

Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation

Hannah Steinert^{1†}, Florian Sochor^{1†}, Anna Wacker¹, Janina Buck¹, Christina Helmling¹, Fabian Hiller¹, Sara Keyhani¹, Jonas Noeske¹, Steffen Grimm¹, Martin M. Rudolph², Heiko Keller³, Rachel Anne Mooney⁴, Robert Landick⁴, Beatrix Suess², Boris Fürtig^{1,*}, Jens Wöhnert^{3,*}, and Harald Schwalbe^{1,*}

HSt and FS are joint first authors.

¹Center for Biomolecular Magnetic Resonance, Institute of Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe-University Frankfurt am Main, Max-von-Laue-Strasse 7, 60438 Frankfurt am Main/ Germany

²Department of Biology, Technical University Darmstadt, Schnittspahnstraße 10, 64287 Darmstadt/ Germany

³Center for Biomolecular Magnetic Resonance, Institute of Molecular Biosciences, Johann Wolfgang Goethe-University Frankfurt am Main, Max-von-Laue-Strasse 9, 60438 Frankfurt am Main/ Germany

⁴ University of Wisconsin – Madison. Department of Biochemistry, 1550 Linden Drive, Madison WI 53706

*to whom correspondence should be addressed: schwalbe@nmr.uni-frankfurt.de, woehnert@bio.uni-frankfurt.de, fuertig@nmr.uni-frankfurt.de

Abstract

In bacteria, the regulation of gene expression by cis-acting transcriptional riboswitches located in the 5'-untranslated regions of messenger RNA requires the temporal synchronization of RNA synthesis and ligand binding-dependent conformational refolding. Ligand binding to the aptamer domain of the riboswitch induces premature termination of the mRNA synthesis of ligand-associated genes due to the coupled formation of 3'-structural elements acting as terminators. To date, there has been no high resolution structural description of the concerted process of synthesis and ligand-induced restructuring of the regulatory RNA element. Here, we show that for the guanine-sensing *xpt-pbuX* riboswitch from *Bacillus subtilis*, the conformation of the full-length transcripts is static: it exclusively populates the functional off-state but cannot switch to the on-state, regardless of the presence or absence of ligand. We show that only

the combined matching of transcription rates and ligand binding enables transcription intermediates to undergo ligand-dependent conformational refolding.

Introduction

Riboswitches are cis-regulatory RNA elements controlling cellular processes, in particular transcription and translation¹⁻³. They exert their function via a conformational switch between mutually exclusive base-paired structures. This conformational switch is induced by binding to a low-molecular weight ligand⁴⁻⁶.

Riboswitches regulating transcription control gene expression by connecting ligand-dependent and ligand-independent RNA folding processes with the concomitant synthesis of the RNA by the RNA polymerase.

It has been increasingly recognized that the folding of RNA is intimately linked to transcription kinetics, which are directly dependent on transcriptional pausing. The synthesis of the RNA is not performed at a constant speed^{7,8}. Instead, slow conformational changes, from a more competent to a less competent transcription state, lead to a variation of the transcription speed of the elongation complex (EC) between 2-30 nt/s^{9,10}. The less competent transcription state can be adopted at any point during transcription and can slow down the EC by inhibiting nucleotide addition¹¹⁻¹⁴. Thus, transcription rates vary during transcription. This elemental pausing occurs frequently, but typically lasts only for 1-6 seconds on average, making the overall transcription movement the sum of more and less competent transcription states averaged over the number of RNA polymerases (RNAP) molecules observed. In contrast to short elemental pausing, long pausing events occur at defined template sequences (pause sites) or upon interaction with protein factors¹⁵. At pause-sites, the RNAP speed decreases by a factor of 100 compared to its maximal transcription speed (max. 75 nt/s, measured on a specialized DNA template)¹⁶. Pause sites can slow RNAP through backtracking of the enzyme relative to the nucleic acids or by the formation of a pause RNA hairpin after the EC has entered an initial elemental pause state which makes elemental pausing an obligatory intermediate from which longer pausing can occur¹⁷⁻²².

Pausing can also facilitate the folding of certain RNA structures and allow a stabilizing interaction between the nascent RNA and the transcription machinery²³. Pausing can be important for effective riboswitch-mediated regulation where co-transcriptional folding is influenced by the transcription speed^{24,25}. The modulation of the available conformational space by altering the transcriptional speed was demonstrated for the *btuB* riboswitch where the use of a mutant RNA polymerase with deficient pause behavior caused the riboswitch to adopt a non-native structure and to be deficient in ligand binding²⁵.

Here, we investigate the *xpt-pbuX* riboswitch from *Bacillus* and identify and characterize conformational sub-states and the kinetics of their inter-state conversion. Our work provides insight into how RNA polymerase residence time and the intramolecular refolding kinetics must be fine-tuned to allow riboswitch-based regulation of transcription. In this contribution, we apply high resolution structural and kinetic techniques to decipher the full co-transcriptional folding pathway, enabling monitoring the behaviour of the decisive structural units during synthesis for the first time.

Results

The guanine-sensing riboswitch from *Bacillus subtilis* (GSW) negatively regulates the transcription of the *xpt*-gene by the premature termination of RNA synthesis³. In this riboswitch, the regulatory output signals of maintaining or repressing transcription require two conformational states with different base-pairing interactions of four complementary sequences P, A, T and H (Figure 1a). In the off-state, the 5'-aptamer-strand (P) pairs with the aptamer-stabilising strand (A) and the switching strand (T) forms the terminator helix together with the terminator strand (H). By contrast, in the putative on-state (Figure 1b), A and T pair and form the antiterminator conformation, and the P and H strands are unpaired.

The free energy landscape of RNA folding features high barriers for the transitions between conformations with alternative base-pairing. These transitions are frequently found to be slower than the timescale of the relevant biological process²⁶. In addition, RNA chains can be trapped in a stable conformation²⁷. The terminator conformation of GSW represents such a “super” stabilised structure. Regardless of the presence or absence of the cognate ligand, the RNA adopts a single conformation in which the terminator helix (TH) is formed as inferred from NMR signals (for assignment, see Figure 1 and Figure Supplement 1) arising from the stable GU base-pairs (U129-G147 and U130-G146) (Figure 1c, d). Addition of the ligand to the GSW induces only changes within the binding core of the aptamer domain²⁸ but the aptamer closing helix (PA) is stably present both in the apo and holo form of the GSW and formation of the antiterminator helix (AT) cannot be detected (Figure 1c,d; Figure 1-Figure Supplement 2).

Since the terminator conformation in the full-length construct is the only detectable state regardless of the presence or absence of ligand, it must be stabilised by at least $\Delta\Delta G > 8 \text{ kJ mol}^{-1}$ compared to the antiterminator conformation at a NMR signal-to-noise greater than 10:1.

86 These observations give rise to two fundamentally important questions: If the full-length riboswitch is trapped
 87 in a single stable structure and not able to refold, which transcription intermediate is capable of undergoing the
 88 regulatory active conformational change^{24,29,30}? If the thermodynamic stable conformation resembles an off-
 89 state, how can the organism ever adopt the on-state to allow gene expression?
 90 We therefore tested for accumulation of distinct transcription intermediates, occurring as shorter RNA
 91 products are produced during the synthesis of full-length riboswitches¹⁹. In multiple round transcription
 92 reactions, we detected several distinct shorter RNA products (Figure 1e). Three of the detected fragments
 93 correspond to the aptamer domain (nucleotides 10-89), to the prematurely terminated form (nucleotides 10-
 94 164), and to the full-length of the mRNA-5'-UTR. Mapping of these RNA fragments to the riboswitch structure
 95 reveals that they correspond to transcription intermediates that include the strands A, T, and H at their 3'-ends,
 96 respectively. Furthermore, two additional fragments of 100 and 140 nucleotides are detected. These latter
 97 RNAs exhibit 3'-terminal stretches of U-residues and therefore represent canonical transcriptional pause
 98 sites^{20,31}.
 99 Based on the detection of these fragments, we truncated GSW after strands A or T to generate transcription
 100 intermediates containing either only the aptamer domain (GSW^{PA}) or the aptamer domain and the subsequent
 101 sequence region of the antiterminator (GSW^{PAT}). Both GSW fragments bind the ligand with similar affinities as
 102 the full-length riboswitch suggesting formation of the unperturbed aptamer domain (Figure 2 and Figure
 103 Supplement 1).
 104 For GSW^{PAT}, we observed two equally intense sets of NMR signals from two long-lived conformations GSW^[PAT]
 105 and GSW^{P[AT]} (Figure 1b, c) that differ with regard to the formation of the two mutually exclusive helices PA or
 106 AT. In GSW^[PAT], helix PA is formed, and in GSW^{P[AT]} the antiterminator helix AT is formed. Upon addition of
 107 ligand, we detect ligand binding and subsequent shifting of the conformational equilibrium towards GSW^[PAT]
 108 (Figure 1c, d). Thus, in contrast to the full-length riboswitch, the antiterminator conformation representing the
 109 functional on-state is a thermodynamically stable state (GSW^{P[AT]}) of this transcription intermediate (GSW^{PAT}) in
 110 the absence of ligand.
 111 In order to evaluate ligand binding and the refolding kinetics of transcriptional intermediates, we mimicked
 112 these processes by sequentially adding short oligonucleotides of sequences T (T^{trans}) and H (H^{trans}) to GSW^{PA}.
 113 Using isotope-filter NMR experiments, selectively ¹⁵N-labelled GSW^{PA} (containing the P and A sequences), and
 114 ¹⁴N-labelled T^{trans} and H^{trans} we investigated the sequential refolding of the mRNA transcription intermediates

with nucleotide resolution (Figure 2). The imino-signal stemming from the aptamer domain in the ligand-free (Figure 2a) and –bound state (Figure 2b) can be readily assigned and constitute the characteristic NMR signals to detect the relevant GSW conformations and their ligand-induced refolding. Without ligand, addition of equimolar amounts of T^{trans} to GSW^{PA} led to formation of the antiterminator conformation mimic containing helix AT and the strand P is unpaired (Figure 2c). This $GSW^{PA}-T^{trans}$ complex was disrupted by the addition of ligand (Figure 2d), restoring formation of helix PA. In contrast, the presence of ligand stabilised GSW^{PA} so that subsequent addition of T^{trans} did not disrupt helix PA in the aptamer domain within the GSW^{PA} -ligand-complex and formation of the antiterminator helix was suppressed. From these findings, we can conclude that ligand binding to GSW^{PA} represents the structural decision point.

Regardless of the presence or the absence of ligand, addition of H^{trans} to the antiterminator conformation in the $GSW^{PA}-T^{trans}$ complex induced a conformational switch characterized by the formation of the terminator helix TH and the helix PA (Figure 2 e and f).

Having characterized these three functional relevant states, we conducted real-time NMR experiments³² to determine the kinetics of RNA refolding to the antiterminator and to the terminator conformation and of ligand binding to the distinct conformations, i.e. the critical events for GSW function (Figure 3). Utilizing the above described labeling strategy, we were able to resolve the NMR signals from the ligand, the acceptor domain and the associating or dissociating single RNA strand. In total, 59 signals could be detected, and the reported kinetics have been measured in triplicate experiments and averaged over structural elements in the riboswitch. Ligand binding needs to lock strand A in the aptamer's PA helix and suppress formation of the antiterminator conformation. Indeed, ligand binding to GSW^{PA} either in the absence or in presence of the terminator helix (Fig. 3 f and g) was fast, with a k_1 of $(2.4 \pm 0.3) 10^{-1} s^{-1}$, and induced a slower tertiary structure rearrangement ($k_2 = (1.8 \pm 0.2) 10^{-2} s^{-1}$) within the aptamer, consistent with previous experiments^{16,17}. Subsequent formation of the terminator helix from single strands was also rapid ($k_1 = (1.4 \pm 0.8) 10^{-1} s^{-1}$; Figure 3h).

