Riboswitches

DOI: 10.1002/anie.201208167

New Insights into Gene Regulation—High-Resolution Structures of Cobalamin Riboswitches**

*Marie F. Soulière, Andrea Haller, Tobias Santner, and Ronald Micura**

B_{12} cofactors · folding · gene expression · metabolites · RNA

Since the discovery of the first mRNA riboswitches around 2002, the concept that small molecules, mostly metabolites, directly bind to the nascent RNA chain and thereby trigger mutually exclusive folding pathways has been confirmed for many bacteria as well as for other organisms.[1] This interaction results in either up- or down-regulation of the corresponding gene.[2] The nature of ligands participating in such a strategy is diverse, ranging from enzyme cofactors and nucleotide precursors, to amino acids and ions. They include adenosylcobalamin (AdoCbl), thiamine pyrophosphate, adenine, guanine, 7-aminomethyl-7-deazaguanine ($preQ_1$), *S*-adenosylmethionine, *S*-adenosylhomocysteine, adenosine triphosphate, lysine, glycine, glutamine, tetrahydrofolate, flavin mononucleotide, cyclic diguanylate, glucosamine-6-phosphate, fluoride and magnesium ions, and most likely others will emerge in the future.

The sizes of metabolite-binding mRNA domains range from 34 nucleotides for $preQ_1$[3] to about 200 nucleotides for AdoCbl.[1a] Strikingly, for almost all above-mentioned riboswitches the three-dimensional structures of the aptamer domains bound to the dedicated ligands have been solved by X-ray crystallography at high-resolution and thus provide a platform to comprehend the molecular mechanism of these systems.[4] However, the 3′-adjoining expression platforms often elude attempts of crystallization, as many ligand-free aptamers do. This reflects the intrinsic dynamic nature of free riboswitches and is also in line with the fact that the expression platforms are much less conserved in sequence than aptamer sequences are.[5] The former have evolved to satisfy structural features that are required to act at different modes of gene expression, predominantly at transcriptional or translational levels, but also through splice-site control, mRNA decay, and likely other mechanisms yet to be discovered.

In this context, it is understandable that the largest and most complex riboswitch known to date (which was actually discovered as one of the first representatives, if not the first one)[1a,d] has resisted three-dimensional resolution and elucidation of its molecular mechanism for a decade. Very recently, the research groups lead by R. Batey[6] and A. Serganov[7] independently obtained high-resolution X-ray structures of cobalamin riboswitches and, together with a functional study by T. Pan[8] and co-workers earlier this year on the ligand-induced folding pathway, these investigations now provide detailed molecular insights into the response mechanism of mRNAs specific for cobalamins.

A first important finding in this recent work is that AdoCbl is not the only B_{12} cofactor derivative[9] that binds to this riboswitch family with high affinity. In cell-based assays, Batey and co-workers discovered that several cobalamin riboswitches from marine cyanobacterial and environmental (*env*) genomes from ocean-surface samples exhibit much higher affinities to methylcobalamin (MeCbl) and aquocobalamin (AqCbl), and clearly discriminate against AdoCbl (Figure 1). This distinction, which is directly related to the

