processing of let-7 pri-miRNAs can be inhibited by Lin28B (Viswanathan et al., 2008), whereas the protein hnRNPA1 (heterogeneous nuclear ribonucleoprotein A1) binds specifically to pri-miR-18a to promote its processing (Guil and Cáceres, 2007). In the cytoplasm, Dicer catalyzes the second endonucelolytic reaction and is essential for miRNA maturation, with its knockdown in human cell lines resulting in aberrant miRNA production and the accumulation of pre-miRNA (Hutvágner et al., 2001). A number of recent reports have described regulatory mechanisms that impinge on Dicer processing. One such regulatory pathway modifies the pre-miRNA via 3′ end uridylylation; when fed into Dicer this can promote or inhibit its maturation (Heo et al., 2009, 2012, and 2008 referenced therein). Other regulatory mechanisms alter Dicer activity to give miRNA isoforms with altered target specificities (Fukunaga et al., 2012). Dicer-mediated processing requires sophisticated regulation and clearly represents another regulatory node.

Structural Basis of Dicing

Dicer is a large (1,922 amino acids in humans) multidomain protein comprised of an N-terminal domain, homologous to DExD/H-box helicases; DUF283 (a dsRNA-binding domain), "platform," and PAZ domain; a major unannotated region; the C-terminal catalytic core, that is the RNaseIII tandem (RNaseIIIa and RNaseIIIb); and a dsRNA-binding domain (dsRBD). Dicer cleaves at a site close to the terminal loop, or pre-element, of the pre-miRNA (~22 nt away from the base of the dsRNA stem) and introduces a staggered cut in its pre-miRNA substrate. The domain arrangement of human Dicer as determined

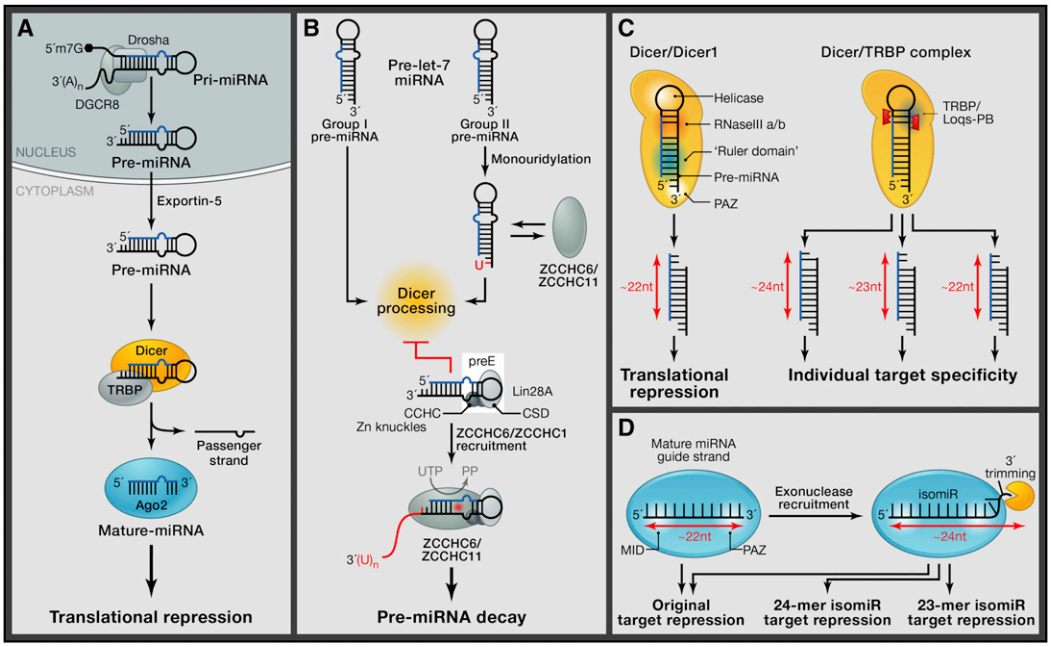

Figure 1. The Regulation of miRNA Function by Means of RNA Length

(A) Canonical maturation pathway of miRNA (see text for details).

(B) Modification of the 3′ end of pre-miRNA by the untemplated addition of uridines catalyzed by cytoplasmic TUTs. Mono-uridylylation facilitates Dicer processing, whereas Lin28 binding and oligo-uridylylation are inhibitory.

(C) Tuning of Dicer cut-site selection (shifted by 1–2 nt) by its binding partners allows the generation of specific isomiRs with altered target specificities.

