Riboswitches: still a lot of undiscovered country

ROBERT T. BATEY

Department of Chemistry and Biochemistry, University of Colorado, UCB 596, Boulder, Colorado 80309-0596, USA

In this era of electronic publishing there are not a lot of compelling reasons to keep one's personal library filled with shelves of dusty print journals. Over the past few years, my office and laboratory have been depopulated of most of these relics with the exception of a few issues of special significance for various reasons: an issue of *Nature* with the first hammerhead structure, issues of *Science* describing structures of the P4-P6 domain from a group I intron and of the 50*S* ribosomal subunit and, of course, some issues containing papers of my own. But, amongst the rump of this once substantial hoard are issues 2 and 5 of volume 1 of *RNA*. Why did these survive? Unlike most of the others, they have dog-eared pages, highlighted references, and post-its with comments about a technique I needed to look into or a result relevant to my research project at the time. As a graduate student at the time of the inception of this journal, *RNA* rapidly became the journal that I *studied* and thus an enormously valuable learning tool that greatly shaped my growth as a scientist. I suspect that I just could not bear to let go of these issues, not because of any one significant piece of research—as in the others—but just the fact that they were old friends that didn't need to go into the recycling bin just yet.

Looking through the somewhat yellowed pages of these issues from the not-so-distant past, I find them remarkable in what I didn't see. On their cover is a brightly colored model of a group I intron from François Michel, but within their pages is an almost complete absence of RNA structure with the exception of a nice article on U-turns in GNRA tetraloops from Art Pardi. Indeed, when *RNA* was founded, very little was known about RNA structure beyond tRNA, tetraloops, and simple A-form duplexes. While there is plenty of research on RNA "oldies-but-goodies" such as tRNA, the spliceosome, and the ribosome in these early issues, entirely absent is the plethora of regulatory RNAs including small interfering RNAs, lncRNAs, CRISPR RNAs, or my current personal favorite, riboswitches. All of these RNAs, and many more, were discovered after the inception of *RNA*, and this journal has been able to chronicle work on characterizing them from discovery through unraveling the details of their mechanisms.

This has certainly been true for our knowledge of riboswitches, which have progressed from a few intriguing genetic elements to an important model system for exploring RNA structure-function relationships over the past twenty years. In the late 1990's, various conserved elements were found in the leader sequences of bacterial mRNAs and identified as regulators of gene expression of unknown mechanism, including the S-box by Tina Henkin, the *btuB* leader by Robert Kadner, and the *thi*-box by Juan Miranda-Ríos. Additionally, bioinformatic analyses of the rapidly increasing number of bacterial genomes uncovered others such as the *RFN* element in the leader of mRNAs associated with riboflavin biosynthesis by Mikhail Gelfand. In the absence of evidence for proteins interacting with these sequences to manage the regulatory response, such as in the *Bacillus subtilis trp* operon previously described by Charles Yanofsky, it was proposed in a prescient review by Gary Stromo that, like RNA aptamers raised through in vitro selection methods, these mRNAs could directly interact with small molecules. Ronald Breaker's group proved this hypothesis in 2002 for cobalamin binding to the *btuB* leader, followed by *S*-adenosylmethionine (SAM) binding the S-box by Tina Henkin, and thiamine-pyrophosphate (TPP) binding the *thi*-box and flavin mononucleotide (FMN) binding to the *RFN* element by Evgeny Nudler. These regulatory elements became known as riboswitches, formally defined as RNAs that bind small organic molecules or ions via an "aptamer domain" that instructs a second domain called the "expression platform" that in turn directs the expression outcome. By 2004, the first atomic-level pictures of how the aptamer domain of purine riboswitches binds their cognate effector emerged concurrently from my own laboratory and that of Dinshaw Patel.

Since these discoveries, riboswitches have emerged as an important model system in the RNA field. In a little over a decade, the number of these elements discovered and validated—the bulk of which has come from the work of Ronald Breaker and colleagues—has surged past twenty-five, with more in the pipeline. This has provided a wealth of feedstock for investigation of RNA structure, folding, and function. Structural analysis of riboswitch aptamer domains in complex with cognate ligand has kept pace with discovery such

that there is a crystal structure of a representative of almost every validated class. These structures currently comprise the majority of known small folded RNA domains (sequences of >200 nt that exhibit helical packing), substantially informing our understanding of RNA architecture. Their small size, high-fidelity folding, and easily observable activity make riboswitch aptamers ideal model systems for investigation of various biological and chemical properties of RNA using an increasing array of biochemical, biophysical, and computational tools.

