IMPACT DES PAUSES TRANSCRIPTIONNELLES SUR LE REPLIEMENT DES RIBORÉGULATEURS

Par

Adrien CHAUVIER

Thèse présentée au Département de biologie en vue
de l'obtention du grade de docteur ès sciences (Ph.D.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, juillet 2017

Le 24 Juillet 2017

Le jury a accepté la thèse de Monsieur Adrien Chauvier
dans sa version finale

Membres du jury

Professeur Daniel Lafontaine
Directeur de recherche
Département de Biologie

Professeur Eric Massé
Membre interne
Département de Biochimie

David Rueda
Professeur
Membre externe
Département de Médecine
Imperial College London

Professeur Vincent Burrus
Président-rapporteur
Département de Biologie

SOMMAIRE

Pour s'adaptater à l'environnement et répondre aux besoins énergétiques, tous les organismes doivent pouvoir contrôler l'expression de certains gènes. La trancription constitue le premier échelon de cette régulation en déterminant le niveau d'ARNm produit au cours du temps. Chez les procaryotes, une seule ARNp est responsable de la synthèse de l'ensemble des ARN de la cellule et celle-ci est soumise à différents types de régulation. Ce contrôle peut s'effectuer à tous les niveaux par des processus faisant intervenir bon nombre de facteurs externes ou intrinsèques à la transcription comme les pauses de l'ARNp.

Les riborégulateurs sont des ARN structurés majoritairement retrouvés dans la région 5' non traduite des ARNm bactériens qui vont réguler l'expression du gène situé en aval. Suite à la liaison d'un métabolite particulier, appelé ligand, le riborégulateur change de conformation induisant une réponse directe qui déterminera si un gène est exprimé ou non.

Au cours de mes travaux j'ai établi le lien qu'il existait entre les pauses de l'ARNp au cours de la transcription et la structure des riborégulateurs. En prenant pour modèles deux riborégulateurs liant le TPP, j'ai identifié une conformation des riborégulateurs qui est réfractaire à la liaison du ligand. Cette structure appelée «Anti-P1» se forme lorsque l'ARNp pause à la fin du riborégulateur, ce qui détermine une fenêtre de liaison co-transcriptionnelle du ligand.

REMERCIEMENTS

Cette section est à mon sens la plus difficile à écrire car je ne souhaiterais oublier personne. Je tiens tout d'abord à remercier particulièrement mon directeur de recherche Daniel Lafontaine pour son soutien moral et intellectuel, sa présence et son amitié tout le long de mon doctorat, les opportunités offertes ainsi que nos discussions (scientifiques) jusqu'à pas d'heure. Mes conseillers et membres du jury, Vincent Burrus et Eric Massé qui ont dû subir mes présentations et rapports interminables pendant 4 ans et demi. Merci au Pr. David Rueda de donner de son temps pour lire ce manuscrit et son expertise quant à son contenu. Les membres de mon laboratoire (par ordre alphabétique, comme ça pas de faché) : Anne-Ma, Antony, Auréliane, Dom, Erich, Fatima, Fred, J-C, J-F, Jo, Julien, Laurène, Max, Marc-André, Marie-Ann, Mel, Pat', Pierre, Seb. Pour tous les fous rires dans le laboratoire et jokes plus ou moins subtiles, mais bien sûr nos fameuses sorties de lab'. Je peux dire que chaque année, je passais Noël en famille le temps d'un repas. Le département de biologie et plus particulièrement les labos Rodrigue et Gévry. Ma branche écologiste Pascal et Michaël. Mon ami de longue date Bruno. Merci à Jean-Philippe (Phil) pour son amitié et qui restera une de mes plus belles rencontres au Québec. L'équipe Gadal à Toulouse pour m'avoir donné le goût de la science au début de mes études.

D'un côté plus personnel je tiens à remercier ma femme Stéphanie Ratovonarivo pour son soutien sans faille et mon père qui m'aura toujours soutenu dans mes choix de vie.

Merci à vous tous d'avoir rendu mon doctorat exceptionnel en tout point !

Tchimbé rèd pa moli !