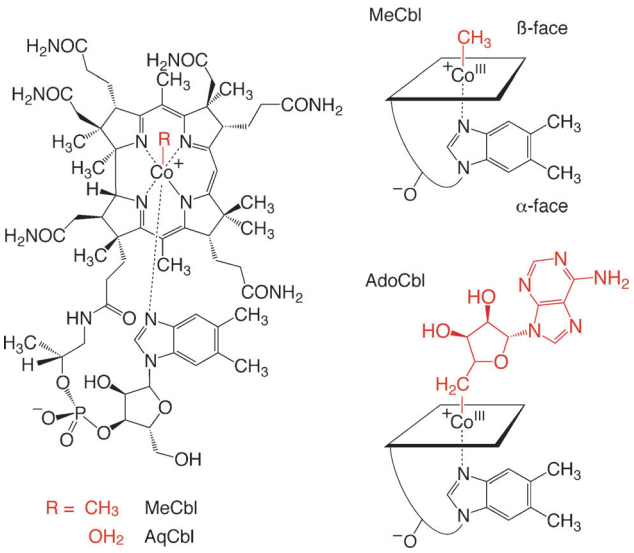

Figure 1. Structural formula of the cobalamins (Cbl) methylcobalamin (R = CH_3, MeCbl), aquocobalamin (R = H_2O, AqCbl), and coenzyme B_{12} (R = 5′-deoxy-5′-adenosyl, adenosylcobalamin, AdoCbl) and symbols used for these Cbls.

[*] Dr. M. F. Soulière, Mag. A. Haller, Mag. T. Santner, Prof. Dr. R. Micura
Institute of Organic Chemistry (IOC) and Center for Molecular Biosciences (CMBI), University of Innsbruck
Center for Chemistry and Biomedicine (CCB)
6020 Innsbruck (Austria)
E-mail: ronald.micura@uibk.ac.at

[**] We thank Robert T. Batey (University of Boulder) and Alexander Serganov (NYULMC) for providing their manuscripts and pdb files prior to publication.

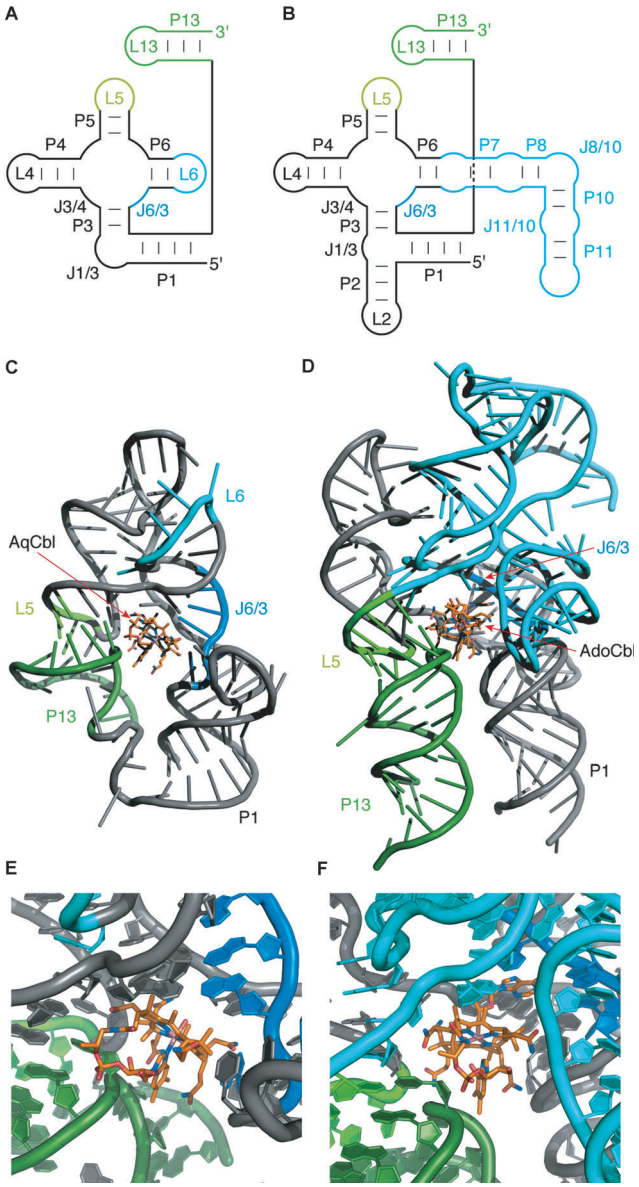

Figure 2. Cobalamin-sensitive riboswitches. Secondary-structure organization for the AqCbl- and MeCbl- (A) and for AdoCbl- (B) binding classes. C) Representation of the AqCbl-bound riboswitch *env8*AqCbl; D) Representation of the AdoCbl-bound riboswitch *Tte*. E, F) Enlargements of the AqCbl-binding pocket (E) and the AdoCbl-binding pocket (F).[6]