(D) Exonucleolytic 3′ trimming of the Ago-loaded mature miRNA alters the length and thus target specificity.

by electron microscopy single-particle reconstruction designates the major unannotated region as a "ruler domain" between the "platform"/PAZ tandem and catalytic core (dsRBD and RNasIIIa/b tandem) thus providing an internal 22 nt gauge (Lau et al., 2012). Moreover, the positioning of helicase domains adjacent to the catalytic core is in accordance with data showing that the helicase of *Drosophila* Dicer-1 (Dcr-1) recognizes the single-stranded terminal loop at a proper distance from the 5′/3′ end, while mutants lacking the helicase domain could not distinguish between long-stem and small-loop pre-miRNA structural variants (Tsutsumi et al., 2011). High-throughput sequencing of Dicer-processed short-hairpin RNAs demonstrated recognition of a single-stranded RNA segment, either the terminal loop or internal bulge (where the loop forms a substructure, such as in pre-let-7 [see Nam et al., 2011]), 2 nt away from the cleavage site (Gu et al., 2012). However, earlier structural and biochemical analyses of human Dicer revealed a basic 5′ phosphate-binding pocket within the platform and the PAZ domain, conserved in higher eukaryotes, that is key to dicing precision (Park et al., 2011). Overall, Dicer measures ~22 nt products by anchoring the open terminus (5′/3′ end) at the platform and the PAZ domain, with the recognition of the terminal loop/bulge by the helicase domain playing a minor role in cleavage site selection. Thus, the complete domain organization of Dicer is crucial for recognition of cognate pre-miRNAs, allowing accurate positioning of the catalytic core and allosteric regulation of Dicer by its substrates and/or binding partners.

Tuning Dicer to Alternatively Dice

Dicer does not function alone but in association with additional proteins (Figure 1C). Mammals produce two related Dicer partners, TRBP (transactivating response RNA-binding protein), and PACT (protein activator of PKR), which contain three dsRNA-binding domains (dsRBD). Biochemical mapping of the interaction between Dicer and TRBP or PACT revealed that both proteins associate via their C-terminal dsRBD with the helicase domains of Dicer (Lee et al., 2006) and close to its catalytic core, suggesting that they may affect its activity. Indeed, two recent papers (Lee and Doudna, 2012; Fukunaga et al., 2012) have revealed that TRBP binding to Dicer alters dicing kinetics and cleavage site selection—the biological significance of which is increased substrate affinity and enzymatic turnover and the generation of miRNA isoforms (isomiRs) (Figure 1D). The association of TRBP, but not PACT, induces and/or enhances the generation of isomiRs that are 1 nt longer in the cases of pre-miR-200a, pre-miR-29, pre-miR-34c, and pre-miR-132 (Lee and Doudna, 2012; Fukunaga et al., 2012). In vitro studies of the TRBP homolog, Loquacious-PB (Loqs-PB), in *D. melanogaster* have also demonstrated that binding of Loqs-PB, but not the isoform Loqs-PA, to Dcr-1 shifts the pre-miR-307a and pre-miR-87 cleavage site, generating longer (1–2 nt) miRNAs but, in contrast, a 1 nt shorter isomiR from pre-miR-316 (Fukunaga et al., 2012). Whereas TRBP/Loqs-PB binding to Dicer may cause conformational changes to the RNaseIII core domains, both TRBP and Loqs-PB have been shown to alter dicing efficiency in a substrate-specific manner (Lee and Doudna, 2012; Fukunaga et al.,

2012). For example pre-let-7, which possesses an ideal stem-loop structure, is diced 11-fold faster than pre-miR-21, which possesses a smaller loop and stem mismatches and bulges (Lee and Doudna, 2012), and alternative dicing of pre-miR-307a by Dcr-1-Loqs-PB requires the recognition of the stem and not the loop for redirecting Dicer cleavage (Fukunaga et al., 2012). As the presence of Loqs-PB and differences in pre-miRNA structure both affect dicing efficiency, it is thought that Loqs-PB/TRBP plays a role in substrate binding and product release, allowing processing of otherwise difficult pre-miRNA substrates (Fukunaga et al., 2012). The abundances of isomiRs with altered seed sequences and target specificities fluctuate during *Drosophila* development. Their functional significance is not fully understood, although Fukunaga et al., (2012) demonstrated that, despite being overlapping, the seed sequences are distinct, with the 23 nt (but not 21 nt) isoform of miR-307a specifically repressing *glycerol kinase* and *taranis* mRNAs, implying a role in organismal growth and metabolism. Beyond alteration of the seed sequence, Dicer-mediated tuning of the miRNA also has consequences for guide strand selection (Lee and Doudna, 2012), thought to be based on which strand has the weaker base pairing at its 5′ end. Lee and Doudna (2012) demonstrated that for miR-200a the strand originating from the 3′ strand of the pre-miRNA duplex (3p-arm) was the preferred guide strand when processed by Dicer alone, whereas a shift of 1 nt induced by TRBP-associated dicing resulted in both strands operating as guide strands.