In the absence of ligand, GSW^{PA} and T^{trans} anneal rapidly to form the antiterminator complex ($k = (9 \pm 5) 10^{-3} s^{-1}$; Figure 3 b). Subsequent formation of the terminator helix TH was observed if H^{trans} was added to the $GSW^{PA}-T^{trans}$ complex (antiterminator conformation) (Figure 3c). Under these conditions, the dissociation of the antiterminator helix AT was rate-limiting and slow ($k_1 = (0.5 \pm 0.2) 10^{-3} s^{-1}$; $k_2 = (1.1 \pm 0.4) 10^{-4} s^{-1}$ (T^{trans} intrinsic unfolding, Figure 3 – Figure Supplement 1)). Importantly, the dissociation of helix P4 was accelerated by the presence of ligand, but in a concentration-independent manner above equimolar ratios (Figure 3 – Figure

Supplement 2; $k = (0.89 \pm 0.09) 10^{-3} \text{ s}^{-1}$; Figure 2d) and was stimulated to a higher degree if ligand binding and annealing of H^{trans} occurred simultaneously (Figure 3e). Under the latter conditions, the formation of the terminator conformation was described by two rate constants, $k_1 = (2 \pm 1) 10^{-2} \text{ s}^{-1}$ reporting on the ligand-induced refolding of the aptamer domain and $k_2 = (1.2 \pm 0.4) 10^{-3} \text{ s}^{-1}$ characterising the rate-limiting dissociation of helix AT.

From these kinetic data (Supplementary Tables 1-7, Figure 3i), we conclude that the AT conformation represents a meta-stable state that slowly refolds into the thermodynamically stable terminator state (which represents the state of lowest free energy). The dissociation of PA is slow compared to the polymerase transcription speed. Therefore, during mRNA synthesis, transcriptional resting states are required to slow transcription to ensure sufficient time for the refolding of conformations [PA]T to P[AT] to enable riboswitch function. We are aware that in this applied experimental strategy some of the kinetics are measured in trans, although the riboswitch is a cis-acting element. However, at the applied concentrations of 0.1 mM, the most relevant conformational transition of ligand-dependent [PA] is slow and if the rate may possibly be overestimated by this method, which would not impact the fundamental conclusion as discussed below.

Transcription rates between 12 to 90 nucleotides per second have been reported in prokaryotes^{11,16}. However, rates of synthesis vary during transcription. 5'-untranslated regions (5'-UTRs) for example often contain several U-rich sequences that have been shown to constitute transcriptional pause sites^{19,33}. Other sequences can also induce pausing of RNA polymerase, in particular in *B. subtilis*³¹. In any case, pausing has been reported to assist co-transcriptional folding e.g. of the FMN- and coenzyme B₁₂-sensing riboswitches^{24,25}. Within the sequence of GSW, a U₆-stretch (U107-U112) is suspiciously located between the aptamer domain and the terminator helix (pause site 1, PS1) and between the A and T stretches of the antiterminator (U135-U141, pause site 2, PS2) (for analysis of the distance between aptamer and terminator sequences in purine riboswitches see Figure 4 – Figure Supplement 1). We analysed the synthesis rate of the RNA polymerases from *E. coli* (EcRNAP) and *B. subtilis* (BsRNAP) in a time-resolved single-round transcription assay (Figure 4). Seven transcribed RNAs could be identified, suggesting that the formation of the shorter products after transcription of the aptamer domain and PS1 accumulate at a later time point during the overnight multi-round transcription. The seven RNA fragments were identified as FL (214 nt), GSW^{PATH} (164 nt), PS2 (141 nt), PS1 (between 107 and 112 nt, Figure 4 – Supplement 1), and three additional paused RNA fragments. Two of these additional pauses were seen with transcription with EcRNAP and were mapped as RNA95 and RNA77 (transcripts 95 nt and 77 nt in length,

respectively). Transcription with BsRNAP through the same sequence showed the same pausing at RNA77,
 pausing at a different position, RNA90 (transcript 90 nt in length), but no pausing at RNA95. The FL and GSW^{PATH}
 fragments accumulated over time and reached their maximum at 300 s and more than 600 s, respectively, for
 both the Ec and Bs RNAPs. The GSW^{PATH} concentrations did not reach their maximal values within the observed
 600 s. For the determination of transcription rates, the maximum values were included in the fitting procedure.
 Upon ligand addition, the intensity of GSW^{PATH} increases whereas the intensity of FL decreased. The intensities
 of the pausing transcription intermediates appear unaffected by ligand addition. The data indicate that the
 riboswitch maintains its ligand-response function during the *in vitro* transcription assays, as the termination
 efficiency is increased by 24.8% for EcRNAP and by 30.3% for BsRNAP. The intensities of the PS2 and PS1 RNA
 also show an exponential growth, reaching maximum value after 120 s, indicating that under low NTP-
 concentrations, the poly-U sequences act more as a polymerase blocking point than a pause-site. This could be
 observed for both RNAPs. The intensity of PS2 is near the level of noise for BsRNAP, making quantitation of the
 amount of transcript difficult for this fragment. However, the pausing values could be determined, and these
 are of higher importance for the simulations. Furthermore, the intensity of PS1 is stronger for BsRNAP
 transcription, indicating that PS1 has a higher influence on the elongation complex for the BsRNAP. In addition,
 the determined pausing-parameters (Supplementary Information Table 8) show that both RNAPs pause longer
 at PS1 than at PS2. This effect is increased for the EcRNAP when ligand is added whereas pausing of the BsRNAP
 shows no ligand dependency. Relative to the pausing observed at PS1 and PS2, pausing at RNA77, RNA90, and
 RNA95 was of shorter duration for both RNAPs. For EcRNAP-mediated transcription in the absence of ligand,
 pausing at RNA77 could not be clearly determined as the relative intensities show a linear than an exponential
 decay. However, as four pausing events are observed for each RNAP, at least in the presence of ligand, even
 these brief pauses can have a high impact on the overall transcription kinetics. The apparent transcription rates
 were determined to range from $0.4 \pm 0.1 \text{ nt s}^{-1}$ and $2.0 \pm 0.5 \text{ nt s}^{-1}$ (FL and GSW^{PATH}) to $1.3 \pm 0.6 \text{ nt s}^{-1}$ and
 $47.0 \pm 53.3 \text{ nt s}^{-1}$ for the synthesis of the shorter transcription intermediates at 37°C (Figure 4). Upon ligand
 addition, the apparent transcription rate decreased by at least a factor of two for PS2 and PS1 but was nearly
 identical for RNA95/RNA90 and RNA77. However, the multiple pausing events during transcription drastically
 decrease the apparent transcription rates of the longer RNAs FL and GSW^{PATH}.

In order to examine the mutual influence of transcription speed and refolding kinetics on the population of
 intermediate riboswitch conformations we performed simulations based on a kinetic Markov-model³⁴ using the

experimentally determined refolding and transcription rates of 20 nt/s. This value is based on the apparent rates of transcription determined in the transcription assays for *B. subtilis* polymerase but taking into account the cellular concentrations of nucleotides, that are approx. 10 times higher than those used in the in vitro transcription reactions³⁵ (Figure 5). In the absence of ligand, there is no major change as the intermediate conformations occur in a linear succession. In both cases, with and without transcriptional pausing, RNA synthesis is fast enough to allow for formation of the functional on-state apo-form conformation P[AT] (Figure 5 b-e). Refolding to the thermodynamic most stable functional off-state conformation [PA][TH] in its apo-form only occurs after the polymerase has passed the regulatory point of decision. In presence of the ligand the conformational space is described as a branched succession of intermediates (Figure 5a). Without pausing of the transcription machinery, riboswitch synthesis is fast compared to its own refolding reactions from apo- to holo-conformations. Therefore, the GSW does not reach a substantial population of conformations with the formed terminator [TH] and the polymerase escapes termination (Figure 5f, g). In contrast, pausing of transcription at pause site 1 opens a time window that allows for refolding of the riboswitch from the apo- to the holo-form. Consequently, the [PA] conformation is formed which is stable over the time window of regulation. In continuation of transcription strand T is synthesized but because it can no longer rapidly associate with strand A to form an anti-terminator conformation, it will form the terminator hairpin conformation [TH] with strand H, as the elongation of the riboswitch by the polymerase occurs with rates not allowing dissociation of the [PA] helix. Therefore, a reduced rate of transcription after formation of the aptamer nucleotides of the riboswitch due to pausing at PS1 in conjunction with a comparative slow conformational transition from [PA]T to P[AT] results in higher population of off-states during synthesis before the polymerase leaves the point of decision (Figure 5h,i).

In order to validate the derived model, we tested whether the pause site PS1 could further increase the time for ligand binding to the aptamer domain and shift the riboswitch towards off-state function in an *in vivo* reporter assay. Successive disruptive mutations (M1-M3) of PS1 correlated with the loss of regulation *in vivo* (Figure 6a), and the addition of further U residues to PS1 (M4) increased riboswitch efficiency. Monitoring the dose-dependence of the wild-type riboswitch compared to the M3 and M4 mutants indicated that the loss of pause site integrity reduced the switching efficiency, whereas extended pausing of the RNA polymerase led to increased repression of reporter gene expression (Figure 6b). Hence, pausing of the polymerase at PS1

230 temporarily decouples the continuous synthesis of the downstream antiterminator sequence, assists in folding
231 of the aptamer domain and allows for sufficient time for the ligand to bind at the structural decision point.
232

Discussion

Single molecule force extension experiments showed that the structurally similar aptamer domain of the adenine-sensing riboswitch folds co-transcriptional in a hierarchical manner and that the closure of the PA is the last step in accommodation of the ligand in the holo-form³⁶. These findings and our proposed model are consistent as both conclude that the formation of the base-paired PA-stem is the committed-step in the off-pathway. Our experiments complement the previous data, however, by investigating the full-length riboswitch, which is important to map out the coexistence of ligand-dependent and ligand-independent functional off-states³⁷. For this, we show that one of the two proposed functionally important conformational states of transcriptional riboswitches, the antiterminator state, represents a metastable conformation. The structural analysis using static and dynamic experiments performed close to the thermodynamically equilibrium indicate that an additional kinetic component is needed for adoption of the antiterminator conformation during transcription to allow full biological function of the riboswitch. Pausing during RNA synthesis and the inherent influence on the conformation of the nascent RNA chain is likely required for formation of the antiterminator. Without this additional kinetic component, the terminator would always form and transcription of the adjacent genes would never occur. This is reminiscent of the notion that transcriptional riboswitches are kinetically driven, as here the folding of the RNA has to compete with the speed of the transcribing polymerase. For the FMN switch from *Bacillus subtilis*, it was shown that FMN binding and subsequent folding into the terminator conformation requires pausing of the bacterial polymerase²⁴. Unexpectedly, our simulations show that without pausing, the antiterminator conformation is always adopted, independently of ligand concentration. Without pausing, effective regulation could not be achieved by the riboswitch because gene expression could not be turned off. Under these conditions, the thermodynamically favored terminator conformation is not adopted. The presence of the multiple pausing events between the formation of the binding competent aptamer and the transcription of the switching T-strand allow the terminator to form and fulfill its regulatory purpose. In addition we showed that the polymerases of *B. subtilis* and *E. coli* utilize different pausing sequences (e.g. RNA95 and RNA90) and therefore differently affect conformational change in the synthesized RNA. Thus, the effective biotechnological application of transcription-regulating riboswitches in heterologous organisms may require a match between polymerase pausing sequences and the conformational space of the riboswitch. Our work significantly extends our understanding of the previously introduced concept of kinetic control of

261 riboswitches, and reveals the importance of matching discontinuous transcription rates to folding rates of
262 transient conformations in order to facilitate RNA-based regulation of gene expression.
263
264

Methods

RNA constructs

RNA constructs GSW^{PATH}, GSW^{PAT}, GSW^{PA}, dGSW and FSW were synthesised by *in vitro* transcription using T7 polymerase and purified as described³⁸. Antisense oligonucleotide S2^{trans} was purchased from Thermo Fisher (Epsom, Great Britain), 7-deazaG147 modified S3^{trans} was obtained from IBA (Göttingen, Germany).