Riboswitches are also emerging as important RNAs in human health and engineering of biological systems. Because of their ability to bind small molecules with high affinity and specificity and their almost complete uniqueness to bacteria, riboswitches have been proposed by Blount and Breaker to be an important new class of targets for antimicrobial therapeutics and are being employed as model systems for structure-based design of molecules targeting RNA. Indeed, they discovered that roseoflavin is a naturally occurring compound that acts by targeting the FMN riboswitch, suggesting that this is a practical route for the creation of new antimicrobials. The utility of riboswitches in synthetic biology is also beginning to emerge. Recent work has exploited these RNAs to select for bacteria that overproduce metabolites, reveal novel aspects of biosynthetic pathways, and as sensors of intracellular metabolites. Despite this rapidly growing body of research on riboswitches, there remain significant underexplored areas as well as unanswered questions. The following is certainly not a comprehensive list of what remains to be learned about riboswitches, but is a set of the questions that represents important avenues of research on these RNAs.

Do eukaryotes use riboswitches or other small-molecule binding RNA elements to regulate gene expression?

To date, there is only one riboswitch class observed eukaryotes: TPP. In plants, this riboswitch is found in the 3′-untranslated region and regulates alternative lengths of the transcript and thereby mRNA stability, while in fungi it is found within introns and controls alternative splicing. Given the well-established nucleotide conservation patterns in other bacterial riboswitch aptamers, other classes of riboswitches should have been discovered in eukaryotic genomes by now. However, discovery of small-molecule binding RNA elements eukaryotes may not be as simple of task as discovery in bacterial genomes. Indeed, only recently have small nucleolytic ribozymes been found to populate many eukaryotic genomes, despite their sequence and structures being well described. Further hope for functional ligand-binding RNAs in eukaryotes comes from the recent discovery of ATP aptamer sequences in the human genome in several introns by Andrej Luptak. Another possibility is that a completely different set of small-molecule binding RNAs have evolved in eukaryotes that await discovery and validation. Conversely, the advantages of riboswitch-mediated regulation in bacteria may not exist in eukaryotes (or even archaea) and thus this mechanism is not used to direct gene expression in these organisms.

How idiosynchratic is RNA architecture?

The wealth of structural information derived from riboswitch aptamers has substantially increased our understanding of how small multihelix RNA domains are organized. As Eric Westhof has pointed out, compact RNA folds are almost invariably built using a set of recurrent structural modules such as the kink-turn, tetraloops, and loop E motifs; riboswitch aptamers are no exception. However, for a number of these modules, their structural diversity remains poorly explored, which likely significantly impedes improvements in computational prediction and modeling of RNA secondary and tertiary structure. Recent work by David Lilley has aimed at a comprehensive understanding of the kink-turn module and its variations such as the kink-junction, in part using a SAM riboswitch aptamer as a host to explore the structure and stability of these variants. Using a similar strategy, other modules need to be examined with the same rigor enabling a more substantial toolbox of assembly units for computational modeling.

As more structures are unveiled, themes have clearly emerged as to how recurrent modules organize multiple helices into compact folds that host an active site. Two of the simplest solutions that have emerged are the pseudoknot and the three-way junction supported by a distal tertiary interaction. For example, the latter solution is observed in the purine, class-I cyclic-di GMP, TPP, and tetrahydrofolate (THF) riboswitch aptamers along with the hammerhead ribozyme. However, even within this seemingly simple fold, there is a maddening degree of variation in the details of their architecture. Put another way, why do paralogous RNAs appear to be rare in biology? Do we need to solve more RNA structures to observe them, or is it just difficult to evolve new functions in RNA through this evolutionary route? While this idiosyncrasy enables surprising RNA structures to continue to emerge, it frustrates efforts to easily predict higher-order RNA architecture or engineer new functional molecules. While certainly this would call for more unique structures to be determined, it can be argued that effort needs to be expended to broaden our understanding of the structures already solved, an avenue of research that many structural biology laboratories are loathe to undertake.

Can riboswitch aptamer domains act as ribozymes?

Riboswitches bind a number important protein cofactors involved in a broad spectrum of chemical reactions including TPP, FMN, coenzyme B_{12}, SAM, and THF. In theory, these RNAs could act as ribozymes. For example, in the SAM clan of riboswitches, the electrophilic methyl group of *S*-adenosylmethionine faces a solvent-filled channel in the

middle of the RNA. If the aptamer could bind and orient a second substrate molecule for a nucleophilic attack on the methyl group, this aptamer could become a methyltransferase. Similarly, the active moieties of THF and TPP are positioned in their respective aptamers in a configuration that could make it accessible to participate in chemical reactions as well. Despite the potential, catalysis by an RNA derived from riboswitch aptamers has yet to be demonstrated.

How does the expression platform sense the occupancy status of the aptamer domain?