Pre-miRNA 3′ Uridylylation as a Means to Regulate Dicer

In addition to regulation exerted via Dicer partner proteins, metazoans have evolved mechanisms for modifying pre-miRNAs, for example the untemplated addition of uridines to their 3′ ends, which alternatively promotes or inhibits dicing. Human Dicer optimally processes pre-miRNAs with a 2 nt 3′ overhang (Park et al., 2011). This preference is explained by the ~20 Å distance between the 5′ and 3′ end-binding pockets in the PAZ and platform domain achieved by a pre-miRNA with a 2 nt overhang (Park et al., 2011). However, it is now clear that not all pre-miRNAs are generated with a 2 nt 3′ overhang. In a recent and remarkable study, it was demonstrated that in chordates some let-7 miRNA family members possess an unpaired uridine, looped out of the stem of nascent pri-let-7 transcripts near the Drosha cut site, that is ignored during processing, giving a pre-miRNA with a 1 nt 3′ overhang (Heo et al., 2012). These pre-miRNAs with unusual structures are categorized as group II pre-let-7, and the majority of the let-7 family members belong to this group (2 nt 3′ overhangs identify group I; Heo et al., 2012). The group II 1 nt overhang is recognized and converted to 2 nt by the cytoplasmic terminal uridylyl transferases (TUTs) ZCCHC6 (TUT7), ZCCHC11 (TUT4), and the nuclear/cytoplasmic TUT2/GLD2 in vitro; however, ZCCHC6 has a dominant function and appears to be principally responsible for pre-let-7 mono-uridylylation in vivo (Heo et al., 2012). In contrast the developmental regulator Lin28A selectively binds to the pre-let-7 miRNA, recruits ZCCHC11 and switches it from a mono- to an oligo-uridylyl transferase (Heo et al., 2009). This not only results in an unfavorable substrate for dicing but is also thought to promote pre-let-7 decay to enhance the suppressive effects of oligo-uridylylation (Heo et al., 2009 and references therein).

Structural Basis of 3′ Uridylylation

The metazoan cytoplasmic TUTs ZCCHC6 (TUT7) and ZCCHC11 (TUT4) are large enzymes (~185 kDa) that possess a single N-terminal C2H2 zinc finger, three CCHC zinc knuckles, and tandem nucleotidyltransferase regions of which the N-terminal copy is catalytically inactive. They belong to a family of cytoplasmic TUTs of which caffeine-induced death suppressor 1 (Cid1) from fission yeast was first demonstrated to uridylate the 3′ ends of mRNAs (Rissland et al., 2007 and references therein). Cid1 exhibits extensive similarity to the tandem nucleotidyltransferase domains and, like ZCCHC6 and ZCCHC11, has a preference for UTP over other NTPs (Rissland et al., 2007; Yates et al., 2012). Recent determination of the crystal structure of Cid1 in complex with UTP revealed that its selectivity is mediated by a single histidine residue conserved in the catalytically active transferase domain of ZCCHC6/ZCCHC11 (Yates et al., 2012). Cid1 is a bona fide ssRNA-binding protein but the binding of structured pre-miRNAs by ZCCHC6/ZCCHC11 is conferred by their zinc knuckles as well as by Lin28A. It was recently demonstrated that the ZCCHC TUTs require intact pre-miRNA and, for ZCCHC11, the N-terminal C2H2 zinc finger, for Lin28-enhanced oligo-uridylylation. However, mutants lacking the N- or C-terminal region, or lacking the most C-terminal CCHC3 zinc knuckle still possessed Lin28-enhanced oligo-uridylylation activity (Thornton et al., 2012). ZCCHC11 was shown to associate briefly with pre-let-7 alone, transferring single uridines, but when in a stable ternary complex with Lin28A-pre-let-7 to promote oligo-uridylylation (Heo et al., 2012 and references therein). The structure of Lin28 bound to the terminal loop (pre-element) of pre-let-7 reveals a remodeling of the RNA with implications for pre-let-7 dicing and uridylylation (Nam et al., 2011); the pre-let-7 pre-element (preE) is base-paired at its center, creating a substructure that is divided into a preE loop, preE stem, and preE bulge (Nam et al., 2011). The cold-shock domain (CSD) of Lin28 inserts itself into the preE loop with a preference for the 9 nt sequence NGNGAYNNN, whereas its two CCHC zinc knuckles bind to the heptad AGGAGAU within the preE bulge (Nam et al., 2011). Mutation of the sequence GGAG to GAGG in the let-7 preE still permitted the stable association of Lin28 but prevented oligo-uridylylation and the formation of a stable ternary complex with ZCCHC11 (Heo et al., 2009). The binding of Lin28 to the preE results in the partial melting of the pre-let-7 RNA duplex close to the Dicer cut site (Nam et al., 2011). The Lin28 CCHC zinc knuckles also bind to the preE bulge, recently shown to be an important structure recognized by Dicer (Gu et al., 2012), and therefore sterically block Dicing at the 5p arm (Nam et al., 2011). As only the terminal pre-let-7 loop was used for structural studies of Lin28-preE complexes, it is unknown how far into the stem the duplex unwinding proceeds (Nam et al., 2011). It is therefore possible that ZCCHC6/ZCCHC11 recruitment to pre-let-7 in part depends on induction of long-range conformational changes in the dsRNA stem (Thornton et al., 2012).