NMR-spectroscopic experiments

Unless stated otherwise, all NMR experiments were performed in NMR buffer (2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate, pH 6.2) containing 10% or 5% D₂O (static or kinetic experiments) on Bruker spectrometers (Rheinstetten, Germany) and analysed with Topspin (Bruker, Rheinstetten, Germany). Hypoxanthine was used as the better soluble ligand. Real-time NMR experiments were performed at 283 K utilising a rapid sample mixing device³⁹. Selectively ¹⁵N-U and ¹⁵N-G,U labelled GSW^{PA} and non-labelled antisense oligonucleotides T^{trans} and H^{trans} were used in combination with a pulse sequence containing an x-filter element to improve the spectral resolution. The RNA concentrations were chosen to be 150 μM after the mixing event; RNAs GSW^{PA}, T^{trans} and H^{trans} were in a 1:1(:1) ratio, 4 equivalents of hypoxanthine were added. Ligand binding and conformational rearrangements were monitored on the resolved imino proton resonances reporting on the corresponding interactions. Association of GSW^{PA} and T^{trans} was performed in a 1:2 ratio in 50 mM potassium chloride, 25 mM potassium phosphate, pH 6.2.

ITC

ITC measurements were performed with a Microcal VP ITC (Northampton, MA USA) at 283 K. A 217 μM solution of hypoxanthine as ligand was titrated to a 15 μM solution of RNA using 20-42 injections. Buffer conditions were 2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate, pH 6.2. Data were analysed with the Origin ITC software (OriginLab, Northampton, MA USA) assuming a single binding site.

Simulations

Kinetic Markovian simulations were performed in order to evaluate the switching behaviour of the guanine sensing riboswitch under different transcription synthesis velocities. The model assumes N discrete states each

representing a distinct length and associated conformation of the RNA chain. The kinetics between the states is described by first order rate equations with a single rate constant k_{ij} for the inter-state transition i to j . For the synthesis steps in the evaluated models, a transcription rate of 20 nt s^{-1} was assumed. The rates connecting conformational states by refolding reactions were employed as determined here from the real-time NMR measurements. Two sets of simulations were performed, in the first set the transcription rate was adjusted to the experimental value for transcription without any pausing events, while the second set simulated a pausing event between states $PA/^{lig}PA$ and $PAT/^{lig}PAT$. Both sets of simulations were performed for absence and presence of cognate ligand. For the pausing events the transcription rate was set to the reduced transcription rate determined in the transcription assays. Thereby, the reduction of the transcription rate over the pause site corresponds to a factor of ≈ 70 in the simulations.

Expression and purification of the *B. subtilis* RNA-Polymerase

Cell cultures

The *B. subtilis* MH5636 strain was grown in Bacillus growth medium (5 g tryptone, 10 g yeast extract, 5 g NaCl, and 10 g dextrose per liter) at 37°C with shaking at 200 rpm to a OD_{600} of 1.0-1.2. The cells were harvested by centrifugation with 5000 rpm for 15 min at 4°C . The pellet was frozen at -80°C .

French Press

Cells (60 grams) were thawed on ice and resuspended in 120 ml Lysis buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 230 mM NaCl, 5 % glycerol, 1 mM β -mercaptoethanol, 1 mM DTT, 1 mM phenylmethylsulfonyl fluoride, 1 ml of a protease inhibitor cocktail (31.2 mg/L benzamide, 0.5 mg/L chymostatin, 0.5 mg/L leupeptin, 0.1 mg/L aprotonin, and 1 mg/L antipain). After cells were fully resuspended, 0.1 mg/mL lysozyme was added to the cells on ice. The cell suspension was lysed in the cold using a French Press at 1000 PSI. The lysate was then centrifuged for 15 min with 12000 rpm at 4°C . The resulting supernatant was transferred to new centrifuge tubes and again centrifuged for 15 min with 12000 rpm at 4°C .

Polyethyleneimine (PEI) and Ammonium Sulfate Precipitation

In the cold, the supernatant was transferred to a 250 mL beaker and 8% PEI was added slowly (0.5 ml added every 5 min) with constant stirring to a final concentration of 0.6 %. After the last PEI addition, the solution was stirred for an additional 30 min and then centrifuged for 15 min at 12000 rpm at 4 °C.

The supernatant was removed and the PEI pellet was washed with 200 mL TGEZ (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 5 μ M ZnCl₂, 5 % glycerol, 5 mM β -mercapthoethanol) + 0.4 M NH₄Cl. The pellet was properly dispensed in the buffer. The suspension was centrifuged for 15 min with 12000 rpm at 4 °C.

The pellet was resuspended in 200 mL TGEZ + 1 M NH₄Cl for protein elution. The pellet was completely resuspended in the buffer and then the solution was centrifuged for 15 min with 12000 rpm at 4 °C.

The supernatant was transferred to a prechilled beaker. Ammonium sulfate was gradually added to a final concentration of 37 % with constant stirring. After the last addition, the solution was stirred overnight at 4 °C.

In the morning, the solution was centrifuged for 15 min at 12000 rpm at 4 °C.

Ni²⁺ chromatography using His-Trap column

The pellet was resuspended in Loading buffer (25 mM Tris-HCl pH 8.0, 2 mM EDTA, 230 mM NaCl, 5 % glycerol, 10 mM imidazole, and 5 mM β -mercaptoethanol) and centrifuged at 12000 rpm for 15 min at 4 °C. The resulting supernatant was then filtered and loaded onto a His-trap column equilibrated in the same buffer at a flow rate of 1 ml/min. After washing with ten column volumes, the protein was eluted using a gradient of 0.01 M to 0.5 M imidazole in the same buffer over 30 min, then from 0.5 M to 0.75 M over 10 min, and then 1 M imidazole. 2 ml fractions were collected.

The fractions were analyzed by 4-12 % gradient PAGE and Bradford assay. The RNAP was eluted in fractions 7-23, which corresponded to buffer with 200-850 mM imidazole. The combined fractions were dialyzed overnight in TGEZ + 100 mM NaCl + 5 mM DTT.

Heparin Hi-Trap column

The dialyzed solution was filtered and then loaded onto the heparin column equilibrated in the same buffer. After washing with the same buffer for 10 column volumes, the protein was then eluted in the same buffer using a gradient to 1 M NaCl.

348 Protein-containing fractions were pooled and dialyzed into storage buffer (10 mM Tris-HCl pH 8.0, 25 %
349 glycerol, 0.1 mM EDTA, 100 mM NaCl, 20 μ M ZnCl₂, 1 mM MgCl₂, and 10 mM DTT) overnight. The total yield of
350 the core enzyme ($\epsilon_{260} = 210025 \text{ M}^{-1} \cdot \text{cm}^{-1}$) was 5 mL of 0.83 mg/mL, 2.5 μ M solution.
351

Purification of *B. subtilis* SigA

Cell culture

E. coli Rosetta 2 DE3 cells were transformed with plasmid pT7_Bs_sigma_A and selected on LB agar plates with antibiotics (20 µg/mL Kanamycin and 20 µg/mL Chloramphenicol) at 37 °C. From this plate, a single transformant was grown at 37 °C in 5 ml LB medium with antibiotics. This culture was then used to inoculate 2 liters of LB with antibiotics. The cultures were grown with shaking at 37 °C to an OD₆₀₀ of 0.5. The cultures were then shifted to 16 °C and the overexpression of SigA (43 kDa) was induced by addition of 1 mM IPTG. Cultures were maintained at 16 °C with shaking overnight. The cells were then harvested by centrifugation at 5000 rpm for 15 min at 4 °C. The 8.5 g pellet was frozen at -80 °C.

Sonication

Cells were thawed on ice and resuspended in 35 ml SigA Lysis buffer (50 mM Tris-HCl pH 8.0, 2 mM EDTA, 233 mM NaCl, 5 % glycerol, 5 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride, and 350 µl of a protease inhibitor cocktail (31.2 mg/L benzamide, 0.5 mg/L chymostatin, 0.5 mg/L leupeptin, 0.1 mg/L aprotinin, and 1 mg/L antipain). The resuspended cells were transferred to a prechilled beaker and lysis buffer was added (0.130 mg/mL) and the cells were incubated on ice for 15 min. Sodium deoxycholate was then added to 0.05 % final and incubated on ice for an additional 15 min. Cells were then subjected to sonication at output 6 with a 20 % duty cycle for 30 min. The sonicated lysate was then centrifuged for 15 min with 12000 rpm at 4 °C.

His-Trap purification

The SigA-containing cell lysate was filtered and then loaded onto a His-Trap column equilibrated in the same buffer plus 10 mM imidazole. After washing with ten column volumes, the protein was eluted using a gradient to the same buffer containing 1 M imidazole.

The protein-containing fractions were pooled and analyzed by 20 % SDS-PAGE. As the fractions contained pure SigA, they were dialyzed into SigA Storage Buffer (10 mM Tris-HCl pH 8.0, 30 % glycerol, 0.1 mM EDTA, 100 mM NaCl, 20 µM ZnCl₂, 1 mM MgCl₂, and 0.1 mM DTT) overnight. The resulting protein solution contained 5 mL of 25 mg/mL or 590 µM ($\epsilon_{260} = 24410 \text{ M}^{-1} \cdot \text{cm}^{-1}$).

Time-resolved transcription assay

The DNA template used for time resolved transcription analysis was composed of either a modified native promoter (*B. subtilis* wt promoter with A-35T, T-35G, G-33A mutations) or the strong lambda pR promoter, followed by the GSW-coding sequence from the *B. subtilis* genome (GSW, Supplementary Information Supplementary Sequence 1). For halted complex formation (HC), 25 nM DNA template and 40 nM RNAP holoenzyme (expression and purification shown in SI) were incubated in transcription buffer (20 mM Tris-OAc pH 7.7, 40 mM KOAc, 5 mM Mg(OAc)₂, and 5 mM β-mercaptoethanol) with 10 μM ATP and UTP, 7.5 μM GTP, and 150 μM ApA (for the *B. subtilis* GSW promoter) or ApU dinucleotide (for the lambda pR promoter). For transcriptions using the E.cRNAP, 10 μCi GTP* were used whereas transcriptions using the BsRNAP used 20 μCi GTP*. The HC of EcRNAP seem to be more stable than the ones with BsRNAP. Additional GTP* and therefore increased incorporation of ³²P-labeled nucleotides into the nascent RNA, made more of the fewer transcribed RNAs visible on the PAGEs. Ligand (Hypoxanthine, Sigma Aldrich) was added to a final concentration of 200 μM. Purified rNTPs were obtained from Promega and dinucleotides were obtained from TriLink BioTechnologies. Synchronized transcription was re-started by adjustment of the ATP, CTP, and UTP concentrations to 150 μM and the GTP concentration to 10 μM, and the addition of heparin to a final concentration of 100 μg/mL. Transcription was stopped at different time points by the addition of an equal volume of 2X STOP buffer (100 mM Tris-HCl pH 8.3, 100 mM Boric acid, 40 mM EDTA, 7 M urea). For determination of the transcription rates at high NTP-concentration, the NTP-concentration was raised to 1 mM. After heating to 95°C for 2 min, the reactions were analyzed by 8% PAGE. The gels of the time-resolved transcription assays show several expected RNA signals (Figure 4): pausing 1 fragment (PS1, 110 nt), pausing 2 fragment (PS2, 141 nt), terminated transcript (Gsw^{[PA][TH]}, 164 nt), and full-length transcript (FL, 214 nt) as well as two additional pausing events at 77 nt (RNA77) and 95 nt (RNA95) for *E. coli* RNAP mediated transcriptions and 77 nt (RNA77) and 90 nt (RNA90) for *B. subtilis* RNAP mediated transcription. The positions of the RNA species were as estimated by comparison on high-resolution denaturing polyacrylamide gels of transcripts to RNA sequence ladders generated using 3' deoxyNTPs. To do this, halted ECs were first generated as described above. To generate sequence ladders, transcription was restarted by addition of ATP, CTP, GTP, and UTP at 250 μM each, and one 3'-deoxyNTP in each of 4 separate reactions at 240 μM dGTP, 300 μM dATP, 350 μM dCTP, or 400 μM dUTP, followed by addition of heparin at 100 μg/mL. After incubation for 10 min at 37°C, samples were mixed with an equal

volume of 2X STOP buffer before being electrophoresed next to transcription products obtained by regular in vitro transcription as described above. To quantitate the relative amounts of the different RNA species, the signal intensities were plotted against time and analyzed using Lab Image 1D software (INTAS, Göttingen, Germany) and Image Quant 5.2 (GE, Freiburg, Germany). The mean apparent transcription rates k_{app} for FL, Gsw^{[PA][TH]}, PS2 and PS1 were derived from the exponential plots as shown in by Ederth et al.⁴⁰. The signal intensities of RNA95, RNA90 and RNA77 were plotted against time and revealed that these fragments were transcribed with apparent first-order kinetics. The k_{app} was derived similar as previously described, by determining the average appearance rate for the first values. Knowing the length of the RNA fragment, the rates in nt s⁻¹ were then calculated from the k_{app} (Supplementary table 7). Determination of $t_{1/2}$, τ , and k_p were performed according to Landick et al.⁴¹.

