JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

NMR Structural Profiling of Transcriptional Intermediates Reveals Riboswitch Regulation by Metastable RNA Conformations

Christina Helmling,[†] Anna Wacker,[†] Michael T. Wolfinger,[‡,∥] Ivo L. Hofacker,[‡,§] Martin Hengesbach,[†] Boris Fürtig,[†] and Harald Schwalbe*[,†]

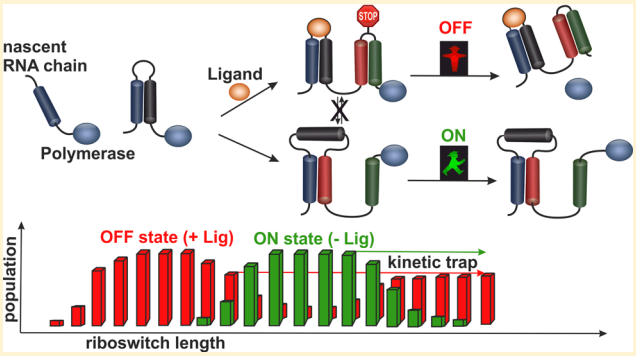
[†]Institute for Organic Chemisty and Chemical Biology, Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-Universität, Frankfurt/M. 60438, Germany

[‡]Department of Theoretical Chemistry and [§]Faculty of Computer Science, Research Group Bioinformatics and Computational Biology, University of Vienna, 1090 Vienna, Austria

[∥]Medical University of Vienna, Center for Anatomy and Cell Biology, Währingerstraße 13, 1090 Vienna, Austria

Supporting Information

ABSTRACT: Gene repression induced by the formation of transcriptional terminators represents a prime example for the coupling of RNA synthesis, folding, and regulation. In this context, mapping the changes in available conformational space of transcription intermediates during RNA synthesis is important to understand riboswitch function. A majority of riboswitches, an important class of small metabolite-sensing regulatory RNAs, act as transcriptional regulators, but the dependence of ligand binding and the subsequent allosteric conformational switch on mRNA transcript length has not yet been investigated. We show a strict fine-tuning of binding and sequence-dependent alterations of conformational space by structural analysis of all relevant transcription intermediates at single-nucleotide resolution for the I-A type 2′dG-sensing riboswitch from *Mesoplasma florum* by NMR spectroscopy. Our results provide a general framework to dissect the coupling of synthesis and folding essential for riboswitch function, revealing the importance of metastable states for RNA-based gene regulation.



INTRODUCTION

Structured parts in untranslated regions of bacterial mRNA can be involved in cis-regulation of the expression of adjacent genes at either the level of transcription or translation. In case of transcriptional regulation, the possible structures that a nascent mRNA chain can adopt are continuously changing as the mRNA increases in length due to the addition of nucleotides to the 3′-end of the RNA chain. During the sequential transcriptional elongation, mRNA can fold into kinetically trapped transient structures formed by those mRNA chains that emerge first from the RNA polymerase. Thus, cotranscriptional RNA folding and refolding is a sequential process[1] that occurs on the same time scale as RNA synthesis *in vivo*.[2] Kinetically trapped transient structures can have lifetimes sufficiently long to play an essential role in many RNA-mediated processes ranging from pre-mRNA splicing,[3−5] viral replication in eukaryotes,[6] and rRNA assembly[7] to gene regulation by riboswitches.[8−12]

Riboswitches represent an important class of regulatory RNA elements. Metabolite binding to transcriptional riboswitches can lead to either premature termination of transcription ("OFF switch") or continuation of transcription ("ON switch"). It has been shown that many transcriptional riboswitches regulate under kinetic control, where both ligand binding to the aptamer domain and the formation of secondary structures that either induce or escape termination (terminator/antiterminator)[13] are cotranscriptional events.[8] Mutually exclusive conformations responsible for the regulatory decision inevitably have to be adopted before the polymerase reaches the termination site. Transcriptional pausing further fine-tunes the regulation efficiency by prolonging the time window for the interconversion between folding intermediates.[9] Structures that regulate the expression of downstream genes do not need to be preserved in thermodynamic stable state(s) of the full-length riboswitch RNA, provided the time required for refolding of a metastable transient state is sufficiently long to escape termination. Only few studies have addressed the sequential process of nucleotide addition and the consequence of available RNA folding space during transcription,[8,14,15] despite the fact that the identification and characterization of folding intermediates is key not only to identify folding pathways but also to understand riboswitch-mediated regulation efficiency *in vivo*. Prior work by Breaker and co-workers describes the

Received: October 5, 2016
Published: January 30, 2017

ACS Publications © 2017 American Chemical Society

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

importance of ligand-binding kinetics in relation to the speed of transcription. According to this study, ligand binding must occur prior to synthesis of the expression platform and never reaches equilibrium during transcription.[16] This work is supported by fluorescence quenching studies on the *pbuE* adenine-sensing riboswitch, suggesting that the ligand does not bind to cause a structural rearrangement in the expression platform in full-length transcriptional ON switches.[17,18] These studies highlight the requirement of rapid ligand binding during transcription, but the folding pathway to form an interaction between aptamer domain and expression platform has not yet been investigated.

Here we aim to gain further insight into mechanistic details of cotranscriptional riboswitch regulation by addressing the precise programming in time regarding both ligand binding and folding rearrangements within the expression platform. We investigate the type I-A 2′-deoxyguanosine (2′dG)-sensing transcriptional riboswitch (dGsw) from *Mesoplasma florum* by NMR spectroscopy as an example for a transcriptional OFF riboswitch acting under kinetic control.[19−21] The downstream gene codes for the enzyme ribonucleotide reductase, converting ribonucleotides into deoxyribonucleotides. According to the current model of riboswitch function, ligand binding to the aptamer domain of the 2′dG switch represses gene expression of ribonucleotide reductase genes due to transcription termination.[19]

To understand how functional states, i.e., the ON and OFF states of dGsw, evolve during transcription, we mimicked the progress of transcription by synthesizing a total of 25 length variants of the riboswitch RNA aptamer domain and expression platform by applying a recently established NMR-based screening procedure.[22] A high-resolution picture of structures adopted by these transcriptional intermediates including a characterization of their ligand binding properties was obtained. The thus-identified lowest free energy states were used to describe thresholds for potential folding events and to reveal distinct sequential time frames for both ligand binding and rearrangements within the expression platform during transcription. Our results suggest that cotranscriptional folding into a kinetically trapped conformation allows transcription to continue and facilitates the ligand-dependent gene expression by extending the lifetime of an intermediate ligand-free structure. This crucial sequence-dependent RNA folding and refolding delays thermodynamic equilibration, allowing the polymerase to escape termination. The results further strongly support the view that riboswitches do not function as binary ON−OFF switches,[23] but finely modulate population ratios with every added nucleotide; in this conformational adaptation, further Mg^{2+}-induced stabilizations for some but not all of the transcription intermediates play an important role.

