Commentary

Do mRNAs act as direct sensors of small molecules to control their expression?

Gary D. Stormo* and Yongmei Ji

Department of Genetics, Washington University Medical School, St. Louis, MO 63110

The paper in this issue of PNAS by Miranda-Rios *et al.* (1) demonstrates the importance of a conserved RNA structure in the regulation of genes involved in thiamin biosynthesis.

Thiamin, also known as vitamin B_1, is a cofactor for many important enzymes and therefore essential for growth. Bacteria have genes for all the enzymes necessary to synthesize thiamin, but if adequate amounts are present in the environment, they can use those rather than make their own, saving the energy and materials that would otherwise be used for synthesis. Such feedback inhibition, where the product of a pathway can repress the expression of the enzymes in the pathway, is quite common in bacteria. Many of the most interesting results in molecular biology over the past 50 years have come from unraveling the mechanisms of such feedback regulation. Although bacteria, at least many species, have the ability to make all of the complex molecules they need for growth from simple compounds, they can also use environmental sources of those molecules and, in doing so, repress synthesis of the genes required to make them. Such a regulatory response requires, at a minimum, a means of sensing the concentration of the product and a mechanism to control the expression of the relevant genes that depends on that concentration. In the case of thiamin regulation, the data suggest that the mRNA for the synthesizing enzymes may itself serve as the sensor and provide the mechanism for regulation.

Although each regulatory feedback loop has its own features, some mechanisms are quite common. Typically, there is a regulatory protein that can sense the level of the product and then bind to the DNA or RNA to affect the expression of the relevant enzymes. For example, the *trp* repressor of *Escherichia coli* binds to DNA only if it is first bound by the amino acid tryptophan (2). Higher concentrations of tryptophan increase the probability of the repressor binding to the DNA where it turns off expression of the genes needed to synthesize tryptophan. This is an example of transcriptional control where regulation affects the activity of the promoter and the amount of mRNA made. Regulation by thiamin is shown to be posttranscription initiation (1). The amount of mRNA initiated from the promoter appears unaffected by the presence of thiamin, but the elongation of the mRNA is attenuated by a terminator structure in the middle of the first gene of the operon, *thiC*. In *Bacillus subtilis*, the *trp* operon is regulated by a similar attenuation process (2, 3). When the TRAP protein binds tryptophan, it can then bind to the mRNA and lead to its premature termination, inhibiting the expression of all the genes required for tryptophan synthesis. *E. coli* also uses an attenuation process as a second level of control, with the ribosome sensing the concentration of tryptophan in the cell (actually the concentration of charged $tRNA^{Trp}$) (2).

In thiamin regulation, the attenuating termination structure forms when translation initiation of the *thiC* gene is blocked (1). When the mRNA is translated by a ribosome, the structure will not form, but thiamin somehow inhibits ribosome initiation, thereby leading to premature termination of the mRNA. Examples of repressing translation initiation are also well known. Several ribosomal proteins are known to bind to their own mRNA and repress their own synthesis if their primary binding target, the ribosomal RNA, is not available (4). More relevant to the case of thiamin is an example where a sensor protein, in the presence of a sufficient effector molecule, blocks translation initiation. The *B. subtilis* TRAP also performs this function at the *trpG* gene, where its binding site overlaps the translation initiation region and, at high concentrations of tryptophan, it binds to the mRNA and blocks ribosome binding (3, 5). So regulation by thiamin appears to have several features that have been observed previously, and the critical open question is the mechanism by which thiamin inhibits translation initiation of the *thiC* gene.

The critical open question is the mechanism by which thiamin inhibits translation initiation of the *thiC* gene.

