Update TRENDS in Biotechnology Vol.23 No.3 March 2005

Research Focus

Engineering regulatory RNAs

Eric A. Davidson[1] and Andrew D. Ellington[1,2]

[1]Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, TX 78712, USA
[2]Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA

RNA has long been a favoured medium for *in vitro* evolution and engineering. Functional RNAs produced *in vitro* can bind small molecules (aptamers), possess catalytic activity (ribozymes) or do both (aptazymes). A plethora of recent work has shown similar strategies used naturally for gene regulation in bacteria. Interest in these natural systems has inspired an effort to engineer and evolve this activity *in vivo*. A recent paper by Isaacs *et al.* describes the engineering and *in vivo* activity of a small RNA that removes translation inhibition by binding the 5′ untranslated region of its target mRNA and making the ribosome-binding site accessible.

Introduction

RNA is known to possess many different functions in the cell, ranging from acting as a passive messenger between the genome and the proteome to actively controlling the expression patterns of various genes. The recent renaissance of RNA biology, in the form of microRNAs and riboswitches, can be explained by noting three aspects of RNA chemistry. First, the simplicity of base-pairing interactions allows new targets to be readily acquired by regulatory RNA molecules. Second, RNA molecules, like proteins, can fold into complex shapes that bind metabolites and catalyze reactions. Third, RNA molecules, unlike proteins, can sample a wide variety of secondary structures and undergo facile, large-scale conformational changes.

The same chemical principles that have made RNA a 'natural' evolutionary choice for metabolic regulation mean that RNA can be used as an engineered regulatory element. Although discoveries in biology usually herald applications in biotechnology, this is no longer the case for RNA regulatory elements. The engineering of conformation-switching aptamers as biosensors predated the discovery of conformation-switching riboswitches in the 5′ untranslated regions (UTRs) of genes that control transcription or translation via ligand binding [1,2]. Similarly, the *in vitro* selection of allosteric ribozymes (aptazymes) preceded the discovery of a natural aptazyme that controls glucosamine-6-phosphate metabolism [3,4]. Further interplay between discoveries in biology and biotechnology will be best applied in the emerging field of synthetic biology, in which natural genetic elements are incorporated into unnatural genetic circuits with startling complexity. Because of the ease of engineering RNA base-pairings, ligand interactions and conformational changes, the synthetic biology 'toolbox' will increasingly be composed of RNA-based gene regulation methods such as those developed by Isaacs *et al.* [5].

Corresponding author: Ellington, A.D. (andy.ellington@mail.utexas.edu).
Available online 26 January 2005

Functional RNAs in nature

RNA has many different functions in the regulation of gene expression in organisms. Many functional RNAs are located in the 5′ or 3′ UTRs of the mRNAs that they regulate. Other small regulatory RNAs are transcribed independently and act through intermolecular interactions.

There are numerous examples of how protein binding to RNA can lead to changes in gene expression. The converse is also true – some RNAs bind to proteins to affect gene expression. For example, 6S RNA is expressed during the stationary phase of *Escherichia coli* growth and binds sigma70 of the RNA polymerase, inhibiting transcription [6]. Small molecule binding to RNA can also regulate metabolism. The recently discovered riboswitches are regulatory RNAs that are located in the 5′ UTRs of bacterial mRNAs and act as metabolite sensors [7]. Ligand binding causes secondary structural changes that lead to regulation of transcription (by premature termination) or translation [by sequestering of ribosome-binding sites (RBSs)]. These regulatory interactions are often negative feedback loops for biosynthetic operons; for example, a coenzyme B12-responsive riboswitch is present upstream of the *btuB* gene in *E. coli* and the *btuB* gene and cob operon in *Salmonella typhimurium* [8]. RNA-gated conformational changes can also function in other ways – in both *Yersinia pestis* and *Listeria monocytogenes* RNA elements in a 5′ UTRs functions as thermometers [9,10]. At temperatures favorable to host colonization the thermosensor RNAs activate translation by unfolding and providing access to the ribosome. In addition, ligand-mediated RNA conformational changes have been shown to modulate ribozyme catalysis. A cleavase located in the 5′ UTR of the *glmS* gene is activated by glucosamine-6-phosphate [4]. Ribozyme cleavage leads to further mRNA degradation and the inhibition of gene expression.