RESULTS

Characterization of the Full-Length 2′-Deoxyguanosine-Sensing Riboswitch. Within the thermodynamic model of transcriptional riboswitches, it is generally accepted that 4 segments of the mRNA undergo conformational rearrangements upon ligand binding (Figure 1). For OFF switches, the segments A and T form the antiterminator structure, leaving segments P and H accessible. Ligand binding stabilizes an interaction between segments P and A, allowing for formation of the terminator hairpin, TH. Such a model requires partial identity between strands P and T and between A and H. The 2′dG riboswitch applies a different scheme of conformational

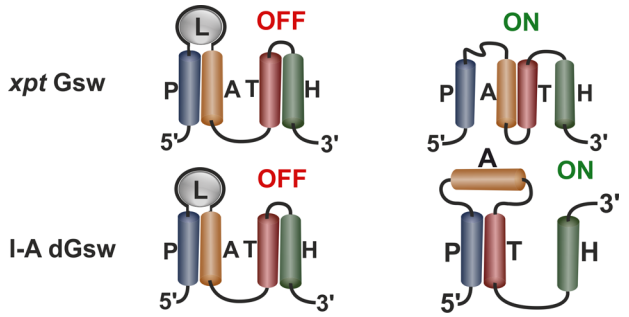


Figure 1. Schematic representation of the ligand-dependent rearrangement of mutually exclusive conformations comparing the I-A 2′dG riboswitch from *M. florum* to the *xpt* guanine-sensing riboswitch from *B. subtilis* as a prime representative of the purine riboswitch class. Structural rearrangements involve four distinct sequence segments in the mRNA chain: 5′-aptamer strand (P), aptamer-stabilizing strand (A), switching strand (T), and terminator stand (H).

rearrangements, where segments P and T form the antiterminator to leave segments A and H accessible. In analogy to the arrangement of the four segments determined for other riboswitches, ligand binding stabilizes the PA interaction to form the terminator hairpin TH. In this allosteric modulation, strand P shares identity with strand H, and strand A shares identity with strand T.

We analyzed the full-length 2′dG riboswitch ($dGsw^{FL}$) by NMR spectroscopy in the presence and absence of ligand. The presence of the terminator helix P4 (TH) in the full-length riboswitch was identified by comparison to a 40 nucleotide (nt) model terminator hairpin (nts 104−144), characterized and described in Figure S1. Signals in the ^{15}N-TROSY spectrum (Figure 2A) assigned to the terminator stem TH (red and green) are present in the absence of ligand and do not shift with the addition of ligand. Additional signals and chemical shift perturbations in the presence of ligand can be assigned to arise exclusively from the aptamer domain, which has been characterized by NMR spectroscopy in detail in a previous study,[20] and forms a more compact three-dimensional structure with tertiary interactions in the ligand-bound form (Figure 2A).

Chemical shift perturbations induced by ligand addition suggest that outside the aptamer domain, base-pairing interactions are neither formed nor disrupted in response to ligand binding. This finding strongly suggests that the conformation containing a fully formed terminator stem (PA-TH) is the thermodynamically favored conformation regardless of ligand binding, but raises the question of how transcription proceeds past the terminator sequence in the absence of ligand.

To characterize the secondary structure of the genetic ON state, $dGsw^{FL}$ was truncated by strand H to allow formation of the antiterminator helix PT in a 122 nt transcriptional intermediate ($dGsw^{ON}$). Helices P2 and P3 could be unambiguously identified in the ON state by comparison to $dGsw^{FL}$. Although complicated by the size of the RNA and the presence of more than a single conformation, we were able to characterize the remaining helical segments only after reassembling the construct from fragments with different ^{15}N-labeling patterns (Figures 2B and S2−S4). Inter- or intrastrand NOE contacts can be distinguished within these constructs by ^{14}N and/or ^{15}N X-filtered NOESY experiments, as shown for one example in Figure 2B. The structural integrity of assembled constructs was verified by spectral comparison to $dGsw^{ON}$. All three assembled constructs exhibit the same secondary

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

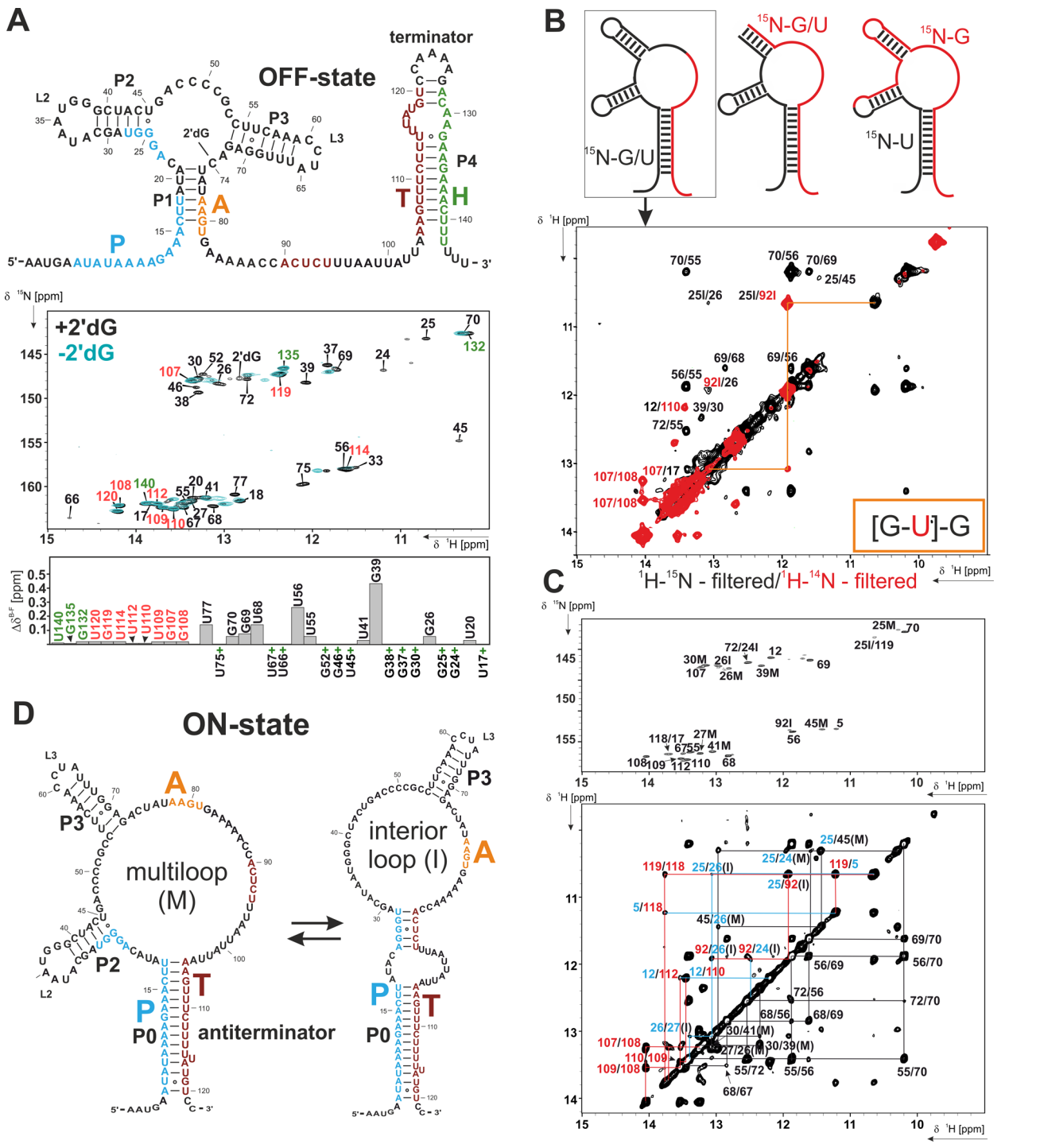


Figure 2. (A) Secondary structure of the OFF state and corresponding ^{1}H,^{15}N-correlation spectra recorded in the absence (200 μM RNA, 12 equiv of Mg^{2+}, cyan) and presence (400 μM RNA, 12 equiv of Mg^{2+}, 2 equiv of 2′dG, black) of ligand. Signals corresponding to the terminator stem (P4) are highlighted in red (strand T) and green (strand H). The spectra were recorded at 800 MHz and 298 K. $\Delta\delta^{B-F}$, chemical shift difference of imino proton signals between the ligand-bound and -free riboswitch forms; green + , imino proton signals appearing only after ligand binding. (B) Strategy for secondary structure assignment of $dGsw^{ON}$ assembled from fragments with different isotope labeling patterns including one exemplary overlay of NOESY spectra ^{14}N-X-filtered (red) and ^{15}N-X-filtered (black) in the direct dimension for the selected construct $dGsw^{0-85-122}$ (400 μM RNA, 900 MHz, 283 K). (C) Assigned ^{15}N-TROSY and NOESY spectrum (750 μM RNA, 800 MHz, 283 K) highlighting NOE contacts between strands P (red) and T (blue) and (D) secondary structures in conformational exchange in $dGsw^{ON}$ (M, multiloop antiterminator conformation; I, interior loop antiterminator conformation).