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS

ADN :	Acide désoxyribonucléique
ARN :	Acide ribonucléique
ARNm :	ARN messager
ARNp :	ARN polymérase
ARNr :	ARN ribosomique
ARNt :	ARN de transfert
ATP :	Adénosine tri-phosphate
ATPase :	Adénosine tri-phosphatase
B. subtilis :	*Bacillus subtilis*
Da :	Dalton
di-AMPc :	Di-adénosine monophosphate cyclique
di-GMPc :	Di-guanosine monophosphate cyclique
E. coli :	*Escherichia coli*
Facteur σ :	Facteur sigma
FMN :	Flavine mononucléotide
FRET :	*Förster Resonance Energy Transfer*
GlcN6P :	Glucosamine-6-phosphate
HDV :	*Hepatitis delta virus*
HMP :	Hydroxy-méthyl-pyrimidine
Kd :	Constante de dissociation
Mg2+ :	Ion magnésium
miRNA :	Micro-ARN
NET-seq :	*Nascent Elongating Transcript sequencing*
NTP :	Nucléotide tri-phosphate
Opéron his :	Opéron histidine
Opéron trp :	Opéron tryptophane
Ops :	*Operon polarity suppressor*
PréQ1 :	7-aminométhyl-7-déazaguanine
RBS :	*Ribosome Binding Site*

RNase :	Ribonucléase
Rut :	*Rho Utilization Site*
S. enterica :	*Salmonella enterica*
SAM :	S-adénosyl-méthionine
SAH :	S-adénosyl-monocystéine
SD :	Shine-Dalgarno
SHAPE :	*Selective 2'-Hydroxyl Acylation analyzed by Primer Extension*
siRNA :	Petit ARN interférent
sm-FRET :	*single molecule FRET*
TPP :	Thiamine pyrophosphate
TRAP :	*Tryptophan Regulated Attenuation Protein*
UTR :	*Untranslated Region*

LISTE DES TABLEAUX

CHAPITRE 1

CHAPITRE 2

LISTE DES FIGURES

INTRODUCTION

CHAPITRE 1

CHAPITRE 2

INTRODUCTION

1. La régulation génique et l'ARN

Chez tous les organismes, réguler l'expression des gènes est primordial pour s'adapter à l'environnement. En effet, exprimer ou réprimer un gène dans certaines conditions détermine l'accessibilité des ressources nécessaires à la survie et constitue le premier niveau de maintien de l'intégrité cellulaire. L'information génétique d'une cellule est renfermée par l'acide désoxyribonucléique (ADN) et ce dernier est répliqué lors de la division cellulaire. L'ADN est transcrit en acide ribonucléique (ARN) par une enzyme spécialisée : l'ARN polymérase (ARNp), pour enfin être traduit en protéine par un complexe composé à la fois d'ARN et de protéines : le ribosome. L'ARN messager (ARNm), ayant une courte durée de vie a été longtemps prédit comme n'étant qu'un intermédiaire entre l'ADN et les protéines, transportant seulement l'information pour la synthèse protéique.

1.1. Historique

Dès 1967, Carl Woese propose que certaines structures complexes d'ARN pourraient avoir une activité catalytique (Woese, 1967). A partir des années 1980, les travaux sur l'épissage (Kruger et al., 1982) et sur le complexe Ribonucléase P (RNase P) (Guerrier-Takada et al., 1983), ont démontré les activités auto-catalytiques de l'ARN. Ces ARN ont été nommés ribozymes (contraction de «*Ribonuclease enzyme*») et par la suite, d'autres ont été identifiés comme le «*Hammerhead*» ou celui du virus de l'hépatite (HDV) (Doherty and Doudna, 2001). Ainsi l'ARN est non seulement capable de véhiculer l'information génétique, mais aussi d'accomplir des activités catalytiques. De ce fait, il a été proposé que l'ARN pourrait être à l'origine de la vie, avec la première utilisation du

terme «monde à ARN» par Walter Gilbert (Gilbert, 1986). Selon cette hypothèse, l'ARN aurait évolué d'une part en une matrice plus stable, l'ADN, pour porter l'information génétique et d'autre part en composantes catalytiques plus efficaces : les protéines.