The Role of the *thi* Box. The mRNA for the *thiCOGE* operon in *Rhizobium etli* contains a 211-base leader with two features that are likely to be important. One is a hairpin structure that is just 5′ of the *thiC*-initiating AUG and overlapping the ribosome-binding site (RBS), where one expects it would inhibit ribosome binding. The other feature is called the "*thi* box," which is a site of 38 bases that is highly conserved in the region 5′ to the start of many genes, from many species, that are involved in thiamin biosynthesis. That the *thi* box is important in the mRNA, rather than the DNA, is demonstrated by comparison of the different occurrences of it. Only about half of the positions are completely conserved, but all of the sites can fold into the same RNA secondary structure because of compensating changes that maintain complementarity. Because this structure is conserved in different thiamin-related operons and different species rather than the RBS structure, it is most likely the one directly involved in the response to thiamin concentration. However, both structures are required for proper regulation. In an mRNA without the *thi* box, translation of the *thiC* gene is inhibited, presumably because of the RBS structure. Somehow the *thi* box relieves the intrinsic inhibitory effect of the RBS structure in the absence of thiamin but allows it to occur in the presence of thiamin. This can be explained by alternative RNA structures, in which the switch between them is modulated by thiamin. But what sets this example apart from other known posttranscriptional regulatory events is the absence of any regulatory protein. This would be simply another interesting example of a sensory-regulatory mechanism, except there is no protein to do the sensing. It is certainly possible that a regulatory protein

See companion article on page 9736.

*To whom reprint requests should be addressed. E-mail: stormo@genetics.wustl.edu.

exists that has not been identified yet, but more intriguing is the possibility, raised by the authors (1), that thiamin concentration is sensed by the mRNA itself, which then controls the rate of translation. At low thiamin concentration, an alternative to the RBS structure occurs that allows normal translation, readthrough of the terminator structure, and synthesis of the thiamin-synthesizing enzymes. However, at high thiamin concentration, the *thi* box is bound, releasing the RBS structure to form, translation to be blocked, and attenuation to occur.

A similar model has been proposed for the regulation of genes involved in synthesis and uptake of cobalamin, a precursor to vitamin B_{12} (6, 7). The *cob* operon of *Salmonella typhimurium* encodes the genes required for synthesis of cobalamin, and the *btuB* gene of both *E. coli* and *S. typhimurium* encodes a transporter for cobalamin. All have been shown to be repressed by cobalamin posttranscriptionally (8, 9). Just 5′ of the AUG for each gene is a RBS hairpin, and each gene has a "B_{12} box" in a long untranslated leader sequence. Additional sequences between the B_{12} box and the AUG have been shown to be important for translation and for regulation by cobalamin. Adenosyl-cobalamin, the active repressor form of cobalamin, has been shown to specifically inhibit ribosome binding to the *btuB* mRNA *in vitro* (6). As in the case with thiamin, no regulatory protein has been identified. However, attempts to measure direct binding of adenosyl-cobalamin to *in vitro* synthesized mRNA for both genes have not succeeded (6, 7). But it was also shown that *in vitro* synthesized *btuB* mRNA was incapable of binding ribosomes, even in the absence of cobalamin, without some cellular factor that is removed by high salt washing of ribosomes (6). One possible explanation for these results is that the *in vivo* mRNA has a different structure than that transcribed *in vitro*, and only the *in vivo* mRNA can bind to adenosyl-cobalamin. A model has been proposed for a complex structure of the *cob* mRNA leader sequence, supported by mutational analysis, chemical structure probing, and computer prediction (7). The conclusion from the cobalamin-regulated genes, as for the thiamin-regulated genes, is that the effector molecule binds to the mRNA, either alone or with some factor, to modify its structure so it no longer binds ribosomes efficiently.

A similar mechanism has been proposed for the regulation of genes used in the synthesis of riboflavin, another vitamin (10). Previous work has shown that the regulatory site for the *B. subtilis* riboflavin operon was within the mRNA sequence and suggested translational regulation (11). Comparison of the riboflavin synthetic genes from many species has led to the identification of a conserved structure in the mRNA leader sequence and the proposal that it binds directly to riboflavin to control translation initiation (10).