While many riboswitch aptamer domains are extremely well characterized in isolation, analysis of their communication with the expression platform significantly lags. The common perspective of how riboswitches regulate gene expression is illustrated by an overly simplistic cartoon representation of the RNA in the ON and OFF states that highlight the two mutually exclusive secondary structures of the regulatory switch dictated by the binding status of the aptamer domain. There are a number of issues related to the details of this mechanism that have yet to be resolved and, in my opinion, represent the most important unanswered questions regarding riboswitches.

First, how do the binding properties of the ligand relate to the regulatory response? Groundbreaking studies by Wickiser, Crothers, and Breaker revealed that riboswitches are *kinetic* entities—that is, the regulatory response is dictated by the association and dissociation rates of ligand binding and not by equilibrium binding affinity. This is related to the fact that in most cases the riboswitch has only a limited timeframe in which to direct the expression machinery and the aptamer often cannot achieve full equilibrium before its occupancy status must inform the expression machinery. The disconnect between binding affinity and regulatory efficacy is most clearly illustrated by the THF riboswitch in which some compounds such 2,6-diaminopurine are able to bind more tightly to the aptamer than THF, but do not efficiently influence the downstream secondary structural switch. Furthermore, riboswitches are sensitive to a host of environmental parameters such as temperature, intracellular nucleotide pools, concentrations of cosolutes in the cellular environment, and small molecules that may compete with the cognate effector, all of which complicate the regulatory response. Indeed, this may be the main evolutionary advantage of riboswitches: the ability to integrate a number of physiological and environmental cues in a single regulatory molecule. Thus, many riboswitches are likely to have a complex regulatory landscape with their response dependent upon a number of influencing factors, not just effector concentration. However, this makes development of a clear model that relates intracellular effector concentration to a regulatory response elusive.

The second aspect of riboswitch mechanism that remains underexplored is the role of co-transcriptional folding. To date, the vast majority of our knowledge of RNA folding comes from studies of intact sequences in which folding is initiated by the addition of magnesium and acquisition of tertiary structure and ground state "native" structure is monitored. However, riboswitches (and likely many other RNAs in biology) have functional folds that are distinctly nonground state. Studies of purine riboswitches by the Lafontaine and Micura groups clearly showed that regardless of the occupancy status of the aptamer, the full-length riboswitch relaxes to a single fold at equilibrium. One of the functional states is a kinetic state—temporary, but long-lived enough to influence expression of the mRNA. Thus, the folding pathway of the RNA must be ultimately studied in the context of transcription, which is very difficult. Recently, a very elegant study by Frieda and Block was able to achieve the first co-transcriptional folding model of an RNA using force extension spectroscopy using an adenine-binding riboswitch to reveal some of the interplay of folding and function. The importance of co-transcriptional folding pathways is likely also to be relevant for understanding the structure of lncRNAs, whose very large size likely makes finding their energetic minimum nearly impossible in a biological context.

Can we apply our wealth of knowledge of riboswitches to real-world applications?

Over the past decade, synthetic biology has developed a strong interest in RNA-based devices. As exquisite sensors of small molecules in the cell, riboswitches have potentially a significant role in facilitating a variety of applications. For example, riboswitches have been employed as a tool to investigate cobalamin metabolism and transport (Yingfu Li) and to improve the biosynthetic production of lysine by *E. coli* (independently by Gyoo Yeol Jung and An-Ping Zeng). Moreover, the Jaffrey and Hammond groups have used riboswitch aptamer domains to the basis of fluorescent sensors of cellular metabolites, paving the way for new metabolomics applications. However, these efforts currently appear to be limited to the aptamers that nature has supplied us, or the few robust aptamers from in vitro selections such as tetracycline and theophylline.

The above begs the question as to whether all of our knowledge of natural riboswitch aptamers could be leveraged to inspire a new generation of artificial aptamers that are more apt to function robustly in the cellular context. One of the principal differences between aptamers obtained from in vitro selection/evolution methods and those observed in nature is that biological aptamers often have tertiary structure facilitating organization of the binding pocket. This tertiary structure is hypothesized to aid in the induced fit binding of the ligand by biasing the population of unbound RNAs toward binding-competent states, i.e., driving ligand recognition toward a "conformational selection" mechanism. The lack of this feature in artificial aptamers might be a substantial limitation

for many of these RNAs when placed in a cellular context. Is it feasible to take a page from the playbook of directed protein evolution and use natural riboswitch aptamers as scaffolds for a randomized selection of nucleotides that is the basis of an initial selection library? This question has only been sparsely explored by a few studies from the Szostak and Inoue groups. Not only might a scaffold derived from a riboswitch aptamer be able to host higher affinity and specificity binding pockets, but greatly facilitate the biochemical and structural characterization of resultant aptamers and coupling to other modules to create functional RNA devices. Ultimately, I believe that the greatest impact of our under-standing of riboswitches will be to illuminate new paths for engineering RNA devices that drive real-world biotechnology applications.

Acknowledgments