presence or absence of a peripheral extension (P8–P11; vida infra), gives rise to define the two classes of cobalamin riboswitches (Figure 2).[10] The two classes possess a common core built around an evolutionary conserved four-way junction responsible for cobalamin recognition. Also, both classes use a "kissing-loop" interaction between L5 of the core receptor and L13 of the accessory region which guides the expression machinery. For cobalamin riboswitches that regulate transcription, L13/P13 usually represents the antiterminator, while for those that regulate translation, L13 contains critical sequences that are involved in release or sequestration of ribosomal binding sites (RBS). The two cobalamin riboswitch classes differ by a large peripheral extension of P6 for the AdoCbl binding class (forming stems P7, P8, P10, and P11), and additionally by the addition of stem P2 between P1 and P3 (Figure 2). The P6 extension and stem P2 are absent or severely condensed for the MeCbl/AqCbl binding class.

Previous sequence-based and chemical-probing-supported predictions of secondary structures[1a,10,11] are correctly reflected in the 3D structures. Two of the newly solved structures concern AdoCbl-binding riboswitches. One of them is derived from a transcription-controlling system from *Symbiobacterium thermophilum* and the RNA investigated is composed of the receptor core and the peripheral P6 extension but lacks P13/L13.[7] This structure nicely complements the second structure of a translation-controlling, full-length system from *Thermoanaerobacter tengcongensis* (*Tte*) which contains the crucial regulatory element P13/L13.[6] In comparison to two structures of the much smaller AqCbl-binding riboswitch class (*env8*AqCbl), again one with and the other one without P13/L13, the modes of cobalamin/RNA recognition become apparent by this thorough structural foundation.[6]

The global architecture of all four ligand-bound RNAs reveals coaxial stacking of stems P1/P3/P6 and P4/P5/P13 (Figure 2). These two stacks are directly joined by L4/L6 interactions in the AqCbl class. The class-specific peripheral extensions (P6 versus P6–P11) adjoin the core and critically affect the conformation of J6/3, which is a key element for ligand recognition. This arrangement serves to form the cobalamin-binding cleft. Importantly, van der Waals surface complementarity between cobalamin and the RNA binding pocket is the major contributing force of the interaction. The ligand does not make use of the full hydrogen-bonding potential for RNA binding; instead hydrophobic packing and electrostatic interactions are most characteristic. In both classes, cobalamin is constricted between the minor grooves of P3/P6 and the helix molded by base pairing between L5 and L13. The corresponding structures of the RNAs that lack P13/L13 are readily superimposable to their full-length counterparts and implicate a relatively rigid scaffold of the open-cleft receptor domain that becomes closed by the docking of the regulatory P13/L13 element through the kissing-loop interaction to L5, thereby encapsulating cobalamin. Ligand discrimination between AqCbl and AdoCbl is handled by a dense stack of purines in the junctions (J3/4 and J6/3) of *env8*AqCbl that limits space and sterically interferes with large substituents at the cobalamin's ß-axial face. The corrin ring is oriented almost perpendicular to the bases of the purine stack. In contrast, the AdoCbl-binding class reveals conformational differences in J6/3 that become possible because of the peripheral extensions (P7–P11). Their complex interaction network enables the riboswitch to accommodate the adenosyl moiety of AdoCbl in a cavity through selective recognition of its Watson–Crick face by the Hoogsteen edge of a conserved adenosine in J6/3.