Regulation of Mature miRNAs at Their 3′ Ends

In addition to alternative dicing, downstream regulatory processes have recently been revealed that alter mature miRNA length and are presumed to modify their target specificity. The

newly described *Drosophila* 3′-5′ exonuclease Nibbler is responsible for trimming Ago-loaded miRNAs (Liu et al., 2011; Han et al., 2011)—in *nibbler* mutant flies five (of nine) miRNAs with known multiple isoforms accumulated longer isomiRs and lost shorter ones (Liu et al., 2011). Nibbler associates with Ago1 in *Drosophila* and is capable of trimming longer, Dicer1-Loqs-PB generated, isomiRs (24 nt) back to the more typical 21–22 nt miRNA, whose 3′ end is then bound and protected by the PAZ domain of Ago (Liu et al., 2011; Han et al., 2011). In mammals, a large subset of miRNAs are shortened at their 3′ ends, as demonstrated in the developing murine brain, but the identity of the exonuclease(s) involved has not been established (Juvvuna et al., 2012). As in flies, these mature miRNAs are trimmed after guide strand selection, but also in an Ago paralog-specific manner (Juvvuna et al., 2012). Thus, of the four Ago paralogs in mammals, Ago2 specifically allows the trimming of miR-124 by 1 nt yielding a 21 nt isomiR (Juvvuna et al., 2012). All Ago proteins possess a PAZ domain and MID domain, that bind to the mature miRNA 3′ and 5′ ends respectively; Ago 2 possesses a parolog-specific PAZ domain variant that corresponds to the substitution K313R-Y314H (human Ago1 numbering) (Juvvuna et al., 2012). The lysine^{313} and tyrosine^{314} contact the penultimate and 3′ terminal phosphates, so it is thought that the Ago-2-specific substitution weakens the RNA-protein interaction to increase 3′ end accessibility for trimming (Juvvuna et al., 2012). Interestingly, 3′ trimming is also base-dependent, with a 3′ A being trimmed more efficiently than 3′ U (Juvvuna et al., 2012).

In addition to mature miRNA 3′ trimming, in vivo gene knockout studies coupled with deep sequencing have revealed that the TUT ZCCHC11 uridylates the 3′ ends of a large number of mature miRNAs (Jones et al., 2012). The addition of a single uridine to miR-126-5p or miR-379 dampened their silencing activity against insulin-like growth factor-1 (IGF-1) mRNAs and had a cumulative effect when multiple uridylylated miRNAs targeted a single transcript (Jones et al., 2012).

Outlook

Deep-sequencing approaches have provided evidence that individual miRNAs are frequently altered in length. The roles of the two endoribonucleases are well appreciated and they are obvious regulatory nodes for altering miRNA expression. Nevertheless, it is now clear that cells regulate Dicer processing to fine-tune miRNA target specificity via subtly altering RNA length (1–2 nt). The alteration of Dicer cut-site selection not only alters the miRNA seed sequence but also influences guide strand selection thus increasing the number of targets a single pre-miRNA may eventually repress. However, the wider biological effects of miRNA tuning have yet to be reported.