Why Propose Direct RNA Binding? Why should we think that these vitamins, cobalamin, riboflavin, and thiamin, might be the effectors themselves, without the aid of other factors such as regulatory proteins? That no regulatory proteins have been found despite significant efforts is suggestive only because one can imagine why they might have escaped detection (1, 6, 7). Equally suggestive is our recent, and growing, appreciation for the range of functions that RNAs can possess. From the discovery of self-splicing RNA (12) to the recent verification of the role of rRNA in peptide bond formation by the ribosome (13, 14), it has become clear that RNAs play many essential roles and may well have provided the initial form of life on earth (15). In addition, RNAs can be made to have many functions besides those found (so far at least) in nature (16, 17).

Vitamins are used as cofactors for many enzymes and might possibly have been used in a similar manner by ribozymes in an RNA world.

Besides catalytic activities, RNAs called aptamers can be selected *in vitro*, by a procedure called SELEX, to bind to target molecules (18, 19). Aptamers with high affinity and specificity can be selected to a wide range of targets, including those with potential for diagnostic and therapeutic applications (20–22). Most relevant to regulation of gene expression and vitamin synthesis is that aptamers have been selected to bind with high affinity and specificity to a large number of small molecules, including amino acids, nucleotides, antibiotics, and enzyme cofactors (23). Among their attributes are very high specificity, as in the ability of an RNA aptamer to distinguish between theophylline and caffeine, which differ by a single methyl group, by 10,000-fold in affinity (24). The aptamer folds into an intricate structure that makes multiple contacts with theophylline but sterically excludes caffeine with an extra methyl group (25). In general, aptamers for small molecules contain bulge and internal loops that can adopt specific binding pockets on interaction with the ligand (26, 27). The conserved structure of the *thi* box contains such an internal loop that could participate in a binding pocket for thiamin (1).

Given the wide variety of small molecule targets for which aptamers have been selected, it is easy to imagine that natural mRNAs might also contain specific binding sites for some small effectors that could be used in regulating gene expression. In a test of that idea, Werstuck and Green selected an aptamer to an organic dye that is nontoxic and cell-permeable (28). The resulting RNA can be folded into a hairpin structure with several internal loops. That aptamer sequence was placed within the untranslated leader of a reporter gene on an expression vector and transfected into mammalian cells. The reporter gene was expressed normally in the absence of the dye but could be switched off specifically by adding the dye to the medium. Other experiments showed clearly that various aptamers could fold properly and function as expected *in vivo*.

It is clear that mRNA sequences could be direct sensors of small molecules and control their own expression in response. But the question in the title, whether they actually do, remains open. Even if thiamin, cobalamin, and riboflavin are shown to be effectors that act directly on the mRNA, there are a number of other interesting open questions. One wonders why such a mechanism is not common, given the simplicity of the regulatory loop and the proven ability of aptamers to bind with high sensitivity and specificity. Perhaps it is common, but more thorough searches are required to find examples. Is this mechanism used for small molecules other than vitamins? There is no reason to exclude its use for any target, as aptamers have been selected for a wide range of small molecules (23). Vitamins are used as cofactors for many enzymes and might possibly have been used in a similar manner by ribozymes in an RNA world, in which case it is easy to imagine that specific RNA pockets for binding them may have ancient origins (15). And finally we ask the question (assuming the proposed structures and mechanisms for thiamin, cobalamin, and riboflavin are correct): Why are they so complicated? One can easily imagine a small RNA structure, the aptamer for the specific effector molecule, being positioned so when it forms, it blocks translation and being "tuned" such that without binding the effector molecule, it is too weak to have a significant effect; however, on binding, it becomes a very tight and inhibitory complex. That is exactly the model shown to be effective by Werstuck and Green (28), but it is quite different from the proposed models for the vitamins (1, 7, 10). The *thi* box of *Mycobacterium tuberculosis* is an exception that does fit the simple model (1). It is just