Small non-translated RNAs are another growing class of regulatory RNAs that are the subject of the engineering efforts reported by Isaacs *et al.* [5]. Some small RNA transcripts in bacteria (sometimes referred to as 'riboregulators') are expressed in response to environmental stimuli, such as oxidative stress, temperature or the presence of toxins [11]. They act through sequence-specific RNA–RNA interactions and are thus the bacterial

 0167-7799/$ - see front matter doi:10.1016/j.tibtech.2005.01.006

equivalent of eukaryotic siRNAs or microRNAs. Riboregulators have been particularly well-characterized in *E. coli* and related bacteria and most of these small RNAs use the protein Hfq as a cofactor [12]. This protein is thought to protect the small RNA from ribonuclease degradation (small RNAs in *E. coli* have an abnormally long half life) and to mediate interactions between small RNAs and their target mRNAs [13].

Like microRNAs, the majority of small regulatory bacterial RNAs inhibit translation, although some have been found that activate translation [11]. The best-characterized mechanism of action involves base-pairing of the small RNA to a target mRNA, usually in the 5′ UTR near the RBS. Hybridization usually leads to RNase E-mediated degradation of the target mRNA and the bound small RNA [13]. However, in at least one example the small RNA selectively inactivates translation of a single gene in an operon, rather than the entire mRNA transcript; spot42 RNA transcripts selectively regulate ribosome-binding to the *galK* gene without affecting other genes in the galactose operon (galETKM) [14]. Most small regulatory RNAs act on more than one target mRNA. The *rhyB* RNA transcripts have at least six target mRNAs, all involved in iron storage and regulation in *E. coli* [15] and *dsrA* RNA has at least two known mRNA targets. Interestingly, different domains of the small RNA transcript bind to each target and yield opposite regulatory results: *dsrA*-binding activates translation of *rpoS* mRNA, but inhibits translation of hns mRNA [16,17].

Synthetic riboregulators

As is usually the case with biological discoveries, the understanding that RNA molecules can act as regulatory elements has numerous engineering applications. There are examples of unnatural aptamers [1], ribozymes [18] and aptazymes [19] that have been harnessed to regulate gene expression and now, as described by Isaacs *et al.* [5], artificial riboregulators that can regulate gene expression.

The work by Isaacs *et al.* represents a novel gene-independent approach to engineering post-transcriptional regulation (Figure 1). Knowledge of natural small regulatory RNAs and their target mRNAs (dsrA:rpoS and hok:sok, a plasmid segregation system using small RNA–mRNA interactions [20]) was the basis for the design of artificial riboregulators. Like their natural counterparts, the engineered RNA–RNA interactions regulate translation by modulating ribosome binding to 5′ UTR sequences. To achieve this two new components were introduced into the *E. coli* regulatory network: a *cis*-repressing sequence in the 5′ UTR of the regulated gene and an independently transcribed *trans*-activating small RNA that can bind to the *cis*-repressing sequence.

The *cis*-repressing RNA is designed so that the ribosome-binding site is sequestered within a stem-loop structure (Figure 1). To ensure the formation of the regulatory structure, Isaacs *et al.* used Mfold (a web server for nucleic acid folding and hybridization prediction) to identify which stem-loops had energies of folding that were well-separated from other conformers. However, no controls were performed with designed RNAs that had not been pre-selected by MFOLD, so it is unclear whether this step was necessary. The loop sequence was also an important design variable because it mediates interactions with *trans*-activating RNA, similar to the way that loops in molecular beacons mediate interactions with mRNA targets. A YUNR sequence [where Y denotes a pyrimidine base (U or C), U denotes a uracil base, N denotes any of the 4 standard bases and R denotes a purine base (A or G)] had previously been noted to be conserved in the loop sequences of other natural stem-loop–small RNA interactions, and had been shown to affect the function of these regulatory pairs, thus this sequence was also included in the design of the artificial riboregulatory circuit [21].

The *trans*-activating RNA was expressed as a small, imperfectly base-paired hairpin that contained the reverse complement of the *cis*-repressing sequence (Figure 1). The hairpin structure sequesters a RBS within the *trans* RNA itself, possibly because of interference from the ribosome. When the *trans*-activating RNA hybridizes to the *cis*-repressing sequence in the 5′ UTR it should disrupt the engineered secondary structure so that the RBS becomes single-stranded and thus available to the ribosome.