1.2. L'ARN acteur de la régulation génique

Chez de nombreux organismes, la grande majorité du génome est constituée de séquences non codantes, c'est-à-dire qu'elles ne vont pas conduire à la production de protéines. Alors que chez l'humain, seulement 2 % du génome code pour des protéines (Mattick, 2004) il a été démontré qu'environ 40 % de celui-ci est transcrit en ARN (Matera et al., 2007), laissant penser que l'ARN non-codant jouerait un rôle tout aussi important que les protéines. Les ARN non-codants ont des fonctions très variées, on pourrait citer les ARN de transfert (ARNt) ou encore les ARN ribosomiques (ARNr) directement impliqués dans une fonction essentielle pour la cellule : la traduction par le ribosome. Le fait que ces ARN ne codent pas pour des protéines n'exclut pas qu'ils puissent jouer un rôle de régulation. En effet, les micro-ARN (miRNA) et les petits ARN interférents (siRNA) sont des éléments dits «régulateurs en *trans*», c'est à dire qu'ils vont agir sur une autre molécule et en sont l'exemple parfait (Matera et al., 2007). Chez les procaryotes, environ une soixantaine de ces petits ARN de régulation ont été identifiés chez *Escherichia coli* (*E.* coli) et dont le tiers interagirait avec la protéine chaperonne Hfq pour les aider à réguler le bon ARNm (Gorski *et al.* 2017). D'autres fonctions régulatrices de l'ARN se déroulent plutôt en *cis,* c'est à dire sur la même molécule d'ARNm. Un exemple de ce type de régulation fait intervenir la liaison d'un autre ARN ou bien encore d'un métabolite (Jones & Ferré D'Amaré 2017).

2. La transcription chez les procaryotes

La transcription est la première étape du processus de l'expression des gènes. Elle consiste à générer un brin d'ARN complémentaire à la matrice d'ADN. Ce mécanisme

est retrouvé chez tous les organismes et est réalisé par des ARNp dépendantes de l'ADN. La transcription est soumise à différents niveaux de régulation au cours des trois grandes étapes qui la composent : l'initiation, l'élongation et la terminaison.

2.1. L'ARN polymérase bactérienne

Chez les procaryotes, une seule ARNp est responsable de la synthèse de l'ensemble des ARN de la cellule. L'enzyme se compose de cinq sous-unités ($\alpha_2\beta\beta'\omega$) avec une masse d'environ 400 000 Daltons (Da). Ces sous-unités sont relativement conservées de la bactérie à l'humain en terme de séquences, structures et fonctions (Ebright, 2000). Les sous-unités ß et ß' interagissent ensemble et forment une structure ressemblant à une pince dont la base est le site actif. α_1 et α_2 interagissent respectivement avec ß et ß' et sont impliquées dans les interactions avec l'ADN ou bien encore avec d'autres protéines comme la protéine Rho (discuté dans la section 2.3.2) (Kainz and Gourse, 1998). A ce jour, de nombreuses informations structurales sur les différentes ARNp sont disponibles grâce à la résolution des cristaux de l'enzyme de *Thermus aquaticus* (Zhang et al., 1999), *Saccharomyces cerevisiae* pour l'ARNp I et II (Cramer et al., 2001; Engel et al., 2013) ainsi que chez les archées (Hirata et al., 2008).

Au cours de la transcription, l'ARNp interagit directement avec les acides nucléiques de l'ADN et de l'ARN définissant ainsi la stabilité du complexe transcriptionnel (Vassylyev et al., 2007a) (Figure 1).