These new crystal structures not only provide molecular insights into the recognition mode between RNA and B_{12} cofactors but also reveal that this riboswitch critically involves tertiary structure contacts at the interface between the aptamer and expression platform to trigger the regulatory

outcome. In stark contrast, most other riboswitches exclusively employ altered patterns at the secondary-structure level. This is of significant interest because RNA tertiary structures tend to fold much more slowly than local secondary structures. This becomes even more important for B_{12} riboswitch species that involve mutually exclusive secondary-structure patterns in the expression platform (e.g. antiterminator versus terminator, or RBS interactions that intervene between sequestration and release through pairing options). Such riboswitch RNAs would be at risk of becoming stuck in either of the two mutually exclusive folding pathways, thereby losing capability for regulation.[2,5]

How the cell can resolve this problem has been shown by Pan and co-workers, who demonstrated that RNA polymerase pausing is a fundamental determinant during transcription of the AdoCbl-sensitive *Escherichia coli btuB* riboswitch.[8] The authors identified pause sites at strategic locations that facilitate the folding and structural rearrangement of the full-length riboswitch, but have minimal effect on the folding of the isolated aptamer domain. Pausing at these regulatory sites blocks the formation of alternative structures and plays a chaperoning role that links the folding of the aptamer domain and the expression platform. To be more precise, in the coenzyme B_{12} bound "gene-off" structure, the anti-anti-RBS region is part of the aptamer, allowing the pairing of the anti-RBS with the RBS regions (Figure 3A). In the apo or "gene-on" structure, the anti-anti-RBS region pairs with the anti-RBS region to form the anti-aptamer structure, leaving the RBS available for ribosome binding (Figure 3B). One of the three identified pause sites (P_A) is located just before loop L13. The nascent transcript outside of the RNA polymerase exit channel therefore directly corresponds to the successfully crystallized domains that lack the regulatory P13/L13 element. This portion of the transcript predominantly adopts the aptamer fold and pausing provides sufficient time to allow recognition of the cobalamin ligand, if present in sufficient amount. A successive pause site (P_B) has been identified that consecutively releases the P13/L13 element from the exit channel and thus furnishes time for docking and formation of the kissing-loop tertiary contacts between L5/L13 before the polymerase proceeds with transcription. The next pause site (P_C) seems to be most relevant for the required refolding into the anti-aptamer fold when no cobalamin ligand is available and the default folding pathway must be entered in order to achieve the opposite gene-regulatory response. Pausing was previously experimentally verified for flavin mononucleotide and thiamine pyrophosphate riboswitches[12a,b] and also proposed for others (e.g. purine riboswitches),[12c,d] and thus might be a general strategy for coping with cotranscriptional folding problems.

All three studies together draw a comprehensive picture of the response mechanism for the most complex riboswitch known to date. The combination of high-resolution structural data and functional assays reveal ligand-dependent RNA folding and refolding in the light of being a cotranscriptional process that depends on transcription rate and distinct pausing of the RNA polymerase.