structure with minor changes regarding artificially implemented changes and local helical destabilizations (Figure S3). The three reconstructed ON states reveal a base-pairing interaction of G24, G25, and G26 that competes with helix P2 formation. This result suggests the existence of two antiterminator conformations in slow exchange: multiloop (M) and interior loop (I) conformations in Figure 2C,D. In both conformations, helix P3 is present, and the antiterminator helix is formed by an

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

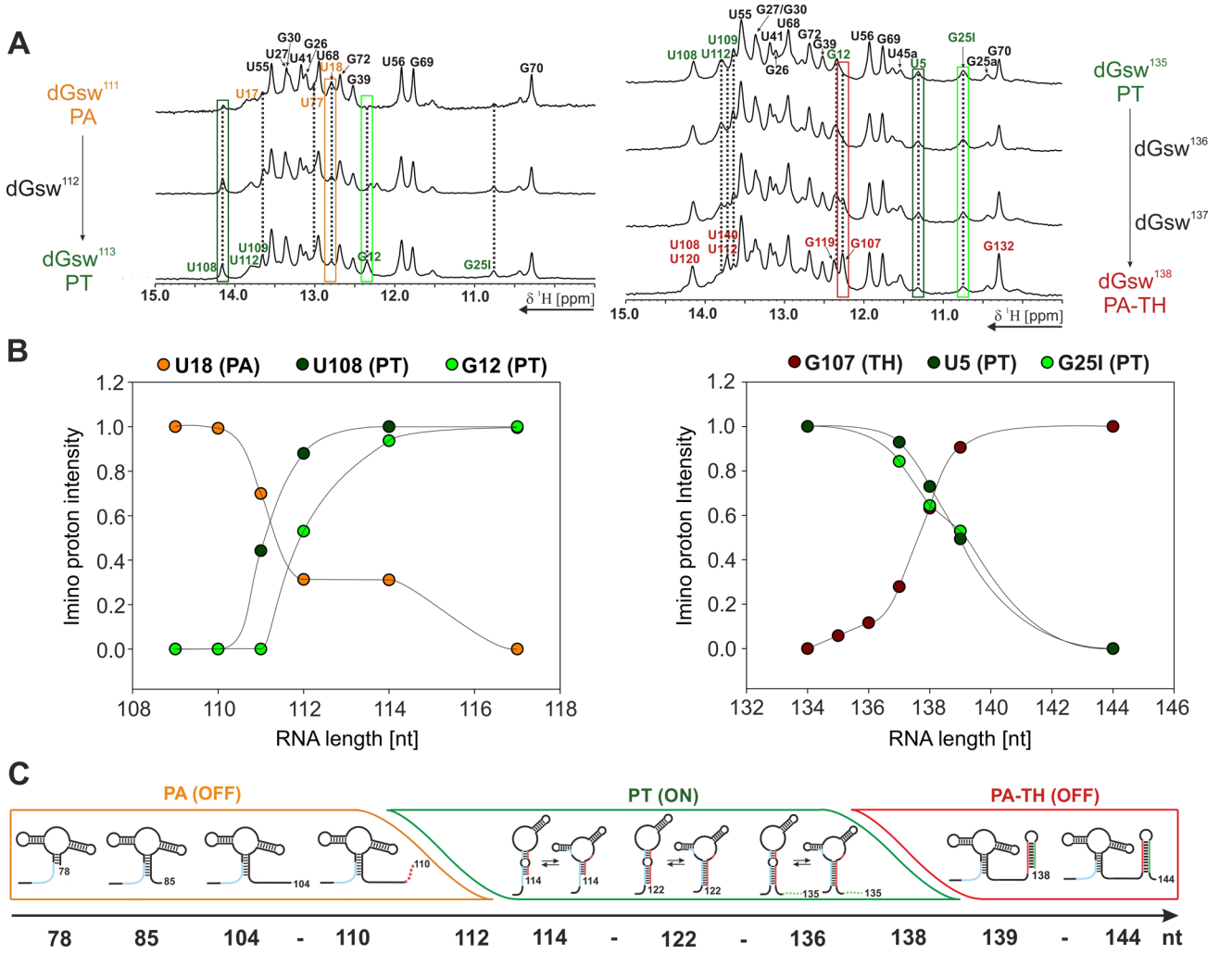


Figure 3. Conformational mapping of transcriptional intermediates in the absence of ligand (Gene expression: ON). (A) 1D NMR spectra of transcriptional intermediates at the transition from helices PA (aptamer domain) to PT (antiterminator helix) (left) and from helices PT to TH (terminator) (right). The samples contain 100 μM RNA and 8 equiv of Mg^{2+} in NMR buffer. The spectra were recorded at 800 MHz (left) and 700 MHz (right) at 298 K. (B) Normalized imino proton intensity of reporter signals for aptamer (PA, orange), antiterminator (PT, green), and terminator (TH, red) as color-coded in panel A with increasing RNA length. (C) Equilibrium- state secondary structures of transcriptional intermediates.

interaction between strands P and T. The multiloop conformation (M) contains helix P2, and only P1 is disrupted. In the interior loop conformation (I), P2 is disrupted to elongate the antiterminator helix PT. The existence of the interior loop conformation was unambiguously verified by reassembling the elongated PT helix in a model construct (Figure S4).

Structural Profiling of Transcriptional Intermediates in the Absence of Ligand (Gene Expression: ON). We investigated which structures form as the riboswitch sequence is extended by single nucleotides by screening the secondary structure adopted by 25 transcriptional intermediates. These RNAs were purified under native conditions but are likely to have reached structural equilibrium by the time of measurement.[22] We identified (with single-nucleotide resolution) at which length the mRNA switches between the helical segments PA, PT, and TH (Figure 3A). Transcriptional intermediates only adopt these three distinct secondary structure elements. Alternative base-pairing interactions could not be detected by NMR (Figure S5). We selected isolated imino proton signals and analyzed the change in intensity upon addition of single nucleotides (Figure 3B). The aptamer domain was identified by PA reporter signals U17 and U18, which dominate from $dGsw^{77}$ to $dGsw^{110}$. $dGsw^{111}$ and $dGsw^{112}$ show reporter signals for both mutually exclusive PA and PT helices (highlighted in orange and green in Figure 3A, respectively). Imino proton spectra of $dGsw^{113}$ up to $dGsw^{136}$ show reporter signals of the antiterminator conformation only. In the absence of ligand, addition of Mg^{2+} to these constructs does not cause shifts in the imino proton pattern. Mg^{2+} alone therefore cannot induce a transition from PA to PT and from PT to PA-TH (Figure S6). For constructs that adopt the PA interaction, addition of Mg^{2+} causes an increase in imino proton signals, which can be assigned to conformational stabilizations, particularly regarding the aptamer domain according to reported data.[20]

We tested how well the experimental findings could be predicted by folding simulations of the 2′dG-sensing riboswitch using the previously described barmap approach.[24] Simulations were performed at 298 K in accordance with the experimental setup. To reduce the computational effort, the sequence was truncated by the first five 5′-nucleotides and compared to the structures previously defined as full-length riboswitch, since