Another emerging regulatory node within the miRNA maturation pathway is found at the 3′ end of both mature and precursor miRNAs. The TUT-mediated remodeling of pre-miRNA 3′ ends during their biosynthesis can both protect from nucleases and promote maturation (mono-uridylylation) or, in embryonic stem cells, promote decay and inhibit processing (oligo-uridylylation), and both these processes can be explained in terms of the structural requirements of dicing. miRNA remodeling is not, however, restricted to such precursor forms. The tailing and trimming of mature miRNAs by competing terminal nucleotidtyltransferases and 3′-to-5′ exonucleases are pervasive and work to modify the efficacy of the miRNA. For example, Ago-loaded RNAs are subjected to 3′ tailing/trimming in a paralog-specific manner in both flies and humans but are also influenced by the target:miRNA complementarity—particularly at the miRNA 3′ end (Ameres et al., 2010).

ACKNOWLEDGMENTS

We apologize to all those whose work we could not cite directly due to space restrictions. We gratefully acknowledge financial support from Cancer Research UK. The Oxford Division of Structural Biology is part of the Wellcome Trust Centre for Human Genetics, Wellcome Trust Core Award Grant Number 090532/Z/09/Z.

REFERENCES

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Science *328*, 1534–1539.

Fukunaga, R., Han, B.W., Hung, J.H., Xu, J., Weng, Z., and Zamore, P.D. (2012). Cell *151*, 533–546.

Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N., and Kay, M.A. (2012). Cell *151*, 900–911.

Guil, S., and Cáceres, J.F. (2007). Nat. Struct. Mol. Biol. *14*, 591–596.

Han, B.W., Hung, J.-H., Weng, Z., Zamore, P.D., and Ameres, S.L. (2011). Curr. Biol. *21*, 1878–1887.

Heo, I., Ha, M., Lim, J., Yoon, M.J., Park, J.E., Kwon, S.C., Chang, H., and Kim, V.N. (2012). Cell *151*, 521–532.

Heo, I., Joo, C., Kim, Y.K., Ha, M., Yoon, M.J., Cho, J., Yeom, K.H., Han, J., and Kim, V.N. (2009). Cell *138*, 696–708.

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, E., Tuschl, T., and Zamore, P.D. (2001). Science *293*, 834–838.

Jones, M.R., Blahna, M.T., Kozlowski, E., Matsuura, K.Y., Ferrari, J.D., Morris, S.A., Powers, J.T., Daley, G.Q., Quinton, L.J., and Mizgerd, J.P. (2012). PLoS Genet. *8*, e1003105.

Juvvuna, P.K., Khandelia, P., Lee, L.M., and Makeyev, E.V. (2012). Nucleic Acids Res. *40*, 6808–6820.

Lau, P.W., Guiley, K.Z., De, N., Potter, C.S., Carragher, B., and MacRae, I.J. (2012). Nat. Struct. Mol. Biol. *19*, 436–440.

Lee, Y., Hur, I., Park, S.Y., Kim, Y.K., Suh, M.R., and Kim, V.N. (2006). EMBO J. *25*, 522–532.

Lee, H.Y., and Doudna, J.A. (2012). RNA *18*, 2012–2019.

Liu, N., Abe, M., Sabin, L.R., Hendriks, G.J., Naqvi, A.S., Yu, Z., Cherry, S., and Bonini, N.M. (2011). Curr. Biol. *21*, 1888–1893.

Nam, Y., Chen, C., Gregory, R.I., Chou, J.J., and Sliz, P. (2011). Cell *147*, 1080–1091.

Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J., and Kim, V.N. (2011). Nature *475*, 201–205.

Rissland, O.S., Mikulasova, A., and Norbury, C.J. (2007). Mol. Cell. Biol. *27*, 3612–3624.

Thornton, J.E., Chang, H.M., Piskounova, E., and Gregory, R.I. (2012). RNA *18*, 1875–1885.

Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., and Tomari, Y. (2011). Nat. Struct. Mol. Biol. *18*, 1153–1158.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. (2008). Science *320*, 97–100.

Yates, L.A., Fleurdépine, S., Rissland, O.S., De Colibus, L., Harlos, K., Norbury, C.J., and Gilbert, R.J. (2012). Nat. Struct. Mol. Biol. *19*, 782–787.