The authors tested a number of *cis*-repressing stem-loop structures with a variety of predicted stabilities for their ability to repress translation of a fluorescence activated cell sorter (FACS) variant of GFP (Green Fluorescent Protein) at medium and high levels of transcription. Those structures that could strongly repress translation at high transcription levels were further tested for *trans*-activation. The constructs repressed expression by 96–98% relative to mRNAs that did not contain a *cis*-repressing stem-loop structure. However, the overall level of expression dropped by 60%, possibly because the engineered stem-loop structure caused transcription termination or was a target for endogenous RNases. This unexpected change in expression levels is a reminder that although our understanding of biology has advanced to the point where it is possible to design complex, functional genetic circuitry, this does not mean that the global effects of such circuitry can be predicted with any certainty.

Strong and specific activation was then achieved by co-expressing the *trans*-activating RNAs for each *cis*-repressing sequence. The two most efficient riboregulators yielded activations of 8- and 19-fold the background rate. The engineered constructs out-performed *dsrA* activation of *rpoS*, which yields approximately a 5-fold increase in activation rate [16]. Two closely related pairs of trans-activating RNAs (~80% sequence identity) and their cognate targets were checked for cross activity. The activation level for mismatched pairs was only slightly above the background rate. This is an important consideration for the design of synthetic biological circuits in which unanticipated cross-talk with natural regulatory circuits, particularly protein-based regulatory circuits, can often cause unforeseen problems.

Part of the importance of this work lies in the fact that the authors used modular design principles when generating riboregulators. Because the *cis* repressive domain functions independently of the gene sequence it regulates