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

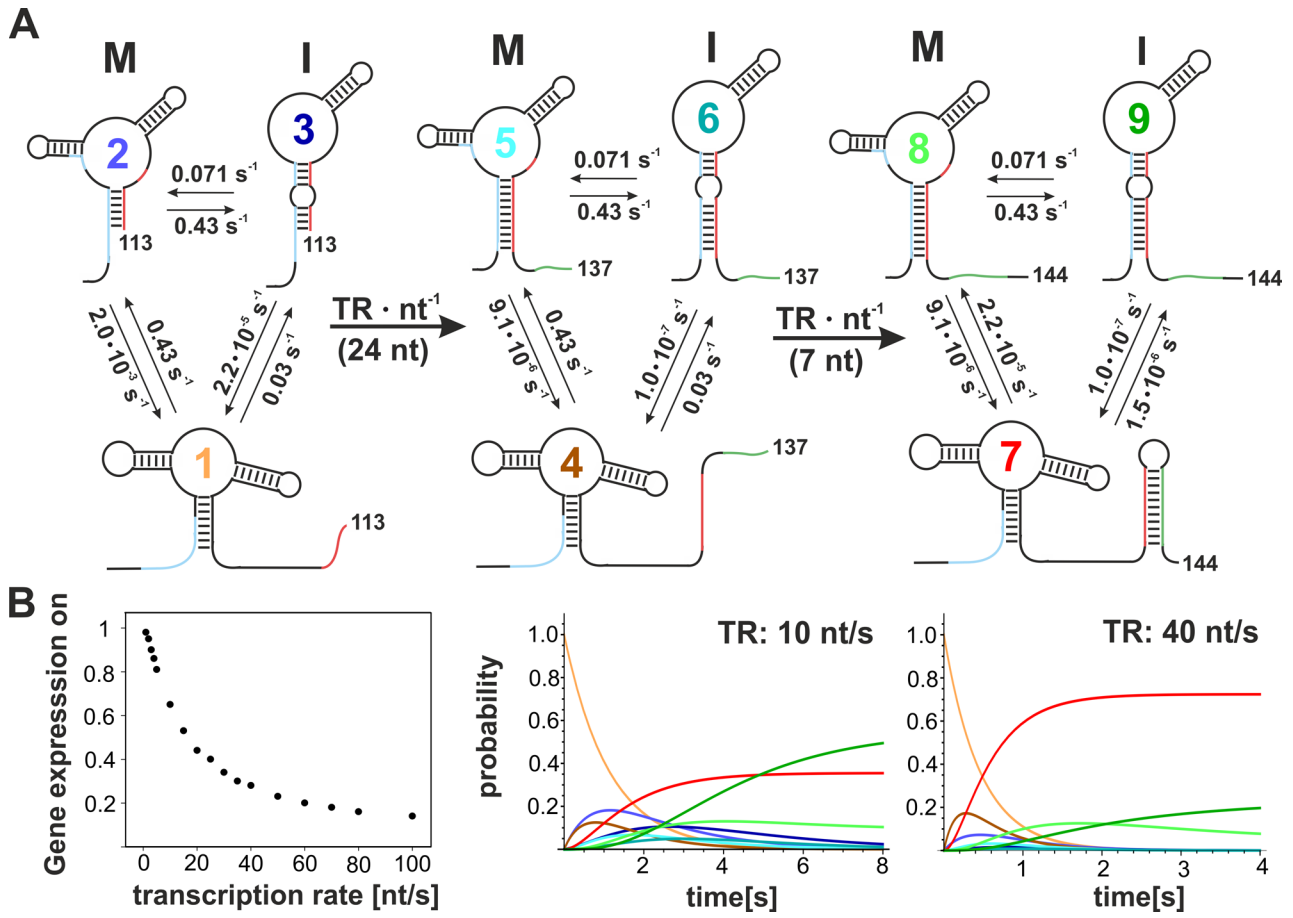


Figure 4. (A) Theoretical model for Markovian simulations of cotranscriptional folding in the absence of ligand assuming 9 states including three distinct RNA lengths: 113, initiation of antiterminator folding (states 1−3); 137, the antiterminator folding pathway can no longer be entered (states 4−6); 144, point of regulatory decision (states 7−9); and three distinct RNA conformations: PA/PA-TH (states 1, 4, and 7); PT, conformation M (states 2, 5, and 8); and PT, conformation I (states 3, 6, and 9). (B) Fraction of the ON state at the regulatory decision point depending on the rate of transcription (left) and exemplary simulation at a transcription rates of 10 and 40 nt/s (right).

these nucleotides were not involved in any observed transient structures. As shown in Figure S7, an RNA chain length of around 50 nt is required for structure formation ($dGsw^{50}$) (P2 and P2* helix). For mRNA chains with lengths between 75 and 90 nt ($dGsw^{75}$ to $dGsw^{90}$) the aptamer structure dominates the structural ensemble. The aptamer domain remains the dominant structure until ~$dGsw^{110}$, but folding predictions suggest the formation of diverse alternative structures in addition to structures maintaining the aptamer fold. At $dGsw^{110}$, the PA segment within the aptamer domain is disrupted to form the antiterminator helix PT until $dGsw^{139}$. The genetic ON state is therefore predicted to maintain thermodynamic stability for ~29 nt. At 140 nt, the most stable RNA conformation is the terminator conformation (PA-TH) representing the ground state of the full-length RNA. The predicted antiterminator conformation is remarkably similar to the experimentally determined multiloop conformation (M) with the exception of helix P5 that cannot be observed by NMR but may not be detectable due to an unfavorable dynamic window. In contrast, the experimentally observed interior loop conformation (I) is not predicted by the simulation.

On the basis of the experimental data, we further simulated the refolding in the expression platform during transcription in a 9-state model (Figure 4). The theoretical model describes the rearrangement from PA (states 1 and 4), which is always transcribed and folded first, to PT (states 2, 3, 5, and 6). Refolding takes place as the mRNA is elongated from 113 to 137 nt, the kinetic window for ON state folding derived from experimental data shown in Figure 3, at the rate of transcription (TR). The RNA chain is then further elongated to the point of regulatory decision (144 nt) to either terminate (state 7) or allow transcription to proceed (states 8 and 9). We calculated refolding rates between all long-lived states observed by NMR based on refolding rates previously derived.[25] In short, real-time NMR experiments have shown an exponential dependence of the number of base pairs required to dissociate on the lifetime of secondary structures in interconformational conversion equilibria. Calculated folding rates from aptamer (PA) to antiterminator (PT) of $k = 0.43\ s^{-1}$ are within the order of the rate by which the subsequent 24 nt are transcribed (12−90 nt/s for *E. coli*).[22−24] In fact, refolding rates to form the antiterminator helix PT do not vary with the length of the RNA transcripts, since initiation of folding always requires dissociation of the helical segment PA as rate limiting step. Figure 4B shows the fraction of nonterminated transcripts in relation to the rate of transcription derived from these cotranscriptional folding simulations. Slow transcription rates extend the 24 nt window of opportunity to fold from PA to PT and therefore increase the amount of transcripts that allow transcription to proceed. At a transcription rate of 10 nt/s, ~35% of the transcripts terminate transcription within the synthesis of nucleotides 113−144 (3.1 s), and antiterminator