Regulation of β -galactosidase reporter gene expression by the wildtype GSW (wt) and pause site mutants (M1 – M4)

For reporter gene assays, the 5'-UTR of the *B. subtilis xpt-pbuX* operon (encompassing nt -120 to +196 relative to the transcription start site) was fused directly upstream of the *lacZ* reporter gene of plasmid pDG268⁴². Pause site mutants M1 – M4 were introduced by PCR mutagenesis. The resulting plasmids were integrated into the *amyE* locus of *B. subtilis* 168. The respective strains were grown in modified CSK minimal medium⁴³ with increasing concentration of guanine. After 5 h growth, β -galactosidase activity was determined as follows: 100 μ L culture aliquots were incubated at 301 K in buffer (60 mM Na₂HPO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgCl₂ and 50 mM 2-mercaptoethanol) with lysozyme and Triton-X treatment. 200 μ L of *o*-nitrophenyl- β -D-galactoside (4 mg mL⁻¹) was added, and the sample was incubated at 301 K. The reaction was stopped by addition of 0.5 mL of 1 M Na₂CO₃ upon appearance of a yellowish colouration. The time for colour change was noted and absorbance at 420 nm and 550 nm was measured. β -galactosidase activity was calculated in Miller units, i.e. $1000 \times [A_{420} - 1.75 \times A_{550}] / [t \times V \times A_{600}]$. Three independent cultures were measured in parallel. The measurements were repeated twice.

Acknowledgements

This work was funded by the German funding agency (DFG) in Collaborative Research Center 902: Molecular principles of RNA-based regulation. H.Sc., and J.W. are members of the DFG-funded cluster of excellence:

macromolecular complexes. Work at BMRZ is supported by the state of Hesse. We thank Roland Hartmann for initial help with the preparation of *B. subtilis* polymerase.

References

1. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747–56 (2002).
2. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043 (2002).
3. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches Control Fundamental Biochemical Pathways in *Bacillus subtilis* and Other Bacteria. *Cell* **113**, 577–586 (2003).
4. Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–5 (2004).
5. Noeske, J. *et al.* An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1372–7 (2005).
6. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729–41 (2004).
7. Maizels, N. M. The nucleotide sequence of the lactose messenger ribonucleic acid transcribed from the UV5 promoter mutant of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3585–9 (1973).
8. Darlix, J. L. & Horaist, M. Existence and possible roles of transcriptional barriers in T7 DNA early region as shown by electron microscopy. *Nature* **256**, 288–92 (1975).
9. Davenport, R. J., Wuite, G. J., Landick, R. & Bustamante, C. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science* **287**, 2497–500 (2000).
10. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell* **115**, 437–47 (2003).
11. Adelman, K. *et al.* Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 13538–43 (2002).
12. Kireeva, M. L. & Kashlev, M. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8900–5 (2009).
13. Strobel, E. J. & Roberts, J. W. Two transcription pause elements underlie a σ^{70} -dependent pause cycle. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4374–80 (2015).
14. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S. A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431–41 (2013).
15. Roberts, J. W. *et al.* Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 319–25 (1998).

16. Vogel, U. & Jensen, K. F. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.* **176**, 2807–13 (1994).
17. Reeder, T. C. & Hawley, D. K. Promoter proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell* **87**, 767–77 (1996).
18. Kireeva, M. L., Komissarova, N. & Kashlev, M. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J. Mol. Biol.* **299**, 325–35 (2000).
19. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7090–5 (2000).
20. Landick, R. The regulatory roles and mechanism of transcriptional pausing. *Biochem. Soc. Trans.* **34**, 1062–1066 (2006).
21. Nudler, E. RNA polymerase backtracking in gene regulation and genome instability. *Cell* **149**, 1438–45 (2012).
22. Hein, P. P. *et al.* RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
23. Pan, T., Artsimovitch, I., Fang, X. W., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9545–50 (1999).
24. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18**, 49–60 (2005).
25. Perdrizet, G. A., Artsimovitch, I., Furman, R., Sosnick, T. R. & Pan, T. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3323–8 (2012).
26. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681–4 (2010).
27. Treiber, D. K., Rook, M. S., Zarrinkar, P. P. & Williamson, J. R. Kinetic intermediates trapped by native interactions in RNA folding. *Science* **279**, 1943–6 (1998).
28. Noeske, J. *et al.* Interplay of ‘induced fit’ and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res.* **35**, 572–583 (2007).
29. Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J. Mol. Biol.* **359**, 754–68 (2006).
30. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. a. Folding of the adenine riboswitch. *Chem. Biol.* **13**, 857–68 (2006).
31. Larson, M. H. *et al.* A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–7 (2014).
32. Buck, J., Fürtig, B., Noeske, J., Wöhnert, J. & Schwalbe, H. Time-resolved NMR spectroscopy: ligand-

- p induced refolding of riboswitches.
- Methods Mol. Biol.*
- 540**
- , 161–171 (2009).
-
33. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination.
- Mol. Cell*
- 3**
- , 495–504 (1999).
-
34. Prinz, J.-H.
- et al.*
- Markov models of molecular kinetics: generation and validation.
- J. Chem. Phys.*
- 134**
- , 174105 (2011).
-
35. Buckstein, M. H., He, J. & Rubin, H. Characterization of Nucleotide Pools as a Function of Physiological State in
- Escherichia coli*
- .
- J. Bacteriol.*
- 190**
- , 718–726 (2008).
-
36. Greenleaf, W. J., Frieda, K. L., Foster, D. A. N., Woodside, M. T. & Block, S. M. Direct observation of hierarchical folding in single riboswitch aptamers.
- Science*
- 319**
- , 630–633 (2008).
-
37. Helmling, C.
- et al.*
- NMR Structural Profiling of Transcriptional Intermediates Reveals Riboswitch Regulation by Metastable RNA Conformations.
- J. Am. Chem. Soc.*
- 139**
- , 2647–2656 (2017).
-
38. Fürtig, B., Richter, C., Wöhnert, J. & Schwalbe, H. NMR spectroscopy of RNA.
- Chembiochem*
- 4**
- , 936–62 (2003).
-
39. Mok, K. H.
- et al.*
- Rapid sample-mixing technique for transient NMR and photo-CIDNP spectroscopy: applications to real-time protein folding.
- J. Am. Chem. Soc.*
- 125**
- , 12484–92 (2003).
-
40. Ederth, J., Mooney, R. A., Isaksson, L. A. & Landick, R. Functional Interplay between the Jaw Domain of Bacterial RNA Polymerase and Allele-specific Residues in the Product RNA-binding Pocket.
- J. Mol. Biol.*
- 356**
- , 1163–1179 (2006).
-
41. Landick, R., Wang, D. & Chan, C. L. Quantitative analysis of transcriptional pausing by
- Escherichia coli*
- RNA polymerase: his leader pause site as paradigm.
- Methods Enzymol.*
- 274**
- , 334–53 (1996).
-
42. Antoniewski, C., Savelli, B. & Stragier, P. The
- spoIIJ*
- gene, which regulates early developmental steps in
- Bacillus subtilis*
- , belongs to a class of environmentally responsive genes.
- J. Bacteriol.*
- 172**
- , 86–93 (1990).
-
43. Martin-Verstraete, I., Débarbouillé, M., Klier, A. & Rapoport, G. Levanase operon of
- Bacillus subtilis*
- includes a fructose-specific phosphotransferase system regulating the expression of the operon.
- J. Mol. Biol.*
- 214**
- , 657–71 (1990).
-
44. Monforte, J. A., Kahn, J. D. & Hearst, J. E. RNA folding during transcription by
- Escherichia coli*
- RNA polymerase analyzed by RNA self-cleavage.
- Biochemistry*
- 29**
- , 7882–90 (1990).

Figure legends

Figure 1: **Conformational states of the guanine-sensing riboswitch depending on transcript length.** a) Secondary structure of the full-length guanine-sensing riboswitch (GSW^{PATH}). In absence (*left*) and in presence of ligand (*right*) the aptamer closing helix (PA) and the terminator helix (TH) are formed, the anti-terminator helix (AT) is not present in either state. The only structural difference between the apo- and holo-states is the formation of a stable ligand binding pocket in the holo-state. The strands involved in the switching mechanism are colour-coded: aptamer strand (P, *black*), aptamer stabilizing strand (A, *blue*), switching strand (T, *red*), and terminator strand (H, *green*). Putative pause sites (PS1-PS3) are indicated, the sequence highlighted in grey is occupied by the polymerase⁴⁴; additionally, stable structured fragments are marked by arrows; b) Secondary structure of a truncated stably structured guanine-sensing riboswitch (GSW^{PAT}). In the absence of ligand two conformational states are populated in a 1:1 ratio each representing a functional (on/off) state of the riboswitch, c) G·U region of NOESY spectra of the full-length GSW^{PATH} (*upper panel*) and the truncated riboswitch GSW^{PAT} (*bottom*). G·U cross peaks are reporters for formation of PA (black/*blue*), P3 (*grey*), AT (*blue/red*) and TH (*red/green*), respectively; d) signal intensities of GSW^{fl} and GSW^{10-134} NOESY cross peaks in absence and presence of ligand. Errors were estimated from the noise of the respective spectra. The full-length riboswitch GSW^{PATH} adopts the terminator conformation irrespective of the ligand. The truncated GSW^{PAT} shows a heterogeneous fold in the absence and in the presence of the ligand, e) 10 % PAGE of the overnight transcription of the full length riboswitch in the absence (-lig) and presence (+lig) of ligand. The transcribed RNA fragments correspond to the full length (FL: 228 nt), the Terminator (T: 172 nt), the second pause site (PS2: 152 nt), the first pause site (PS1: 124 nt) and the aptamer (APT: 96 nt).

Figure 1 - **Figure Supplement 1: Assignment of GSW constructs.** All ^1H , ^1H -NOESY spectra were recorded at 283 K in 2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate (pH 6.2). a) Module design for chemical shift assignment of the full-length GSW with in the divide and conquer approach. The full-length GSW and the modules were measured by NMR spectroscopy separately. *Left*: GSW^{fl} (GSW^{PATH} , *grey*) in the terminator conformation, the aptamer domain GSW^{10-89} (GSW^{PA} , *blue*) and the terminator hairpin $\text{GSW}^{121-155}$ (TH, *green*) are indicated. *Right*: GSW^{fl} (GSW^{PATH} , *grey*) in the antiterminator conformation and the P4 module (AT, *red*). b) Overlay of the full-length GSW^{fl} (GSW^{PATH} , *black*), the aptamer domain GSW^{10-89} (GSW^{PA} ,

blue) and the terminator hairpin GSW¹²¹⁻¹⁵⁵ (TH, green) shows that GSW^{fl} (GSW^{PATH}) adopts the terminator conformation with formed aptamer and terminator hairpin. Assignment of the fragments can be transferred to GSW^{fl} (GSW^{PATH}). c) Left: Overlay of the truncated GSW¹⁰⁻¹³⁴ (GSW^{PAT}, black), the aptamer domain GSW¹⁰⁻⁸⁹ (GSW^{PA}, blue) and the P4 module (AT, red) reveals conformational heterogeneity of GSW¹⁰⁻¹³⁴ (GSW^{PAT}) with either helix P1(PA) (GSW^{10-134A}, GSW^{[PA]T}) or helix P4(AT) (GSW^{10-134B}, GSW^{[AT]T}) formed. Right: Conformations GSW^{10-134A} (=GSW^{[PA]T}) and GSW^{10-134B} (=GSW^{[AT]T}) of the truncated GSW¹⁰⁻¹³⁴ (=GSW^{PAT}). d) Spectra of the aptamer domain GSW¹⁰⁻⁸⁹ (GSW^{PA}, blue), the terminator hairpin GSW¹²¹⁻¹⁵⁵ (TH, green) and the P4 module (AT, red). Assignments are annotated in the spectra; the sequence of the P4 (AT) module is given.