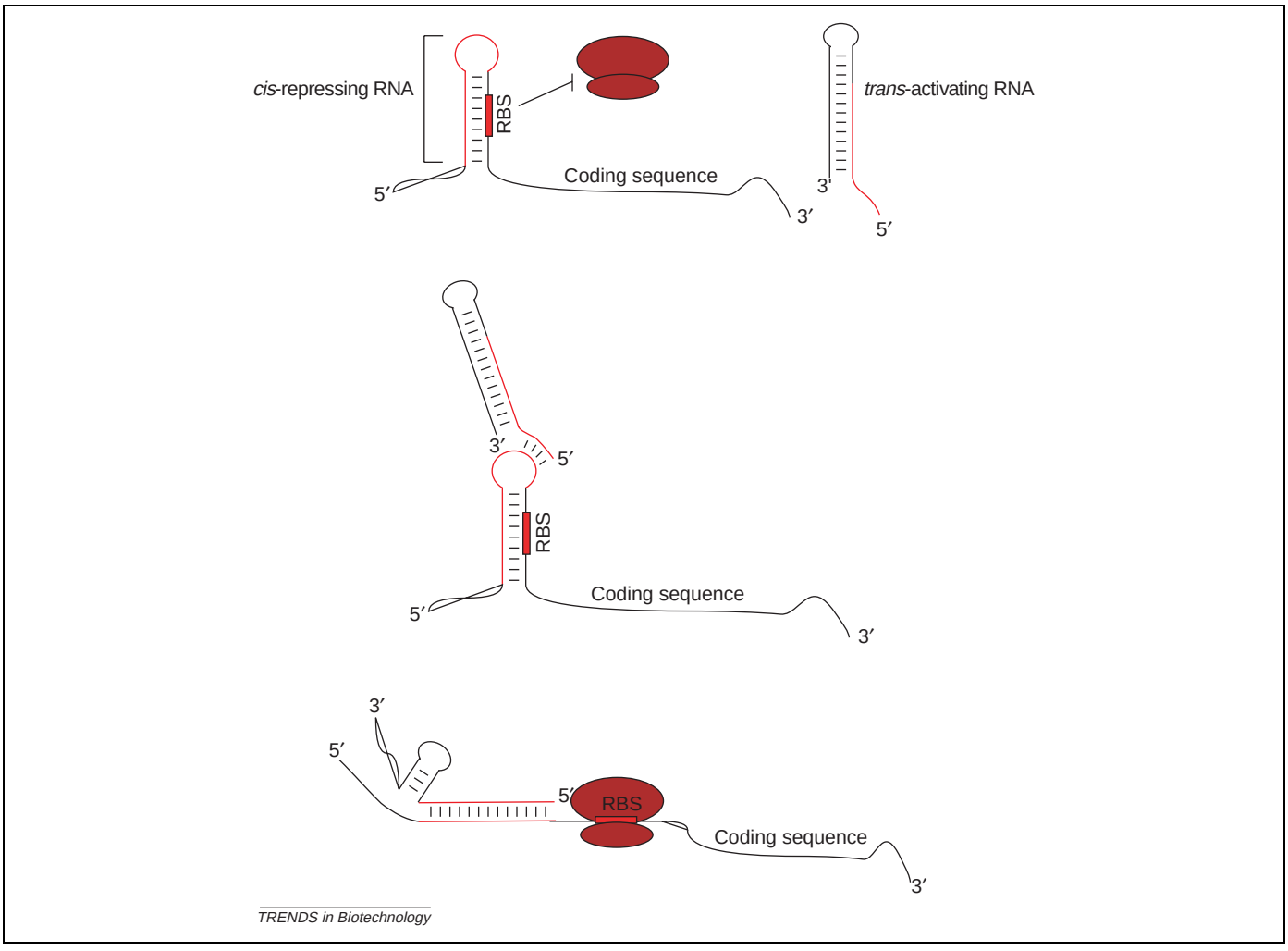

Figure 1. Mechanism of action of engineered riboregulators. The single-stranded 5′ end of the *trans*-activating RNA interacts with the YUNR consensus sequence in the loop of the *cis*-repressing RNA to unfold the stem-loop and allow ribosome binding to the RBS.

or the promoter sequence it is transcribed from, it can potentially be placed near the RBS of almost any gene to repress translation. This feature contrasts with other RNA-based attempts at gene regulation, such as conventional anti-sense approaches and targeted *trans*-acting hammerheads. Such modularity is an important consideration when developing parts for the 'toolbox' of synthetic biology. Parts such as promoters, operators and post-transcriptional regulatory systems should function independently of the coding sequence that they regulate if they are to be broadly applicable as engineering tools.

Modular design also facilitates the development of multiple orthogonal regulators that can be used in parallel with one another, allowing post-transcriptional regulation of many genes independently. However, the extent to which the design rules can be multiplexed is currently unclear. Only one sequence and closely related variants were studied, and it is possible that the general scheme for regulation is sensitive to sequence and structural perturbations. For example, attempts to increase the *trans* RNA concentration by the modular addition of a 'stabilizer' element to the 5′ end stopped all activation.

Concluding remarks

Given the efficiency and potential versatility of the RNA regulatory mechanism described by Isaacs *et al.*, and following up on our original point that natural and engineered molecules and mechanisms closely mimic one another, the obvious question is why do organisms not use RNA-based regulation more frequently than protein-based regulation? The answer might be that we have not yet discovered the full panoply of RNA-based regulation – just as several years ago we did not know about riboswitches that regulated translation. However, another answer, supported by the data presented in the cited paper, is that RNA might be more limited by informational parsimony than proteins are, at least given natural evolutionary mechanisms. Consider a situation in which a gene is regulated by multiple factors in parallel (Figure 2). For each new RNA regulatory element added upstream of a gene ~50 residues would be required given the engineered *cis* regulatory sequences that have been explored. Contrast this with the fact that many transcription factor binding sites can be ~6–12 residues in length. For synthetic biologists with access to design principles and DNA synthesizers, the introduction of multiple, long regulatory sequences presents no problem. For natural