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

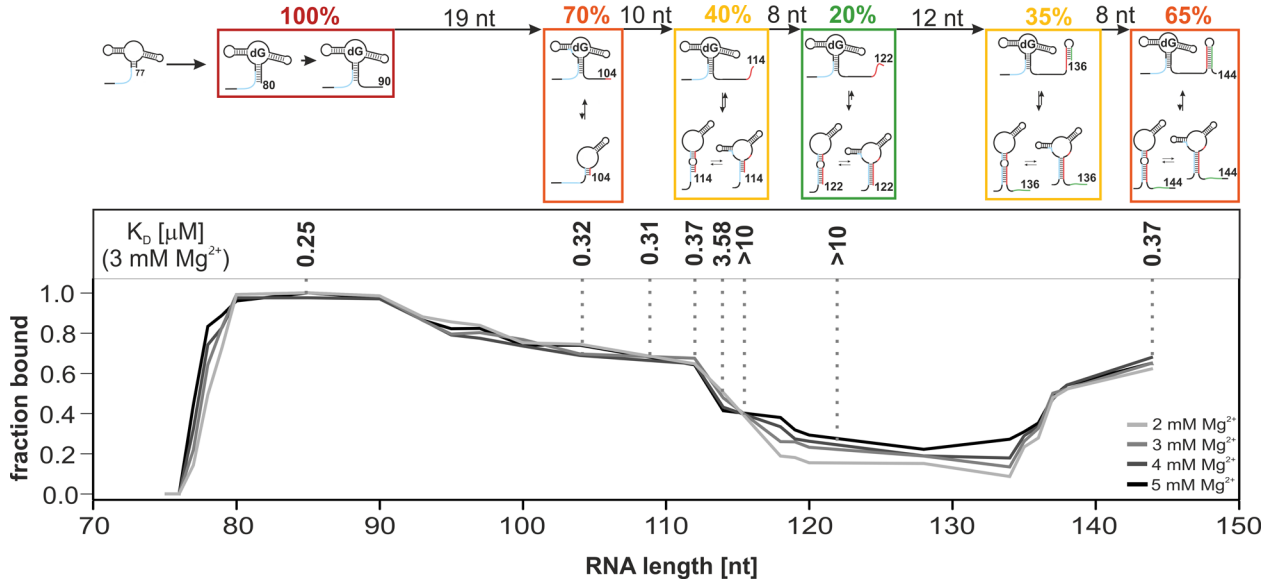


Figure 5. Ligand-bound fraction of dGsw depending on the RNA length. The *y*-axis shows the normalized integral of the imino proton of bound ^{15}N-labeled 2′dG (1 mM) in the presence of ∼500 μM unlabeled RNA for transcriptional intermediates at 2 mM (white), 3 mM (light gray), 4 mM (dark gray), and 5 mM (black) Mg^{2+}. X-filtered spectra to resolve the imino proton of 2′dG were recorded at 600 MHz and 298 K. A graphical representation of secondary structures corresponding to equilibrated transcriptional intermediates in the presence of ligand is shown above. K_D values determined by ITC at 3 mM Mg^{2+} for eight selected transcriptional intermediates are indicated accordingly.

conformations M and I are not at thermodynamic equilibrium. In fact, calculated rates suggest fast folding to antiterminator conformation M followed by 10-fold slower equilibration to antiterminator conformation I. Once the antiterminator conformation M is adopted, refolding back to form the aptamer domain is slow (2.0×10^{-3} s^{-1}), and this rate further decreases with increasing RNA length to 9.1×10^{-6} s^{-1}.

Ligand Binding by Transcriptional Intermediates (Gene Expression: OFF). We assessed ligand binding to each transcriptional intermediate to trace the ligand-binding-competent segment on the mRNA chain and to evaluate if ligand binding remains constant as the mRNA is extended by single nucleotides. The fraction of the ligand-bound state was determined for each transcriptional intermediate by quantification of the imino proton NMR reporter signal of the ^{15}N-labeled ligand 2′dG (Figure 5). The population of the ligand-bound state exceeds 80% for transcription intermediates $dGsw^{80}$–$dGsw^{90}$. For transcription intermediates of more than 93 nt, ligand binding decreases up to a chain length of ∼134 nt. Between $dGsw^{113}$ and $dGsw^{117}$, ligand-binding affinity decreases to ∼20–30%, in line with strand T synthesis to form the antiterminator helix PT. Similarly, K_D values obtained from ITC of eight selected transcriptional intermediates show a decrease in affinity between $dGsw^{85}$ and $dGsw^{104}$ from 250 to 370 nM and for $dGsw^{114}$ to 3.6 μM (Figure S9). The K_D's for $dGsw^{117}$ and $dGsw^{122}$ exceed 10 μM and could not be reliably determined by ITC. Affinities increase again along with strand H synthesis to stabilize the competing terminator helix TH in longer transcripts. However, ligand binding only reaches 65% in the full-length riboswitch (PA-TH) in comparison to the 100% binding for the aptamer domain (PA) alone. The impact of Mg^{2+} addition on ligand binding is strongest for aptamer constructs with a truncated helix P1, in agreement with a Mg^{2+}-induced preorganization of the aptamer domain toward the ligand-bound state,[20,26] which promotes ligand binding. In contrast, once the aptamer domain is fully transcribed to form the complete PA interaction, significant Mg^{2+} dependence of ligand binding can no longer be detected. For all constructs containing the antiterminator helix PT ($dGsw^{112–136}$), an increase in the Mg^{2+} concentration shifts the equilibrium toward the ligand-bound state; however, this is only up to 30–40% in the presence of 5 mM Mg^{2+} and $[Mg^{2+}]$:[RNA] = 10:1.

DISCUSSION

Allosteric Conformational Switch. We characterized the antiterminator (P^AT) and terminator conformation (PA-TH) of the 2′dG-sensing riboswitch, showing that the allosteric conformational switch differs for dGsw compared to other members of the purine-sensing riboswitch class. Typically, the switching sequence, defined as sequence shared between aptamer domain and expression platform, corresponds to strand A and is located at the boundary between aptamer domain and expression platform. In this design of allosteric states, the separation in nucleotides between the two complementary strands A and T is small; for the *xpt* guanine-sensing riboswitch from *B. subtilis*, for example, it does not exceed 33 nt (Figure 1A).[27] By contrast, in the 2′dG-sensing riboswitch studied here, strand P corresponds to the switching sequence and pairs with segment T to form the antiterminator helix, leading to a larger separation in nucleotides between complementary strands. In the *xpt* riboswitch, 5′–3′ synthesis of strand T allows association of part of the antiterminator helix AT during transcription prior to PA melting. If strand P is involved, as it is the case for dGsw, then either melting of PA precedes PT association or the strand T needs to be fully synthesized to initiate refolding.

Furthermore, dGsw exhibits structural heterogeneity in the apo state, also found previously for the three-state *add* adenine-sensing riboswitch.[28] Our NMR analysis of $dGsw^{ON}$ strongly supports the existence of two slowly interconverting, long-lived ON states leading to gene expression. Interior loop conformation I also dissociates helix P2 in addition to helix P1 to form the antiterminator helix PT. Kinetic rate calculations suggest fast folding to the multiloop conformation, disrupting