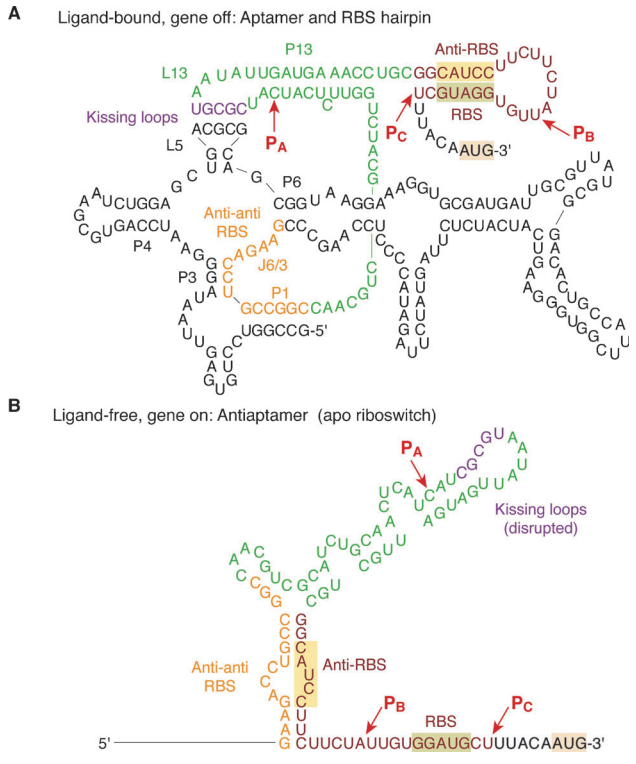

Figure 3. Proposed secondary structures of the "gene-off" (A) and "gene-on" (B) forms for the AdoCbl riboswitch from *E. coli btuB*. Three pausing sites during transcription were identified that are critical for folding, ligand recognition, and secondary-structure rearrangement of the expression platform if ligand concentration is insufficient.[9]

Received: October 10, 2012
Published online: January 7, 2013

[1] a) A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043–1049; b) A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, E. Nudler, *Cell* **2002**, *111*, 747–756; c) W. Winkler, A. Nahvi, R. R. Breaker, *Nature* **2002**, *419*, 952–959; d) X. Nou, R. J. Kadner, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7190–7195.

[2] a) K. E. Deigan, A. R. Ferré-D'Amaré, *Acc. Chem. Res.* **2011**, *44*, 1329–1338; b) A. Serganov, D. J. Patel, *Curr. Opin. Struct. Biol.* **2012**, *22*, 279–286; c) S. Blouin, J. Mulhbacher, J. C. Penedo, D. A. Lafontaine, *ChemBioChem* **2009**, *10*, 400–416; d) T. M. Henkin, *Genes Dev.* **2008**, *22*, 3383–3390; e) H. Schwalbe, J. Buck, B. Furtig, J. Noeske, J. Wöhnert, *Angew. Chem.* **2007**, *119*, 1232–1240; *Angew. Chem. Int. Ed.* **2007**, *46*, 1212–1219; f) J. C. Cochrane, S. A. Strobel, *RNA* **2008**, *14*, 993–1002; g) A. D. Garst, A. L. Edwards, R. T. Batey, *Cold Spring Harb. Perspect. Biol.* **2011**, DOI: 10.1101/cshperspect.a003533; h) A. Haller, U. Rieder, M. Aigner, S. C. Blanchard, R. Micura, *Nat. Chem. Biol.* **2011**, *7*, 393–400.

[3] a) A. Roth, W. C. Winkler, E. E. Regulski, B. W. K. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz, D. Iwata-Reuyl, R. R. Breaker, *Nat. Struct. Mol. Biol.* **2007**, *14*, 308–317; b) T. Santner, U. Rieder, C. Kreutz, R. Micura, *J. Am. Chem. Soc.* **2012**, *134*, 11928–11931.

[4] A. Serganov, *RNA Biol.* **2010**, *7*, 98–103.

[5] A. Haller, M. F. Soulière, R. Micura, *Acc. Chem. Res.* **2011**, *44*, 1339–1348.

[6] J. E. Johnson, Jr., F. E. Reyes, J. T. Polaski, R. T. Batey, *Nature* **2012**, *492*, 133–137.

[7] A. Peselis, A. Serganov, *Nat. Struct. Mol. Biol.* **2012**, *19*, 1182–1184.
[8] G. A. Perdrizet II, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3323–3328.
[9] K. Gruber, B. Puffer, B. Kräutler, *Chem. Soc. Rev.* **2011**, *40*, 4346–4363.
[10] A. Nahvi, J. E. Barrick, R. R. Breaker, *Nucleic Acids Res.* **2004**, *32*, 143–150.
[11] S. Gallo, M. Oberhuber, R. K. Sigel, B. Kräutler, *ChemBioChem* **2008**, *9*, 1408–1414.
[12] a) J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **2005**, *18*, 49–60; b) T. N. Wong, T. Pan, *Methods Enzymol.* **2009**, *468*, 167–193; c) J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, *Biochemistry* **2005**, *44*, 13404–13414; d) U. Rieder, C. Kreutz, R. Micura, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10804–10809.