Figure 1 – Figure Supplement 2: Conformation of full-length GSW. GSW^{fl} without (black) and with ligand (grey) adopts the terminator conformation. Experimental conditions are 2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate (pH 6.2). The resonance assignment of the terminator hairpin P6 is depicted in *bold letters*. In presence of ligand, reporter signals (U47, U49 and U51) from the binding pocket are detected. However, ligand binding induces slight chemical shift changes (arrows) but no conformational changes are observed in particular not for TH.

Figure 2: Models for intermediates during transcription. All NMR spectra were recorded at 283 K in 2 mM MgCl₂, 50 mM KCl, 25 mM potassium phosphate (pH 6.2). 1 equivalent of each RNA (GSW¹⁰⁻⁸⁹ (=GSW^{PA}), S2^{trans} (=T^{trans}) and S3^{trans} (=H^{trans}), respectively) was used, 4 equivalents of ligand were added. The following selective labelling scheme was explored: ¹⁵N-G,U GSW¹⁰⁻⁸⁹, ¹⁴N-G,U S2^{trans} and ¹⁴N-G,U S3^{trans}. Signals originating from GSW¹⁰⁻⁸⁹ and S2^{trans} or S3^{trans} were separated using in x-filter 1D experiments ¹⁴ (top). The ¹H, ¹⁵N-HSQC spectra (bottom) report on the interactions in the aptamer domain. a) In the elongated aptamer domain GSW¹⁰⁻⁸⁹, the helix PA was formed (U17-G79) and the loop-loop interaction reporter G32 was detected. b) GSW¹⁰⁻⁸⁹ and ligand (1:4) ligand binding was monitored by appearance of signals U47, U49 and U51. c) GSW¹⁰⁻⁸⁹ and S2^{trans} (1:1): Sequence S2^{trans} caused the PA reporter signals U17(P1) and G79(P1) to decrease, AT formation was followed by appearance of signals of G79(P4), U81(P4) and G82(P4). d) GSW¹⁰⁻⁸⁹, S2^{trans} and ligand (1:1:4): Addition of ligand to GSW¹⁰⁻⁸⁹-S2^{trans} resulted in dissociation of the complex (decreasing signals for G79(P4), U81(P4) and G82(P4) signals) and reformation of the PA helix (U17(P1) and G79(P1) signals). Ligand binding reporters U47, U49 and U51 were detected. However, in presence of 4 equivalents of ligand, AT helix reporter

signals were significant. e) GSW¹⁰⁻⁸⁹, S2^{trans} and S3^{trans} (1:1:1) Addition of S3^{trans} to GSW¹⁰⁻⁸⁹-S2^{trans} resulted in complete dissociation of the complex (G79(P4), U81(P4) and G82(P4) signals) and reformation of the PA helix (U17(P1) and G79(P1) signals). In contrast to ligand addition, 1 equivalent of S3^{trans} was sufficient to disrupt the antiterminator mimic. f) GSW¹⁰⁻⁸⁹, S2^{trans}, S3^{trans} and ligand (1:1:1:4): Ligand binding to GSW¹⁰⁻⁸⁹ in presence of the terminator helix P6 (=TH) equals ligand binding to GSW¹⁰⁻⁸⁹ (GSW^{PA}) alone (b).

Figure 2 – **Figure Supplement 1: ITC measurements and K_d values of GSW^{PA} (a), GSW^{PAT} (b) and GSW^{PATH} (c).**

ITC measurements were performed with a Microcal VP ITC (Northampton, MA USA) at 10°C. A 217 µM solution of ligand (hypoxanthine) was titrated to a 15 µM solution of RNA using 25-42 injections. Buffer conditions were 2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate, pH 6.2. The data was analyzed with the Origin ITC software (OriginLab, Northampton, MA USA) assuming a single binding site. The K_d values are given in µM.

Figure 3: **RNA refolding and ligand binding kinetics.** a) Schematic overview of the performed kinetic experiments (indicated with letters b-h) to characterise different transcription intermediates. The aptamer domain GSW^{PA} is depicted in *black*, switching sequences A (*blue*), T (*red*) and H (*green*) are colour-coded. Dashed sequences are neglected in the antisense oligonucleotide approach. b-h) results of the kinetic experiments, lettering according to a). ¹⁵N- (*left*) and ¹⁴N-filtered (*right*) 1D spectra before (*top*) and after (*bottom*) the kinetic experiments are depicted. An exemplary time trace reporting on the formation of each structural motif involved in the rearrangement is given, respective signals are marked with red arrows. i) Rates obtained from signal traces of resolved imino proton resonances. Letters refer to the kinetic experiments as shown in a). For all structure motifs (ligand binding, PA, AT and TH formation, respectively) several signals (number indicated below) were analysed, averaged rates are given in *bars*, rates of individual base pairs are indicated with *diamonds*. Colour-coding refers to AT association (*green*), AT dissociation (*red*), ligand binding (*dark grey*), aptamer formation (*blue*), terminator association (*orange*) and H^{trans}-intrinsic unfolding (*light grey*) which is irrelevant for riboswitch function. Residues with a single rate were fitted mono-exponentially, a bi-exponential fit function was applied for residues with two distinguishable rates. For exact values and errors see Supplementary Tables 1 to 7.

Figure 3 - **Figure Supplement 1: Folding of sequences T^{trans} and H^{trans}**. a) Imino proton region of the ¹H-NMR spectra of the oligonucleotide sequences T^{trans} (*top*) and H^{trans} (*bottom*). NMR spectra were recorded at 283 K in 2 mM magnesium chloride, 50 mM potassium chloride, 25 mM potassium phosphate (pH 6.2). The sequence T^{trans} shows intrinsic folding, intrinsic interactions of H^{trans} are minor.

Figure 3 - **Figure Supplement 2: Ligand-independent dissociation of helix AT in** kinetic experiment d. Rates obtained from signal traces of resolved imino proton resonances.

Figure 4: Time resolved Transcription: 8% PAGE of time resolved transcriptions using the *E. coli* (EC holo) and the *B. subtilis* (BS holo) RNAPs in the absence (-ligand) and presence (+ligand) of ligand. The time points of transcription stops are indicated in (s). The positions of ³²P 5'-labeled markers (pBR322 MspI digested) are indicated on the left hand side. Prominent paused and terminated bands are indicated on the right hand side. 7 major RNA-fragments could be identified: The run-off transcript or full-length RNA (FL), the premature termination fragment (Gsw^{PATH}), the second pause-site (PS2), the first pause-site (PS1) and three pausing fragments (RNA95 and RNA77 for *E. coli* RNAP transcriptions and RNA90 and RNA77 for *B. subtilis* RNAP transcriptions). Over time, both RNAPs transcribe the DNA-template, generating RNA-fragments of increasing size. A pausing event is characterized by signal increase and by a fast increase of the signal followed by a slower decrease (e.g. RNA77). FL and Gsw^{PATH} show a strong accumulation over time and when ligand is added, the signal intensity of FL is decreased whereas the signal intensity of Gsw^{PATH} is increased.

Figure 4 - **Figure Supplement 1: 3'-mapping and GTP* increase:** Transcription was performed using the *E. coli* RNAP in the absence (EC) and presence of 3'-deoxy ATP (dA), CTP (dC), GTP (dG) and UTP (dU), respectively and compared to transcriptions using the *B. subtilis* RNAP with different amounts of radioactively labeled GTP (GTP*) or DNA template in the absence and presence of ligand. The gel shows several transcription abortion products which end with a 3'-deoxy U and which migrate slightly faster than the 110 RNA fragment. When compared to the sequence, this poly-U stretch corresponds to the bases T107 to T112. However, it can't be clearly stated on which nucleotide the 110 RNA ends. It was therefore decided to call this fragment 110.

Figure 4 - Figure Supplement 2: Intensity plots of the normalized pause signals: **a** The signal intensities of the *E. coli* RNAP transcripts PS2, PS2, RNA95 and RNA77 were analyzed as shown by Landick et al⁴¹. in the absence (black) and presence (red) of ligand and plotted over time. For normalization, the intensity of an RNA signal (RNA) was divided by the sum of all RNAs of the same length and longer (RNAP).

b Pausing plots of the *B. subtilis* RNAP transcripts. The steeper the pausing-plot, the shorter the dwell-time (τ) of the pause site. Pause-sites with high τ have a higher impact on transcription kinetics. Addition of ligand seems to have a higher impact on the pausing of the *E. coli* RNAP. However, the differences of the plots are within the errors.

Figure 4 - Figure Supplement 3: Distance between aptamer and terminator hairpin in purine riboswitches.

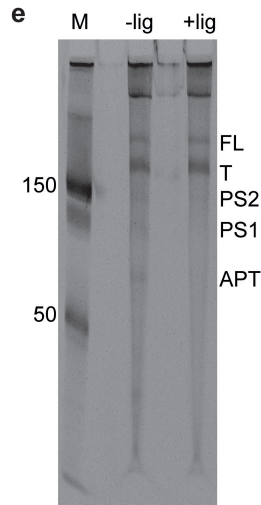
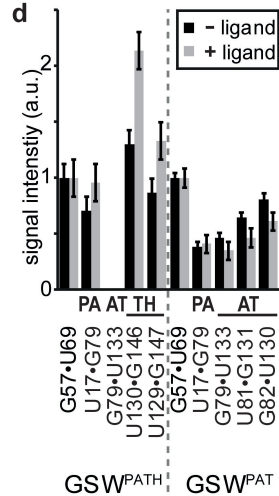
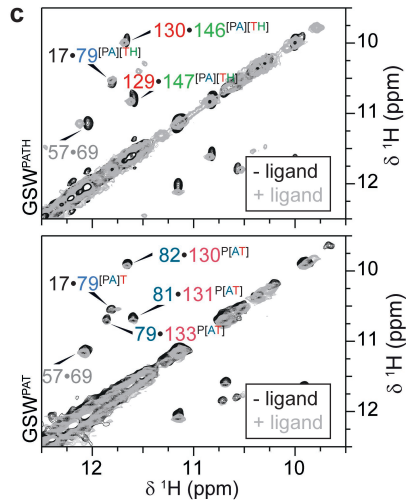
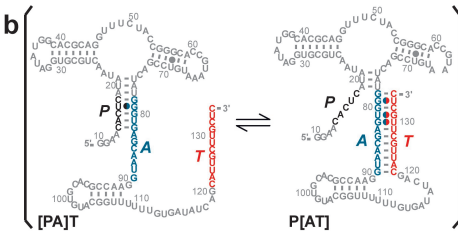
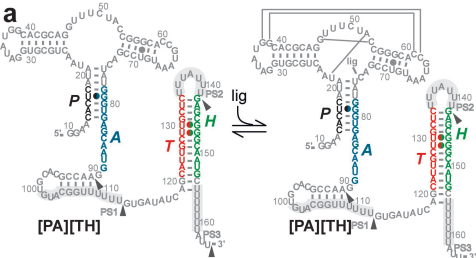
The list of 133 riboswitches was taken from the seed dataset of the Rfam database entry for the Purine riboswitch family (RF00167). Riboswitch sequences were retrieved from the EMBL European Nucleotide Archive (ENA). The Rfam database contains only the aptamer regions of the purine riboswitches but lacks the expression platform. Therefore, the database does not annotate the terminator location. The web service ARNold, which uses Erpin⁴⁵ and RNAmotif⁴⁶ to predict terminators, was run on all riboswitch sequences to determine the terminator position. Sequences upstream of the coding sequence (CDS) as identified by annotations in ENA were used as input for ARNold. ARNold identified terminators between the aptamer and the CDS in 35 of the 133 sequences. The distances between aptamer and terminator were calculated based on the secondary structure annotation from Rfam and the output from ARNold. Predicted terminators inside the aptamer sequence were excluded from the data.