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647–2656

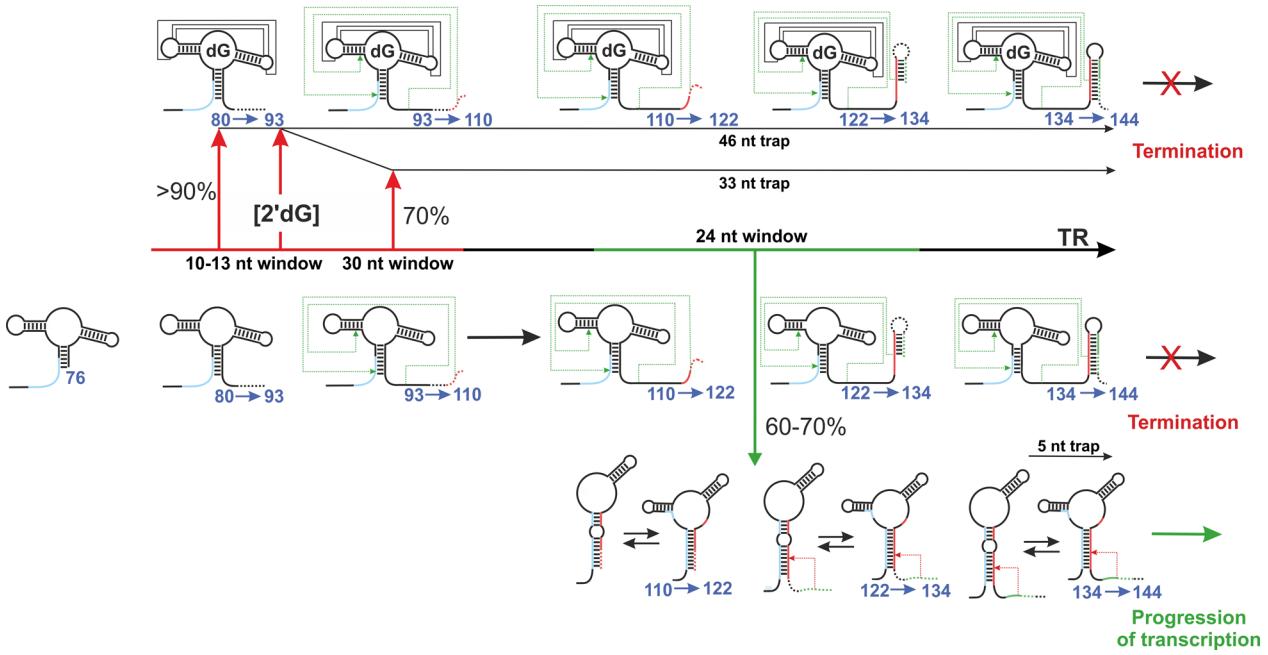


Figure 6. Schematic representation of the transcriptional progress and potential folding pathways. Transcriptional regulation by riboswitches is described by two distinct folding windows occurring at different stages during mRNA transcription. The window of opportunity for ligand binding always occurs early during transcription directly after the aptamer domain is transcribed (red pathway). High riboswitch OFF function (>90%) requires ligand binding to the aptamer domain within the transcription of 10−13 nt followed by kinetic trapping of the ligand-bound state for the transcription of the subsequent 51 nt. Thermodynamic equilibration during the transcription of $dGsw^{93}$−$dGsw^{110}$ leads to a loss of riboswitch function. To achieve at least 70% termination at the regulatory decision point, the ligand-bound aptamer domain must be kinetically stabilized during the transcription of 33 nt (~$dGsw^{111}$−$dGsw^{144}$). Antiterminator folding occurs at a later state during transcription provided that the aptamer domain is not stabilized by ligand binding. Folding can be initiated from $dGsw^{113}$ on and must take place within the transcription of the following 24 nt. Transcription termination assays suggest that ~60−70% of the transcripts fold to the antiterminator within this time frame. If the antiterminator does not fold during synthesis of nucleotides 113−137, then the terminator will fold to abort transcription even in the absence of ligand. Solid black lines represent tertiary interactions within the aptamer domain. Green arrows represent base-pairing interactions competing with displayed secondary structures.

only helix P1, followed by a 10-fold slower folding into the interior loop conformation, disrupting also helix P2. According to cotranscriptional folding simulations, the time frame of transcription is too short to allow full equilibration between the two antiterminator conformations.

ON Function. According to both NMR analysis and structure predictions, the terminator conformation (PA-TH) represents the lowest free energy state of $dGsw^{FL}$ even in absence of ligand. This finding requires that in the folding pathway of the mRNA without ligand the antiterminator conformation ($P^{A}T$) is a metastable state that has to be kinetically trapped during transcription. Equilibration from the antiterminator conformation $P^{A}T$-H to the terminator conformation PA-TH occurs only past the point of transcriptional termination and has no effect on the genetic decision.

Screening of potential folding pathways through transcriptional intermediates by NMR suggest that the aptamer folds first and retains stability until $dGsw^{110}$. This picture is supported by theoretical folding simulations, according to which the aptamer domain represents the dominant structure within this segmental length. In addition, the simulations propose the formation of additional states, which may be present but are not detectable by NMR. The alternative aptamer conformation containing a helix P2*, as found in structure predictions, cannot be detected in solution by NMR. Typically, helices P2 and P3 are perfectly tuned to allow parallel alignment for loop−loop interactions, which cannot easily be accounted for in structure predictions. Experiments and predictions provide very similar results, i.e., the thermodynamic equilibrium shifts from PA to PT at ~$dGsw^{113}$, and this antiterminator state $P^{A}T$ is stable over a ~23−24 nt segment. In the context of transcription, we conclude that an mRNA chain of at least 113 nt has to be synthesized before refolding from PA to PT can be initiated (Figure 6, green pathway). Since the terminator conformation (PA-TH) is the most stable conformation in the full-length construct even in absence of ligand, it is imperative for this refolding event to occur during transcription of nucleotides 113−137; otherwise, the genetic decision would inevitably be OFF. According to reported transcription rates for bacterial polymerases at 37 °C (12−90 nt/s),[29−31] the window of opportunity for antiterminator folding would be limited to 300 ms to 2 s. Transcription assays performed in the absence of ligand with *E. coli* polymerase show a fraction of ~70% of the RNA elongating to the full-length RNA *in vitro* (Figure S8). Combining this result with theoretical rate calculations of conformational interconversions suggests a transcription rate of 8−12 nt/s, which is at the lower end of reported transcription rates. However, *mfl* polymerase operates at lower temperatures (~25 °C) which may account for the differences.

It requires 139 nt to form a stable terminator hairpin TH. Since termination of transcription occurs while a poly-U stretch located between $dGsw^{140}$ and $dGsw^{145}$ is synthesized (Figure S8),[19] refolding to the terminator conformation PA-TH from the antiterminator state $P^{A}T$-H cannot be initiated before nucleotide 139 is released from the 8−10 bp long RNA/DNA

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

duplex, formed within the transcription bubble.[32,33] Therefore, termination must take place after the transcription of two additional nucleotides. Combining these results with theoretical rate calculations, which suggest very slow refolding rates from the antiterminator conformation $P^{A}T$-H to the terminator conformation PA-TH in the range of $10^{-4}-10^{-6}$ s^{-1}, our results clearly explain that the polymerase escapes the termination site once the antiterminator helix PT is formed. This implies that ligand binding has to stabilize the helical segment PA prior to strand T synthesis to avoid formation of the antiterminator helix PT.

OFF Function. Ligand binding is most effective directly after strand A is transcribed to form the complete aptamer domain including the subsequently synthesized 10 nt ($dGsw^{80}-dGsw^{90}$). By NMR, no alternative structure that would inhibit ligand binding can be detected before the antiterminator conformation dominates from ~$dGsw^{112}$ on. Nevertheless, ligand binding continuously decreases by ~30% before strand T to form the antiterminator helix PT is even synthesized.