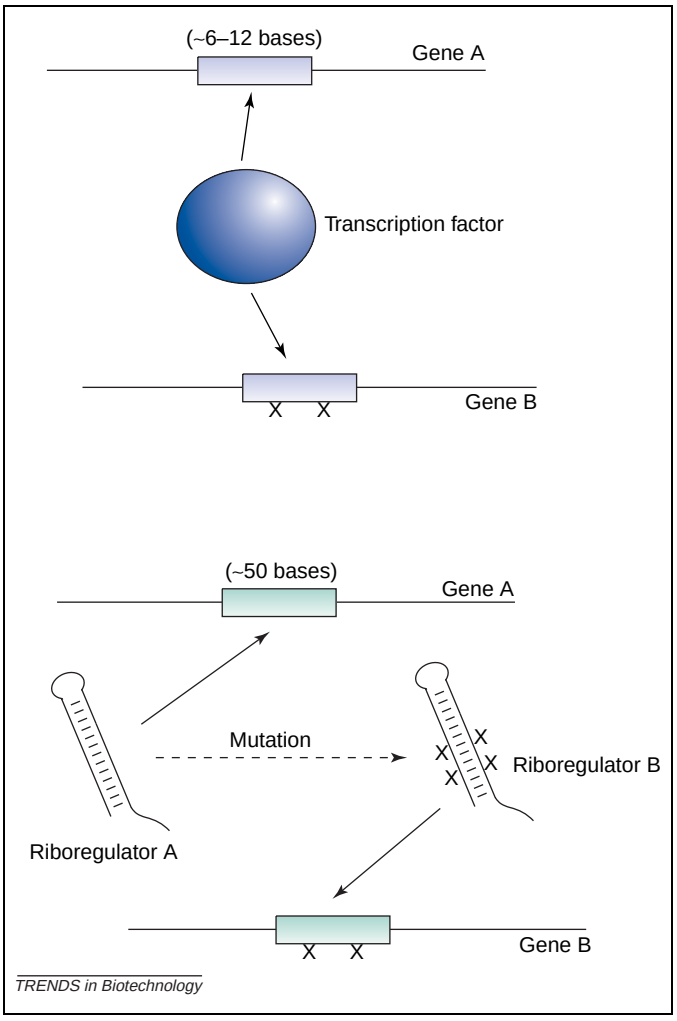

Figure 2. Acquisition of protein or RNA regulators during natural selection. (a) Acquisition of protein regulators. A given DNA-binding protein can frequently bind to a new, short site following the introduction of a small number of mutations. In this figure, the blue transcription factor binds to a new site upstream of a different gene based on two mutational changes, represented as 'X's'. (b) Acquisition of RNA regulators. For a given RNA regulator to bind to a new site, it must not only find a similar site (a relatively rare event), but in all likelihood must itself mutate (X's) to complement the site. Thus, the overall probability of acquiring a new site is much lower than for protein regulators.

selection, the probability of evolving a new regulatory site upstream of a gene is inversely related to the size of the regulatory site. The shorter the site, the easier the evolutionary process. Moreover, the sequence-specificity of RNA regulatory mechanisms, a positive feature outlined previously with respect to cross-talk, will of necessity limit the acquisition of 'partially functional' sites during evolution. Transcription factor binding sites that are 80% identical to one another might be functional; RNA binding sites with similar identities might not be. Thus, RNA regulatory mechanisms and elements, despite their inherent functionality, might be more difficult to acquire and distribute than protein regulatory mechanisms. That said, such an analysis could be belied by the fact that multiple microRNA binding sites are often found in tandem in the 3′ UTRs of genes. There are almost certainly subtle features of particular regulatory mechanisms and the information required to engender regulatory mechanisms that remain to be discovered and analyzed.

Acknowledgements

The authors are sponsored by the Defense Advanced Research Projects Agency (DARPA) grant number DS793.

References

1 Werstuck, G. and Green, M.R. (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science* 282, 296–298
2 Winkler, W. *et al*. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952–956
3 Tang, J. and Breaker, R.R. (1997) Rational design of allosteric ribozymes. *Chem. Biol.* 4, 453–459
4 Winkler, W.C. *et al*. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281–286
5 Isaacs, F.J. *et al*. (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotechnol.* 22, 841–847
6 Wassarman, K.M. and Storz, G. (2000) 6S RNA regulates E. coli RNA polymerase activity. *Cell* 101, 613–623
7 Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5, 451–463
8 Nahvi, A. *et al*. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* 32, 143–150
9 Hoe, N.P. and Goguen, J.D. (1993) Temperature sensing in *Yersinia pestis*: translation of the LcrF activator protein is thermally regulated. *J. Bacteriol.* 175, 7901–7909
10 Johansson, J. *et al*. (2002) An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* 110, 551–561
11 Gottesman, S. (2004) the small RNA regulators of *Escherichia coli*: roles and mechanisms. *Annu. Rev. Microbiol.* 58, 303–328
12 Valentin-Hansen, P. *et al*. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol. Microbiol.* 51, 1525–1533
13 Masse, E. *et al*. (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* 17, 2374–2383
14 Moller, T. *et al*. (2002) Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.* 16, 1696–1706
15 Masse, E. and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4620–4625
16 Lease, R.A. *et al*. (1998) Riboregulation in *Escherichia coli*: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12456–12461
17 Lease, R.A. and Belfort, M. (2000) Riboregulation by DsrA RNA: trans-actions for global economy. *Mol. Microbiol.* 38, 667–672
18 Yen, L. *et al*. (2004) Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 431, 471–476
19 Thompson, K.M. *et al*. (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol.* 2, 21
20 Gerdes, K. *et al*. (1997) Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.* 31, 1–31
21 Franch, T. *et al*. (1999) Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *J. Mol. Biol.* 294, 1115–1125