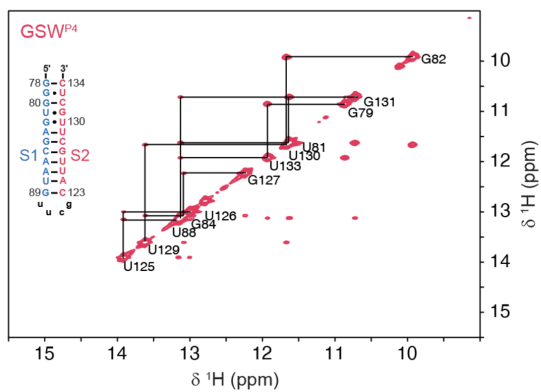
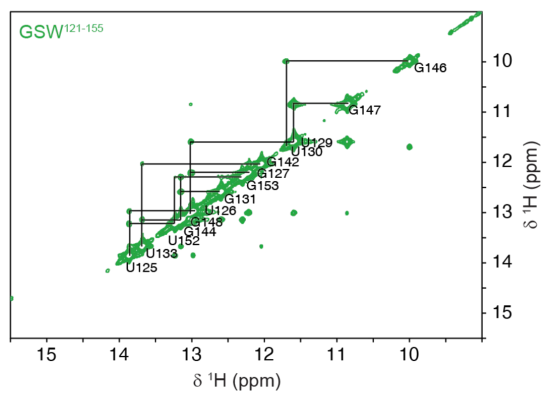
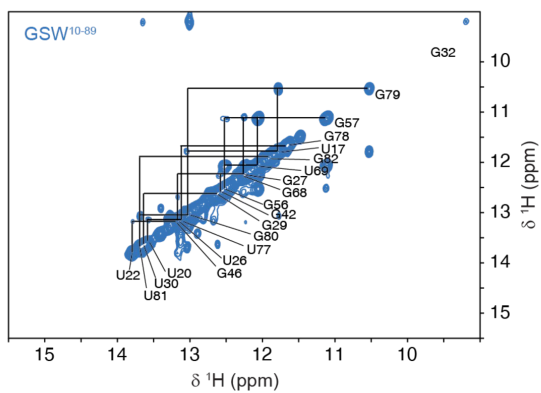
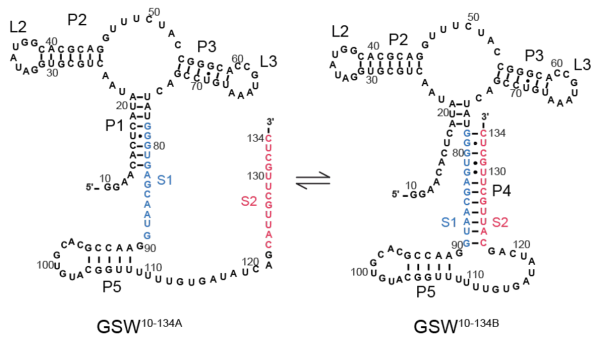
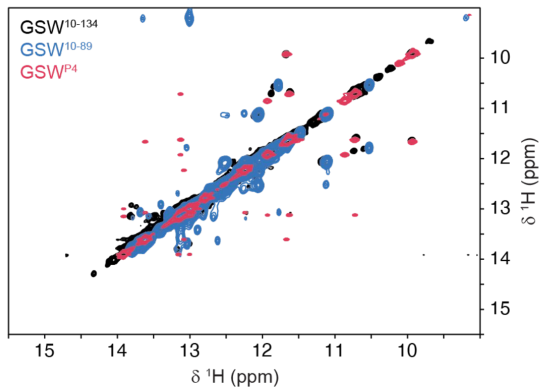
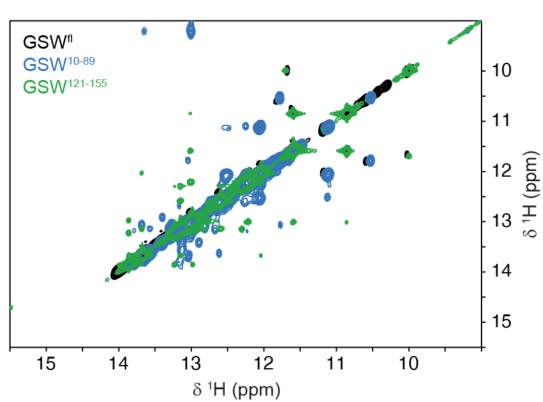
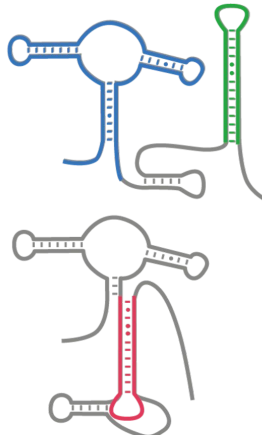
Figure 5: Simulation of co-transcriptional folding pathways **a)** Conformational states of GSW in context of the transcription progress are shown from left to right. K_d values of the transcript intermediates are indicated. Ligand binding can occur as soon as the aptamer domain is transcribed and consequently locks A (blue) in the PA helix resulting in the population of a single conformation during the transcription process and subsequently in transcription termination. The aptamer domain is synthesized first. As transcription continues, the free mRNA adopts the metastable antiterminator conformation stabilized by AT interaction (blue and red, respectively) which refolds to the terminator conformation after the riboswitch is completed. Pause site PS1

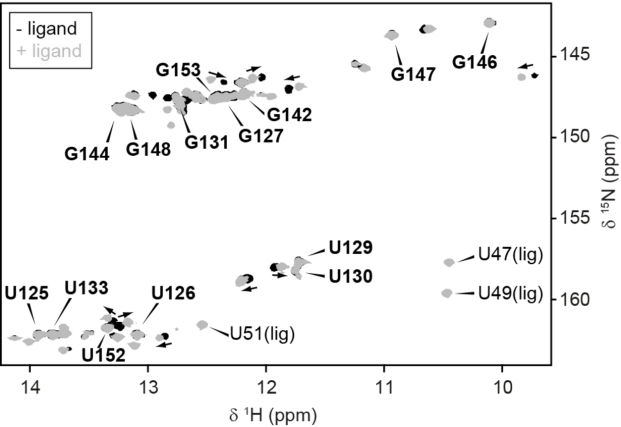
(*grey cylinder*) increases the available time window for ligand binding. Co-transcriptional refolding processes are indicated with their respective life times by the *black arrows*. Population of the conformational states under different conditions (the time plotted on the x-axis is the time of transcription with $t=0$ at nucleotide position 75) (b) without ligand and without pausing, (d) without ligand and with pausing at PS1 and (f) with ligand and without pausing and (g) in presence of ligand and with pausing at PS1. Ratio of the probabilities to populate the on- versus the off-state as derived by kinetic simulation of the switching mechanism in absence (c, e) and in presence (g, i) of ligand. If the polymerase does not pause between the synthesis of the stretches A and T, GSW adopts the on-state irrespective of ligand (c, g) whereas pausing at PS1 enables GSW to respond to its ligand and dominantly adopts the off-conformation in the presence of ligand (i).

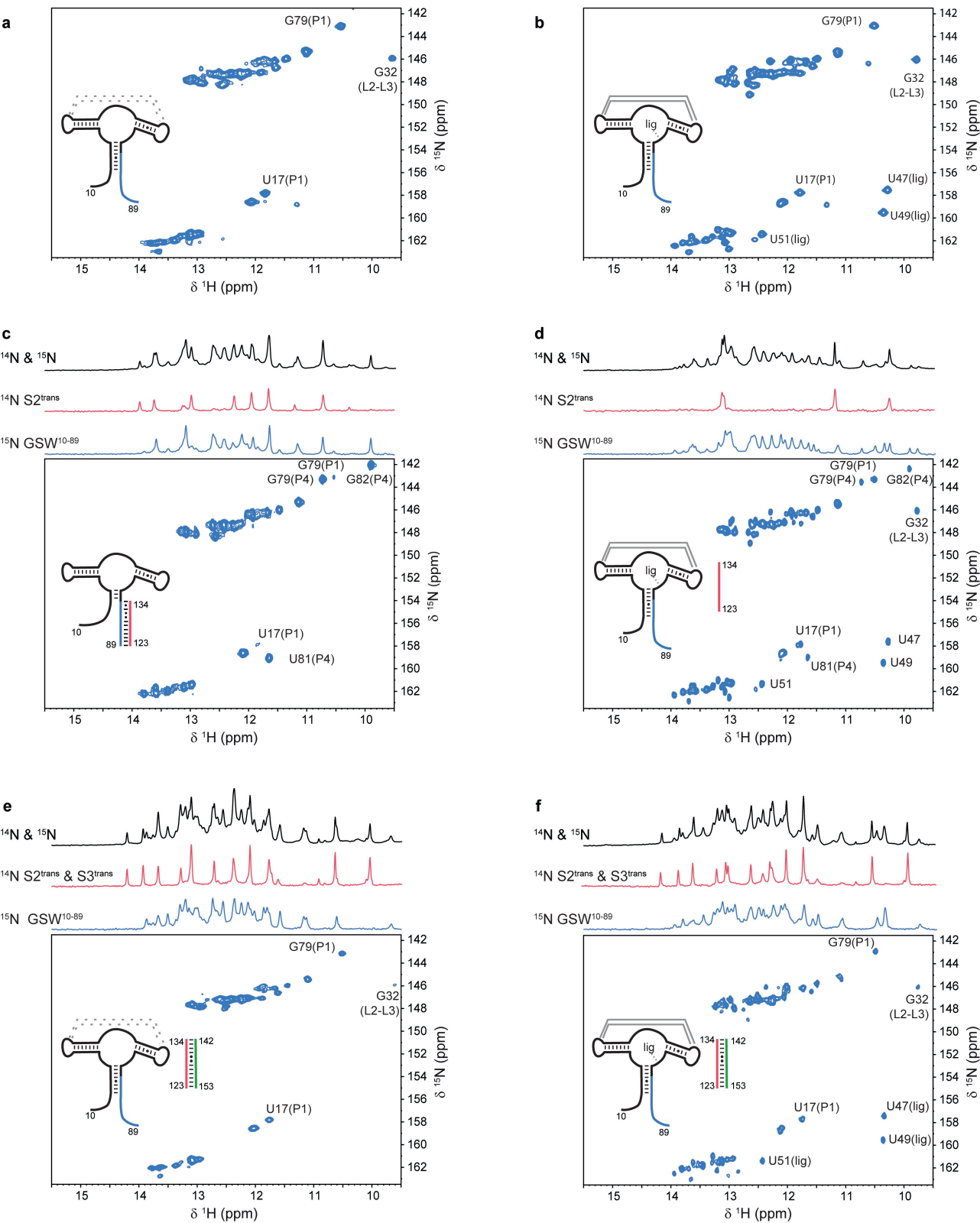
Figure 6: *in vivo* Pause site characterisation a) Regulation of β -galactosidase reporter gene expression by wt-GSW and pause site 1 mutants. Nucleotide exchanges or insertion of residues to generate the PS1 mutants M1–M4 are indicated. Enzyme activity for cells grown in absence (*black bars*) or presence of 0.5 mg mL^{-1} guanine (*grey bars*), respectively; the dynamic range corresponds to the ratio of enzyme activity in absence and presence of ligand. b) Dose-dependent repression of β -galactosidase expression for wt GSW (*black*), M3 (*green*) and M4 (*purple*). Nucleotide exchanges to generate the PS1 mutants M3 and M4 are indicated. Cultures were grown with increasing concentrations of guanine. Deletion of PS1 reduces riboswitch efficiency, whereas it is increased in case of elongation of PS1. However, both mutations do not significantly alter the half maximal effective concentration EC50.