The time point during transcription at which ligand has to mandatorily bind depends on the required extent of regulation. To achieve over 80% switching to a regulatory OFF state (at saturating 2′dG concentrations), the ligand must bind to the aptamer within the transcription of 10−13 nt to lock the PA interaction for the synthesis of the following 51 nt. If 60−70% of mRNA termination is sufficient for the bacterium to vary the ribonucleotide reductase concentration according to its demand, then the window for ligand binding is extended from 10−13 nt to about 30 nt, and the kinetic trap would be reduced from 51 to 34 nt (Figure 6, red pathway). *In vitro* transcription termination assays show a half-maximal modulation of transcription termination at [2′dG] = 2 μM, with the maximal degree of 60% termination reached at a concentration of ~500 μM.[19] This *in vitro* determined termination efficiency is in remarkable agreement with the degree of ligand-binding determined by our screening approach. *In vitro*, the effect of RNA-folding proteins is not taken into account. In addition, not only the transcription speed of the native RNA polymerase but also spatial restraints during transcription may differ *in vivo*,[34] which can affect the attainable degree of termination. However, our results support the idea that regulatory function by riboswitches may represent not a binary ON−OFF switch[23] but rather vary the mRNA concentration to sufficiently adapt the gene product concentration to the cellular demand. The seemingly limited dynamic range of the riboswitch of 60−70% might be essential so that the riboswitch responds to concentration changes of only one of the four substrates of ribonucleotide reductase, and full deactivation of enzyme expression may even be lethal.

CONCLUSION

In summary, our data support the emerging view that folding of riboswitches, for that matter, or any cis-acting RNA, is a process that has to be assessed within the kinetic framework of transcription.[8,11,14,34] Our approach shows that the identification and characterization of intermediate structures is essential for understanding the coupled kinetic and thermodynamic folding pathway and ultimately the mechanism of regulation.[14,35] Identification of transcriptional intermediates and their structural and functional mapping with nucleotide resolution, as shown in this study, provides a basic concept of how folding pathways can evolve and narrows down distinct windows in the context of transcription in which regulatory decisions must be made (Figure 6).

MATERIALS AND METHODS

RNA Preparation. All RNA constructs were prepared by *in vitro* transcription with T7 polymerase from PCR products as reported previously.[22] As DNA template for PCR, a plasmid containing a 5′-Hammerhead and the sequence of $dGsw^{144}$ was used. Templates for $dGsw^{75}-dGsw^{144}$ were prepared by adjusting the reverse primers accordingly. For rapid conformational screening of $dGsw^{75}-dGsw^{144}$, the RNA was purified under native conditions by buffer exchange in centrifugal concentrators.[22]

For NMR secondary structure characterization of the allosteric conformational switch, $dGsw^{144}$ and $dGsw^{122}$ were purified by PAGE with 10% polyacrylamide gels (29:1 (w/w) acrylamide/bis-(acrylamide), 7 M urea). The RNA was excised from the gel and eluted with 0.6 M NaOAc (pH 5.5). The RNA was first precipitated with 5 volumes of ethanol (−80 °C) and then twice with 5 volumes of 2% (w/v) LiClO4 in acetone. The RNA was denatured at high concentrations (0.2−0.5 mM) and folded by dilution with 5 volumes of water (0 °C) and incubation on ice for 1 h. Folded RNA was exchanged into NMR buffer (25 mM potassium phosphate, 50 mM potassium chloride, pH 6.2).

Transcriptions assays with *Escherichia coli* polymerase were performed using the T7A1 promotor. Transcription reactions contained 5 vol % PCR product, 0.1 U/μL *E. coli* RNA polymerase holoenzyme (NEB), 150 mM KCl, 10 mM $MgCl_2$, 0.01% Triton X-100, 500 μM each NTP, 5 μCi [γ-^{32}P]UTP, and 1 mM DTT in 40 mM Tris·HCl (pH7.5) in presence and absence of 200 μM 2′dG. Reactions were incubated for 2 h at 37 °C and analyzed by 10% PAGE. Gels were visualized using a PhosphorImager and analyzed with the software ImageQuaNT.

Screening of Ligand Binding. First, 1 mM ^{15}N-labeled 2′dG was added to ~500 μM samples obtained from buffer exchange of 2.5 mL transcriptions. ^{1}H−^{15}N-filtered experiments were recorded to monitor the ligand imino proton signal integral at varying Mg^{2+} concentrations (2, 3, 4, and 5 mM).

NMR Spectroscopy. NMR experiments were recorded on a Bruker AVIII HD 600 spectrometer equipped with a 5 mm *z*-axis gradient TXI-HCN cryogenic probe prodigy and on AV700, AV800, and AV900 spectrometers equipped with a 5 mm *z*-axis gradient TXI-HCN cryogenic probe. Data were processed with the software Topspin 2.1−3.5 (Bruker Biospin) and evaluated using the software Sparky 3.114.[36] All NMR experiments were performed in NMR buffer in 95% H_2O and 5% D_2O. 1D ^{1}H NMR spectra for conformational screening and NOESY spectra were recorded using the jump-return water suppression scheme.[37] ^{15}N-HSQC spectra were recorded implementing the soft WaterGATE water suppression technique,[38] and ^{15}N-TROSY experiments were recorded with implementations proposed by Brutscher et al.[39,40]

ITC Measurements. ITC measurements were performed with a Microcal VP ITC (Northampton, MA USA) calorimeter. Measurements were performed in NMR buffer at 298 K. The ligand 2′dG was titrated in 20× excess (for exact concentrations, see Figure S9). The ligand was injected in a 1−3 μL volume spaced by 120−240 s for equilibration within a total of 30−50 injections to reach saturation. ITC data were fit with a model for a single set of identical binding sites using the Origin ITC software (OriginLab, Northampton, MA USA).

Coarse-Grained Simulations of Cotranscriptional Folding. Cotranscriptional folding can be modeled computationally on the level of secondary structures, for which a well-established energy model exists.[41] This model is suitable for these simulations, since the refolding behavior of dGsw is characterized by large rearrangements in secondary structure. Tertiary structure would be required to model aptamer binding from first principles. To overcome this issue, we relied on experimentally derived ligand-binding energies[42] and assign a free energy bonus to every binding competent secondary structure.

Cotranscriptional folding was analyzed by means of the previously described barmap approach.[24] In short, this method performs a series

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

of kinetic folding simulations on the energy landscapes corresponding to every elongation step. For each length of the molecule, a coarse-grained representation of the energy landscape using the program barriers is computed,[43] such that each macrostate is represented by a local minimum, as well as a mapping from macrostates of one landscape to macrostates of the next one.[24] On each coarse-grained landscape, folding dynamics can be computed by numerically solving the Markov process using the treekin program.[44] At each elongation step, the population density computed by treekin is transferred from the old landscape to the elongated one using the previously computed mapping. Rates used in the simulations are fixed only up to a prefactor that has to be gauged by comparison with experiment. We therefore simulated kinetic folding of the bistable RNA previously described.[25] Comparison with the experimentally determined refolding rate suggests that 1 s corresponds to approximately 100 000 internal time units. We therefore introduced an elongation step every 4000 time units, corresponding to a transcription rate of 25 nt/sec.

Simulations were performed at 298 K, corresponding to the experimental setup. For the coarse graining, we allowed a maximum of 999 macro states. For clarity of visualization, we performed an additional coarse graining as a post processing, which clustered any macrostates separated by an energy barrier of less than 4.5 kcal/mol. Further details are given in the Supporting Information.

Folding Rates and Markovian Simulations. Interconversion rates between folding intermediates were calculated based on the number of base pairs that need to dissociate in order to initiate refolding. The dependence of folding rates on the number of base pairs to break was adapted from kinetic traces measured by real-time NMR for bistable RNA systems at 298 K.[25] The dependence of the ON state population with the rate of transcription in the absence of ligand was obtained from kinetic Markovian simulations. The model assumes 9 states, representing three conformations (aptamer, multiloop antiterminator, and interior loop antiterminator) at three distinct RNA lengths ($dGsw^{113}$, initiation of ON state folding; $dGsw^{136}$, the ON state folding pathway can no longer be entered; and $dGsw^{144}$, full riboswitch length). The kinetics between the states is described by first-order rate equations with a single rate constant k_{ij} for the interstate transition i to j.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b10429.