5′ of the initiating AUG and includes the RBS, appearing as a fusion of the *thi* box and RBS structures seen in the other genes. It may be important that the *M. tuberculosis* site has an extra stem and internal loop that may constitute the full binding site missing in the other examples. We have recently scanned the bacterial genomes in GenBank and have found at least 40 examples of *thi* box sites upstream of *thiC* homologs. In all but a couple of cases, such as *M. tuberculosis*, they are separate from the RBS. In many cases, but not all, there is an obvious RBS structure as well, but they are not highly conserved in sequence or structure. Assuming the *thi* box is the site of direct interaction with thiamin (with or without a protein factor), this model requires the response to binding to be transmitted to the RBS, presumably by a structural rearrangement of the mRNA leader sequence, as is proposed for cobalamin regulation (7). The advantage to such a complicated system is not clear, compared with the more direct *M. tuberculosis* model. Perhaps unraveling the real mechanism will help explain it. If it turns out to be a general rule that direct sensing by mRNA of small molecule effectors uses complicated structural rearrangements, the motifs will be more difficult to identify by sequence comparison methods. But the constantly increasing sequence databases provide the potential to discover ever more interesting and complex regulatory systems.

1. Miranda-Rios, J., Navarro, M. & Soberon, M. (2001) *Proc. Natl. Acad. Sci. USA*, **98,** 9736–9741. (First Published July 24, 2001; 10.1073/pnas.161168098)
2. Hartl, D. L. & Jones, E. W. (2001) *Genetics: Analysis of Genes and Genomes* (Jones and Bartlett, Sudbury MA), pp. 499–505.
3. Babitzke, P. (1997) *Mol. Microbiol.* **26,** 1–9.
4. Nomura, M. (1999) *J. Bacteriol.* **181,** 6857–6864.
5. Du, H., Tarpey, R. & Babitzke, P. (1997) *J. Bacteriol*. **179,** 2582–2586.
6. Nou, X. & Kadner, R. J. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7190–7195. (First Published June 13, 2000; 10.1073/pnas.130013897)
7. Ravnum, S. & Andersson, D. I. (2001) *Mol. Microbiol.* **39,** 1585–1594.
8. Ravnum, S. & Andersson, D. I. (1997) *Mol. Microbiol*. **23,** 35–42.
9. Nou, X. & Kadner, R. J. (1998) *J. Bacteriol*. **180,** 6719–6728.
10. Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. (1999) *Trends Genet*. **15,** 439–442.
11. Kil, Y. V., Mironov, V. N., Gorishin, I. Y., Kreneva, R. A. & Perumov, D. A. (1992) *Mol. Gen. Genet*. **233,** 483–486.
12. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982) *Cell* **31,** 147–157.
13. Green, R. & Noller, H. F. (1997) *Annu. Rev. Biochem.* **66,** 679–716.
14. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000) *Science* **289,** 920–930.
15. Gesteland, R. F. & Atkins, J. F. (1993) *The RNA World* (Cold Spring Harbor Lab. Press, Plainview, NY).
16. Jaeger, L. (1997) *Curr. Opin. Struct. Biol.* **7,** 324–335.
17. Kurz, M. & Breaker, R. R. (1999) *Curr. Top. Microbiol. Immunol.* **243,** 137–158.
18. Tuerk, C. & Gold, L. (1990) *Science* **249,** 505–510.
19. Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **346,** 818–822.
20. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. (1995) *Annu. Rev. Biochem.* **64,** 763–797.
21. Brody, E. N. & Gold, L. (2000) *J. Biotechnol.* **74,** 5–13.
22. Hesselberth, J., Robertson, M. P., Jhaveri, S. & Ellington, A. D. (2000) *J. Biotechnol.* **74,** 15–25.
23. Famulok, M. (1999) *Curr. Opin. Struct. Biol.* **9,** 324–329.
24. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. (1994) *Science* **263,** 1425–1429.
25. Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. & Pardi, A. (1997) *Nat. Struct. Biol*. **4,** 644–649.
26. Hermann, T. & Patel, D. J. (2000) *Structure Fold. Des.* **8,** R47–R54.
27. Hermann, T. & Patel, D. J. (2000) *Science* **287,** 820–825.
28. Werstuck, G. & Green, M. R. (1998) *Science* **282,** 296–298.