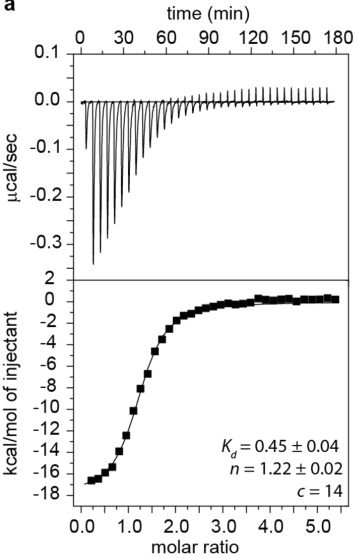
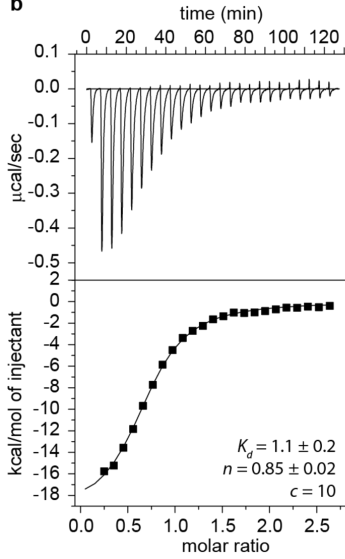
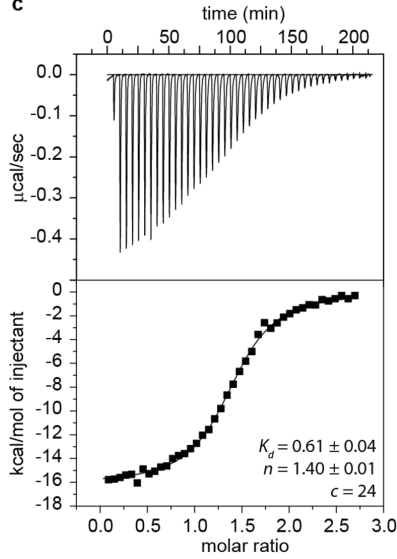
Supplementary File 1: This file contains rates obtained by Kinetic NMR experiments, the exact nucleotide sequence of the DNA template used in transcriptional assays, as well transcription rates and pause site characteristics.

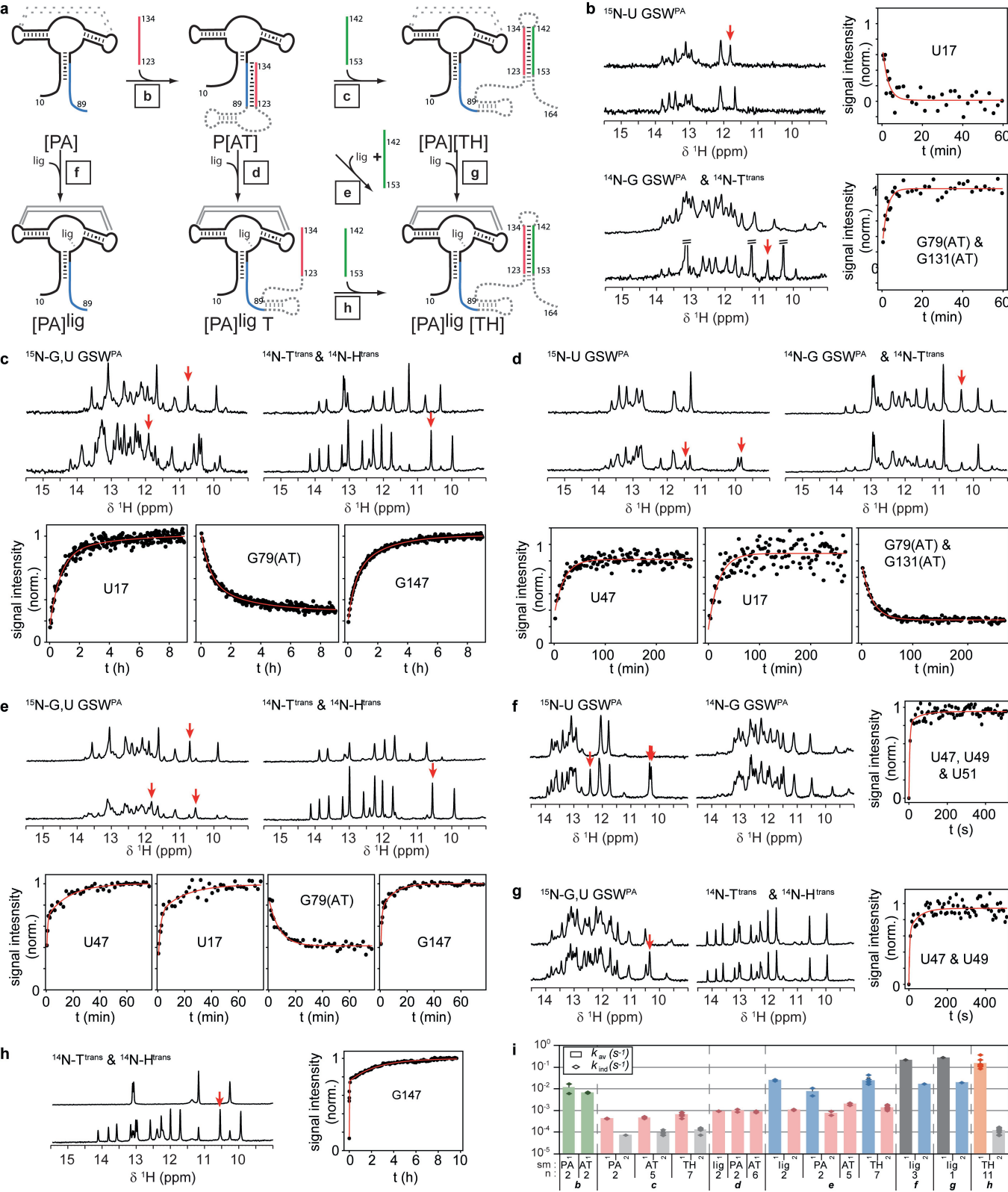


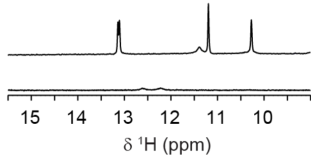
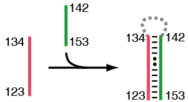
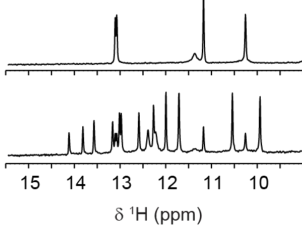
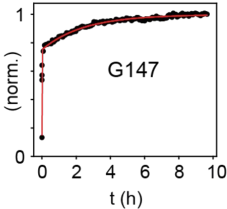


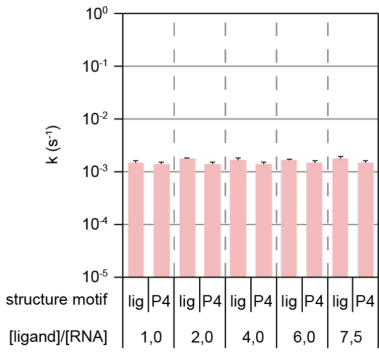
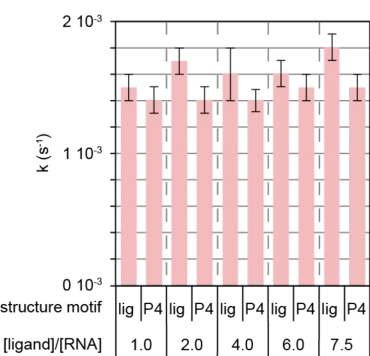




a**b****c**



a $S2^{trans}$ $S3^{trans}$ **b** ^{14}N - $S2^{trans}$ & ^{14}N - $S3^{trans}$ signal intensity
(norm.)



Ec RNAP

Bs RNAP

- ligand

+ ligand

Time [s]

Time [s]

- ligand

+ ligand

Time [s]

Time [s]

0 5 10 20 30 60 120 300 600 600 + 1 mM NTP

0 5 10 20 30 60 120 300 600 600 + 1 mM NTP

M

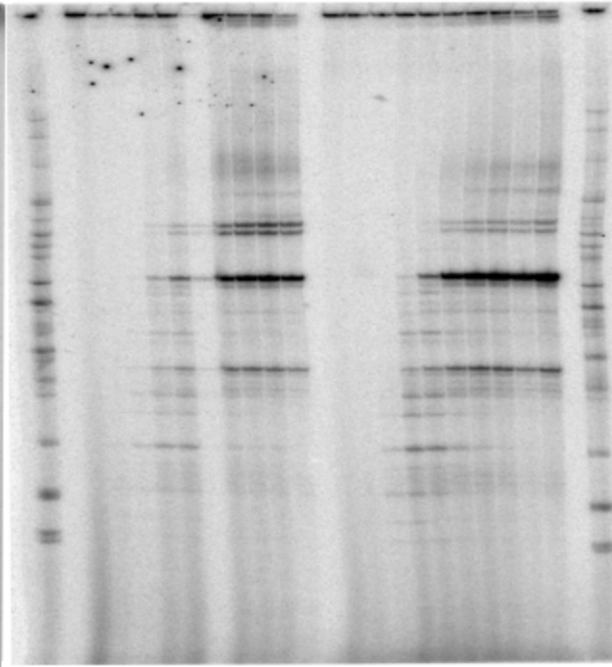
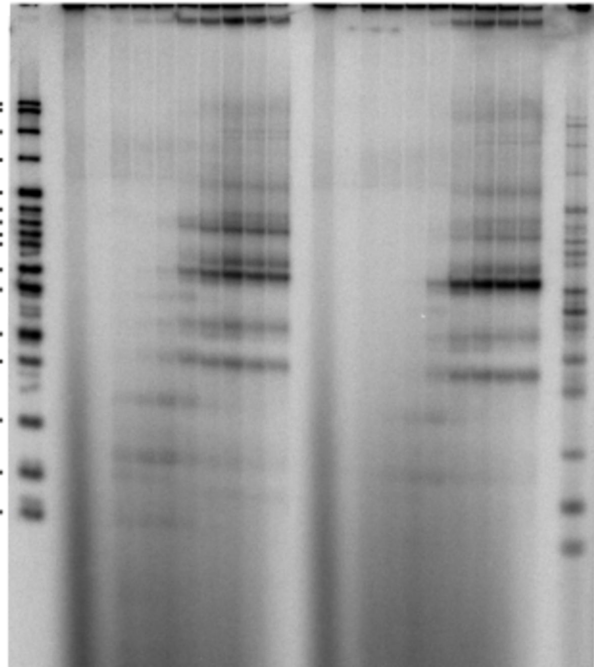
M