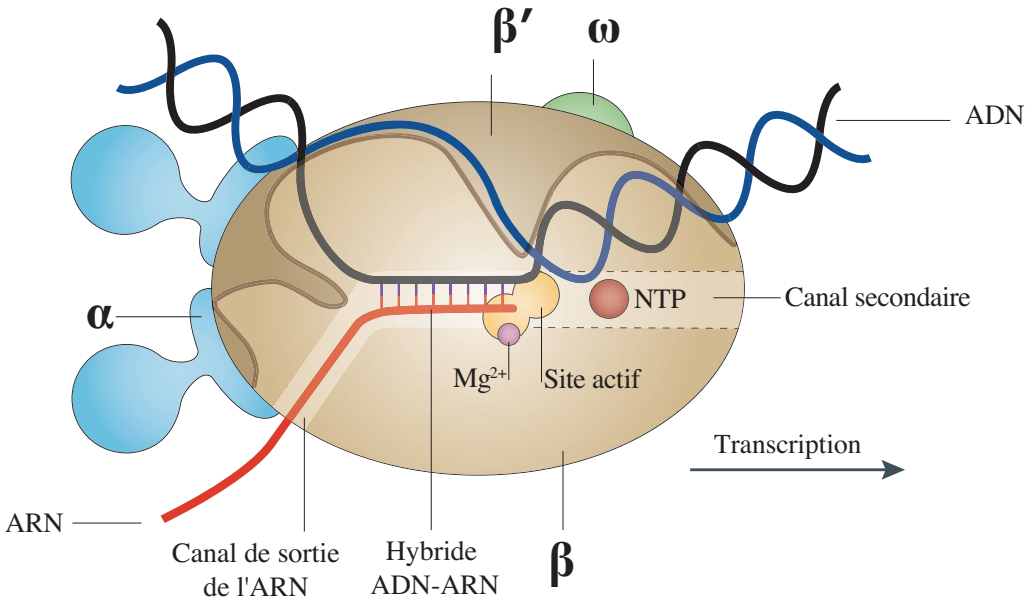


Figure 1 : L'ARN polymérase bactérienne.

Schéma montrant les différentes composantes du complexe transcriptionnel en cours d'élongation. NTP : nucléotide triphosphate, Mg2+ : ion magnésium. Adaptée de (Santangelo and Artsimovitch, 2011).

La «pince» formée par les sous-unités ß et ß' fixe l'ADN en aval et facilite l'ouverture de la double-hélice (Zhang et al., 1999). Le brin dit matrice de l'ADN passe ainsi dans le canal primaire pour atteindre le site actif de l'enzyme pour la synthèse d'ARN. Par la suite, environ 8-9 nucléotides d'ARN interagissent avec l'ADN au sein de l'ARNp formant l'hybride ADN-ARN. Enfin, l'ARN est expulsé de l'enzyme par le canal de sortie de l'ARN au niveau de la sous-unité ß et l'ADN retourne sous forme de double-hélice en arrière de l'ARNp (Vassylyev et al., 2002).

La «core» enzyme telle que définie est active mais incapable d'initier la transcription sans la fixation d'un autre facteur : le facteur sigma (facteur σ). Le complexe formé est ce que l'on appelle l'holoenzyme, qui va reconnaître la région promotrice sur l'ADN pour initier la transcription (Burgess et al., 1969; Travers and Burgess, 1969).

2.2. Initiation de la transcription

La phase d'initiation de la transcription commence par la fixation de l'holoenzyme sur les régions hexamèriques caractéristiques du promoteur : les régions -35 et -10 (par rapport au premier nucléotide transcrit). Les séquences consensus de ces régions (TTGACA et TATAAT respectivement) sont le plus souvent retrouvées mais ne sont pas exclusives (Vvedenskaya et al., 2015). Cette fixation de l'ARNp conduit à la formation d'un complexe dit «fermé» au promoteur, où l'ADN est double brin et n'a pas encore été inséré dans le site catalytique de l'enzyme (Saecker et al., 2011). Ensuite, un déroulement de l'ADN s'opère autour de la région -10 menant à la conformation dite «ouverte» au niveau du promoteur. Dans cette conformation, l'ADN est plié et s'enroule autour de l'holoenzyme (Cellai et al., 2007). Enfin, une torsion de l'ADN est observée, ce qui conduit à un état de stress pour l'ARNp due à l'énergie libérée. Ce stress peut être résolu d'un côté par une transcription abortive (Rosenthal et al., 2008) et un retour de l'ARNp en conformation ouverte au promoteur, ou bien par un relâchement de l'ADN en amont du site transcriptionnel, et un échappement de la région promotrice avec l'incorporation subséquente des nucléotides (Sreenivasan et al., 2016). Après la synthèse de 9-12 nucléotides d'ARN où 8-9 nucléotides sont appariés dans l'hybride ADN-ARN, le complexe passe de la phase d'initiation à la phase d'élongation de la transcription (von Hippel, 1998). Cette transition est caractérisée par l'échappement de l'ARNp de la région promotrice, la dissociation du facteur σ et la formation d'une structure ternaire très stable. Le facteur σ dissocié peut ensuite aller se fixer sur une autre ARNp permettant ainsi l'initiation à de multiples unités de transcription (Mooney et al., 2005).