NMR characterization of the antiterminator conformation, detailed secondary structure screening of transcriptional intermediates, effect of Mg^{2+} on the antiterminator conformation, simulations of cotranscriptional folding, transcription assays, and ITC data of transcriptional intermediates (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*schwalbe@nmr.uni-frankfurt.de

ORCID

Harald Schwalbe: 0000-0001-5693-7909

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Dr. Christian Richter for excellent assistance with the NMR spectrometers. This work was supported by DFG (Collaborative research center 902), Funds of Chemical Industry (FCI) (C.H.), and the Austrian Science Fund (FWF) project F43 RNA regulation of the transcriptome (M.T.W.). H.S. and M.H. are members of the DFG-funded cluster of excellence: macromolecular complexes. BMRZ is supported by the state of Hesse.

■ REFERENCES

(1) Mahen, E. M.; Watson, P. Y.; Cottrell, J. W.; Fedor, M. J. *PLoS Biol.* **2010**, *8*, e1000307.
(2) Brehm, S. L.; Cech, T. R. *Biochemistry* **1983**, *22*, 2390.
(3) Smith, D. J.; Query, C. C.; Konarska, M. M. *Mol. Cell* **2008**, *30*, 657.
(4) Pan, J.; Woodson, S. A. *J. Mol. Biol.* **1998**, *280*, 597.
(5) Fong, N.; Kim, H.; Zhou, Y.; Ji, X.; Qiu, J.; Saldi, T.; Diener, K.; Jones, K.; Fu, X.-D.; Bentley, D. L. *Genes Dev.* **2014**, *28*, 2663.
(6) Diegelman-Parente, A.; Bevilacqua, P. C. *J. Mol. Biol.* **2002**, *324*, 1.
(7) Lewicki, B. T.; Margus, T.; Remme, J.; Nierhaus, K. H. *J. Mol. Biol.* **1993**, *231*, 581.
(8) Wickiser, J. K.; Winkler, W. C.; Breaker, R. R.; Crothers, D. M. *Mol. Cell* **2005**, *18*, 49.
(9) Perdrizet, G. A.; Artsimovitch, I.; Furman, R.; Sosnick, T. R.; Pan, T. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3323.
(10) Heilman-Miller, S. L.; Woodson, S. A. *RNA* **2003**, *9*, 722.
(11) Pan, T.; Sosnick, T. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 161.
(12) Lai, D.; Proctor, J. R.; Meyer, I. M. *RNA* **2013**, *19*, 1461.
(13) Adhya, S.; Gottesman, M. *Annu. Rev. Biochem.* **1978**, *47*, 967.
(14) Frieda, K. L.; Block, S. M. *Science* **2012**, *338*, 397.
(15) Lutz, B.; Faber, M.; Verma, A.; Klumpp, S.; Schug, A. *Nucleic Acids Res.* **2014**, *42*, 2687.
(16) Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, *44*, 13404.
(17) Lemay, J.-F.; Desnoyers, G.; Blouin, S.; Heppell, B.; Bastet, L.; St-Pierre, P.; Massé, E.; Lafontaine, D. A. *PLoS Genet.* **2011**, *7*, e1001278.
(18) Rieder, R.; Lang, K.; Graber, D.; Micura, R. *ChemBioChem* **2007**, *8*, 896.
(19) Kim, J. N.; Roth, A.; Breaker, R. R. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 16092.
(20) Wacker, A.; Buck, J.; Mathieu, D.; Richter, C.; Wöhnert, J.; Schwalbe, H. *Nucleic Acids Res.* **2011**, *39*, 6802.
(21) Pikovskaya, O.; Polonskaia, A.; Patel, D. J.; Serganov, A. *Nat. Chem. Biol.* **2011**, *7*, 748.
(22) Helmling, C.; Keyhani, S.; Sochor, F.; Fürtig, B.; Hengesbach, M.; Schwalbe, H. *J. Biomol. NMR* **2015**, *63*, 67.
(23) Baird, N. J.; Kulshina, N.; Ferré-D'Amaré, A. R. *RNA Biol.* **2010**, *7*, 328.
(24) Hofacker, I. L.; Flamm, C.; Heine, C.; Wolfinger, M. T.; Scheuermann, G.; Stadler, P. F. *RNA* **2010**, *16*, 1308.
(25) Fürtig, B.; Wenter, P.; Reymond, L.; Richter, C.; Pitsch, S.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 16222.
(26) Buck, J.; Noeske, J.; Wöhnert, J.; Schwalbe, H. *Nucleic Acids Res.* **2010**, *38*, 4143.
(27) Serganov, A.; Yuan, Y.-R.; Pikovskaya, O.; Polonskaia, A.; Malinina, L.; Phan, A. T.; Hobartner, C.; Micura, R.; Breaker, R. R.; Patel, D. J. *Chem. Biol.* **2004**, *11*, 1729.
(28) Reining, A.; Nozinovic, S.; Schlepckow, K.; Buhr, F.; Fürtig, B.; Schwalbe, H. *Nature* **2013**, *499*, 355.
(29) Bremer, H.; Yuan, D. *J. Mol. Biol.* **1968**, *38*, 163.
(30) Manor, H.; Goodman, D.; Stent, G. S. *J. Mol. Biol.* **1969**, *39*, 1.
(31) Adelman, K.; La Porta, A.; Santangelo, T. J.; Lis, J. T.; Roberts, J. W.; Wang, M. D. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 13538.
(32) Kent, T.; Kashkina, E.; Anikin, M.; Temiakov, D. *J. Biol. Chem.* **2009**, *284*, 13497.
(33) Malinen, A. M.; Turtola, M.; Parthiban, M.; Vainonen, L.; Johnson, M. S.; Belogurov, G. A. *Nucleic Acids Res.* **2012**, *40*, 7442.
(34) Lutz, B.; Faber, M.; Verma, A.; Klumpp, S.; Schug, A. *Nucleic Acids Res.* **2014**, *42*, 2687.
(35) Chauvier, A.; Lafontaine, D. A. *Methods Mol. Biol.* **2015**, *1334*, 109.

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656

(36) Goddard, T. D.; Kneller, D. G. *Sparky* 3.114; University of California: San Francisco, CA, 2008.
(37) Sklenář, V.; Bax, A. *J. Magn. Reson. (1969-1992)* **1987**, *74*, 469.
(38) Piotto, M.; Saudek, V.; Sklenář, V. *J. Biomol. NMR* **1992**, *2*, 661.
(39) Favier, A.; Brutscher, B. *J. Biomol. NMR* **2011**, *49*, 9.
(40) Lescop, E.; Kern, T.; Brutscher, B. *J. Magn. Reson.* **2010**, *203*, 190.
(41) Turner, D. H.; Mathews, D. H. *Nucleic Acids Res.* **2010**, *38*, D280.
(42) Badelt, S.; Hammer, S.; Flamm, C.; Hofacker, I. L. *Methods Enzymol.* **2015**, *553*, 193.
(43) Flamm, C.; Hofacker, I. L.; Stadler, P. F.; Wolfinger, M. T. *Z. Phys. Chem.* **2002**, *216*, 155.
(44) Wolfinger, M. T.; Svrcek-Seiler, W. A.; Flamm, C.; Hofacker, I. L.; Stadler, P. F. *J. Phys. A: Math. Gen.* **2004**, *37*, 4731.

DOI: 10.1021/jacs.6b10429
J. Am. Chem. Soc. 2017, 139, 2647−2656