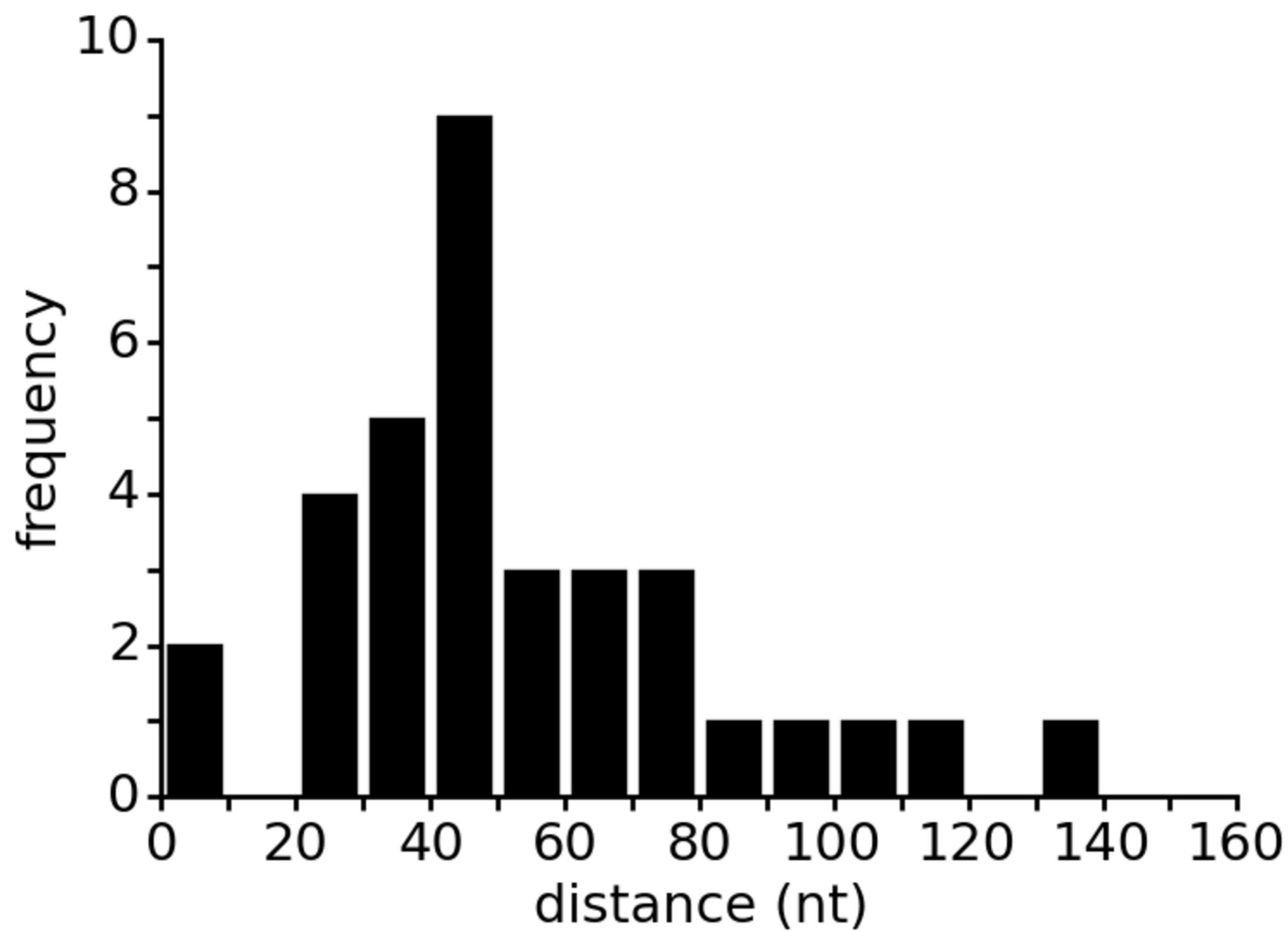
M

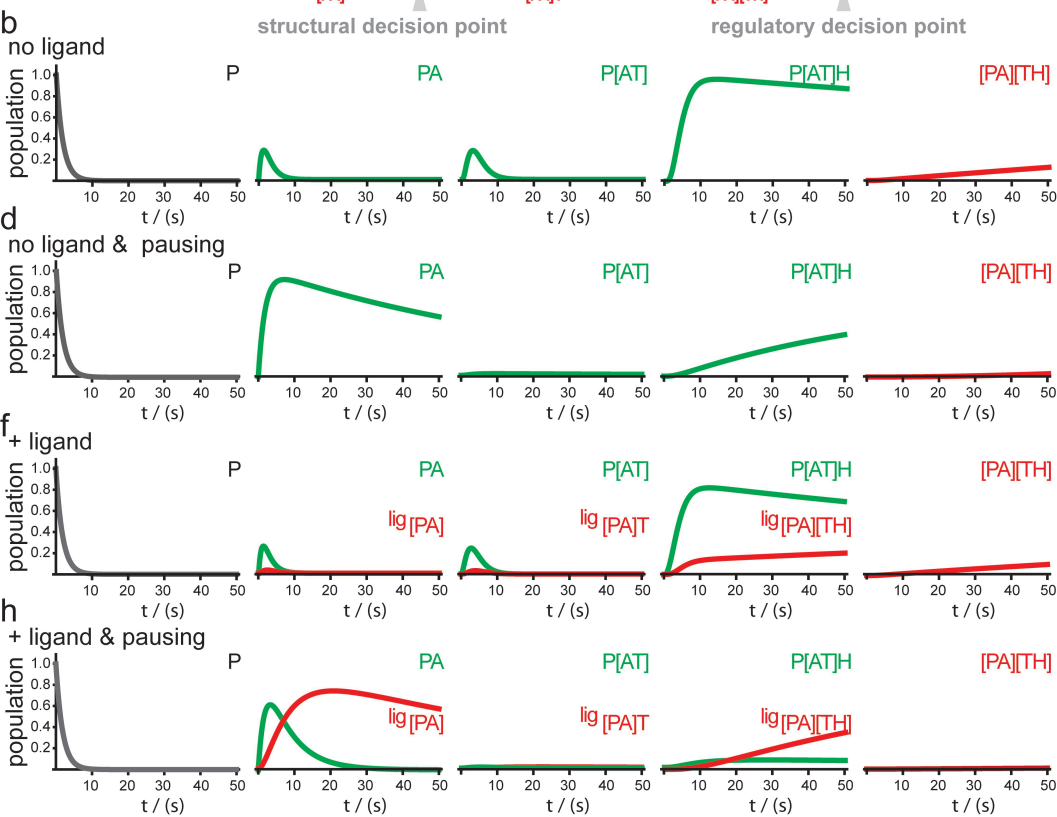
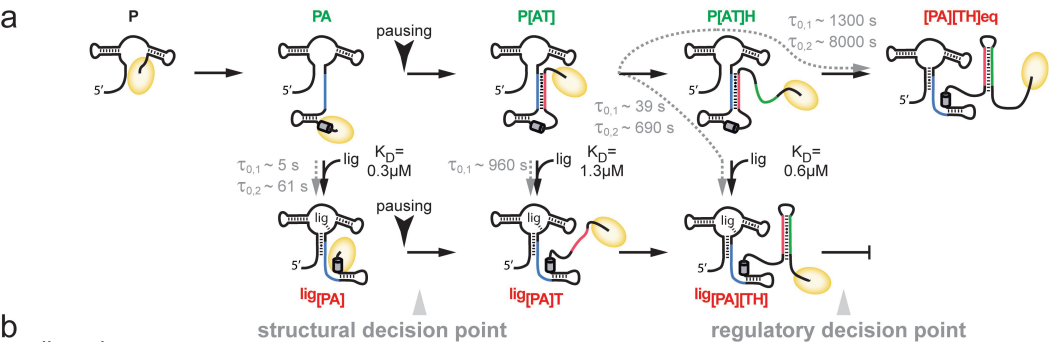
M

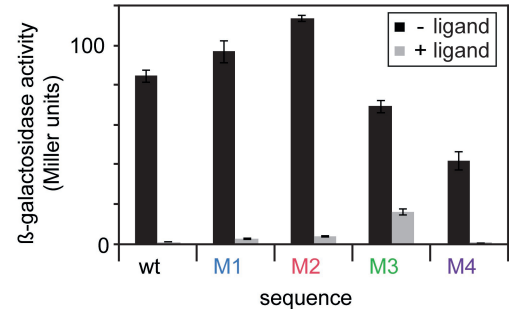
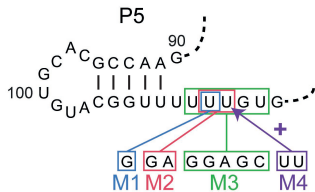
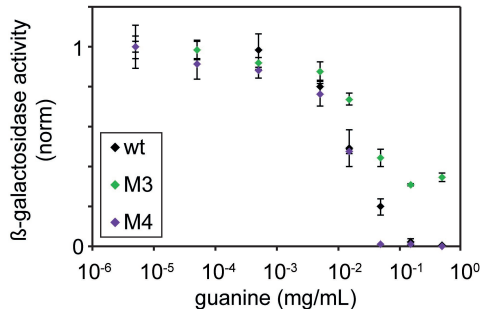
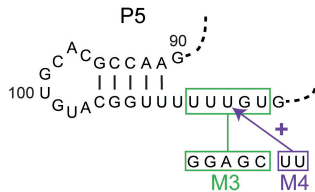
527
404
307
238/242
217
201
190
180
160
147
123
110
90
76
67

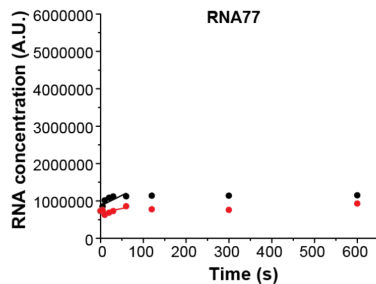
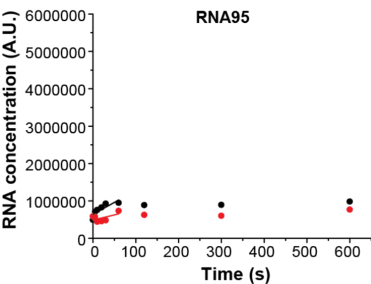
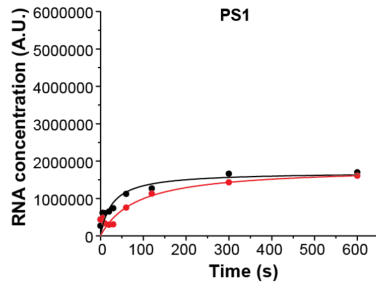
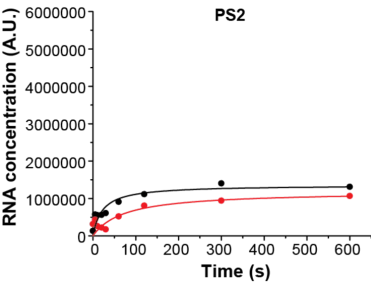
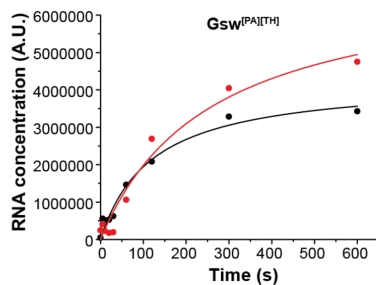
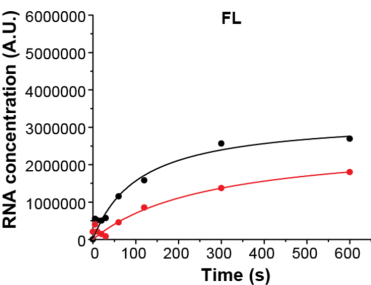
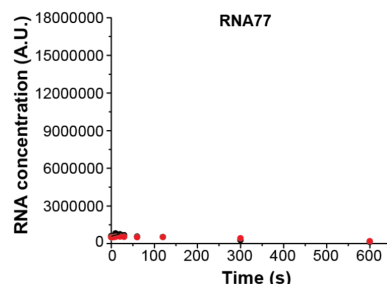
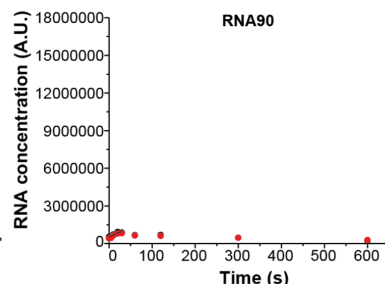
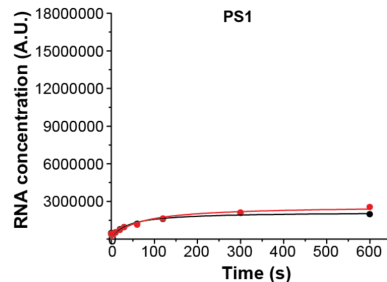
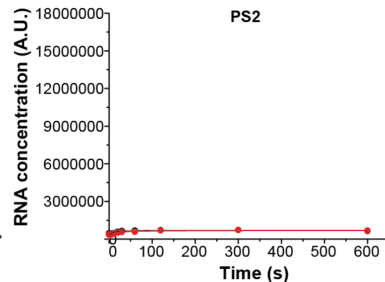
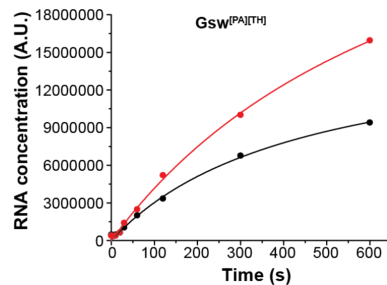
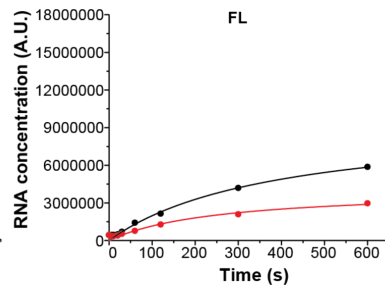
214: FL
164: Gsw[PA][TH]
141: PS2
110: PS1
95: RNA95
90: RNA90
77: RNA77







a**b**

a**b**

BS