2.3. Elongation de la transcription

A l'échelle d'un seul nucléotide, l'élongation de la chaîne d'ARN consiste en la formation d'un pont phospho-diester qui s'établit entre le nucléotide positionné dans le site catalytique de l'ARNp et l'hydroxyde en 3' de l'ARN. Dans un premier temps l'ARNp est dans un état dit «pré-transloqué» où l'extrémité 3' de l'ARN occupe la moitié en aval du site catalytique (Vassylyev et al., 2007b). Ensuite l'enzyme passe dans un état

dit «post-transloqué» où l'hydroxyde en 3' de l'ARN passe dans la moitié en amont du site catalytique permettant ainsi la fixation du nucléotide subséquent pour continuer la synthèse d'ARN. Enfin, la catalyse de la réaction permet l'établissement du pont phospho-diester et un relargage du pyrophosphate qui est produit pendant la synthèse d'ARN (Yuzenkova and Zenkin, 2010). L'ARNp en cours de translocation est souvent comparée à un rouage mécanique dont le mouvement peut se produire de façon bilatéral. L'addition de nucléotides substrats provoque l'avancement unilatéral de l'enzyme alors que la mauvaise incorporation de nucléotides, peut conduire à un reculement de l'ARNp (Bar-Nahum et al., 2005). Pendant la phase d'élongation, l'enzyme est très stable sur la matrice d'ADN et peut transcrire sur de grandes distances sans interruption. Cependant, de nombreux signaux comme l'interaction avec des facteurs de transcription ou les pauses de l'ARNp (développé par la suite) vont influencer la progression de l'enzyme.

2.3.1. Les pauses de l'ARN polymérase

2.3.1.1. Définition

L'ARNp peut littéralement effectuer un arrêt temporaire lors de la synthèse de l'ARN, c'est ce que l'on appelle une pause transcriptionnelle.

Une pause est caractérisée par une isomérisation du site actif de l'ARNp dans une conformation inactive, et ce changement de conformation permettrait le passage dans un état que l'on appelle «pause élémentaire» (Figure 2). Cet état peut conduire à différents événements comme la terminaison, l'arrêt ou bien encore l'élongation de la transcription et serait à l'origine de tous les types de pauses (Weixlbaumer et al., 2013). À ce jour, différentes classes de pauses ont été définies en fonction du signal provoquant cet arrêt temporaire de l'ARNp.

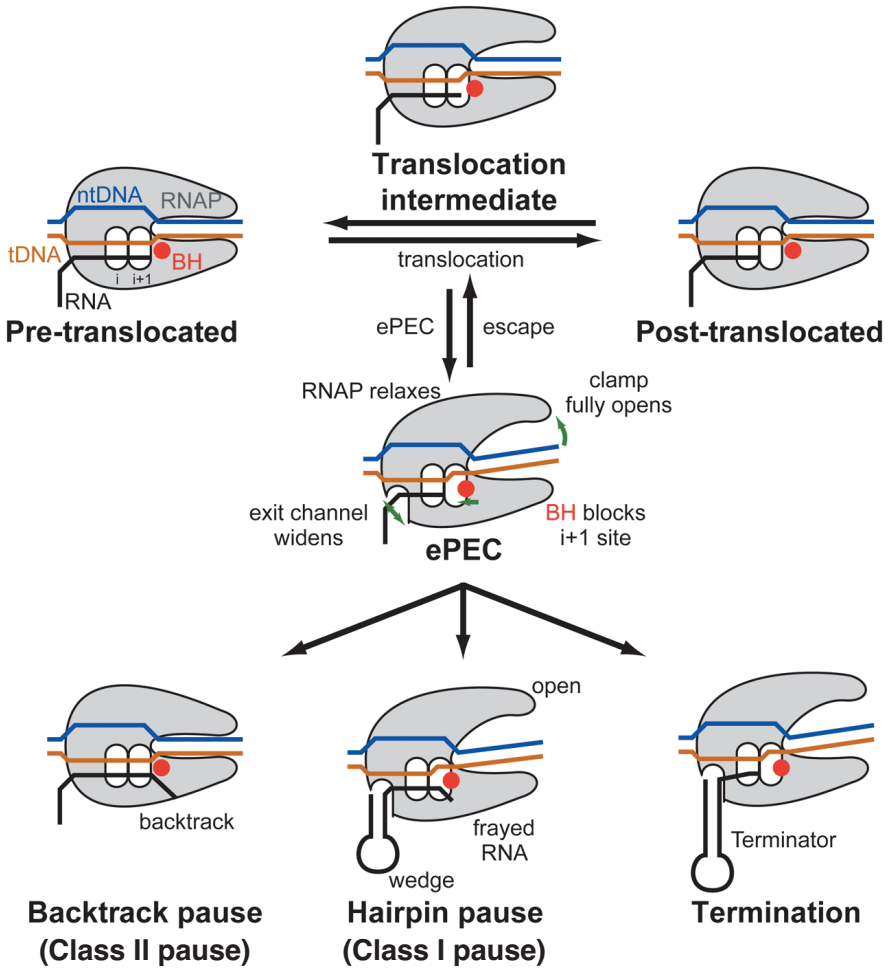


Figure 2 : Modèle de pause transcriptionnelle

L'état de pause élémentaire (ePEC) se caractérise par une ouverture de la pince fixatrice de l'ADN (clamp) et un élargissement du canal de sortie de l'ARN. Cet état de pause élémentaire peut conduire aux pauses de type «backtrack», dépendantes d'une tige boucle ou bien encore à la terminaison de la transcription. Adaptée de (Weixlbaumer et al., 2013).

2.3.1.2. Pause de type tige-boucle

La formation d'une tige-boucle au sein de l'ARN nouvellement synthétisé peut conduire à la pause de l'ARNp et définie les pauses de classe I (Artsimovitch and Landick, 2000).

Pendant une pause de type tige-boucle, un réarrangement allostérique du site catalytique de l'ARNp est observé suite à l'interaction de l'enzyme avec l'ARN (Toulokhonov and Landick, 2003). Ce type de pause a été le premier découvert et les différents éléments nécessaires pour sa formation sont bien caractérisés. Dans ce contexte, la pause de l'ARNp est déterminée par différents facteurs comme la taille de la tige (environ 11 nucléotides d'ARN), l'identité du nucléotide où se situe la pause (préférentiellement un uracile) ou bien encore la distance entre la tige et l'ARNp (Chan and Landick, 1993; Chan et al., 1997; Lee et al., 1990).

Ces pauses ont un rôle capital pour la régulation des opérons contrôlant la synthèse d'acides aminés comme l'histidine (opéron his) (Chan and Landick, 1993), ou bien encore le tryptophane (opéron trp) (Landick and Yanofsky, 1987). Chez *Bacillus subtilis* (*B. subtilis*), une pause au niveau de l'opéron trp facilite le recrutement de la protéine TRAP (Tryptophan Regulated Attenuation Protein) qui va contrôler l'atténuation de la transcription en favorisant par la suite une structure secondaire d'ARN qui empêche la fixation du ribosome (Yakhnin and Babitzke, 2010; Yakhnin et al., 2008)

2.3.1.3. Pause de type «backtrack»

Comme mentionné précédemment, l'ARNp ne progresse pas de façon unilatérale au cours de l'élongation. Le «backtrack» consiste en un mouvement rétrograde de l'ARNp sur la matrice d'ADN, résultant en un désengagement de l'extrémité 3' de l'ARN du site catalytique de l'enzyme (Komissarova and Kashlev, 1997a; Nudler et al., 1997). La pause de type «backtrack» ou pause de classe II, est causée par une instabilité au sein de l'hybride ADN-ARN dans la bulle de transcription (Kireeva and Kashlev, 2009). Cet état de l'hybride est connu pour induire des pauses ou encore d'inactiver le complexe transcriptionnel (arrêt de transcription) (Komissarova and Kashlev, 1997b). La pause de type «backtrack» a souvent pour origine un manque dans l'incorporation du nucléotide subséquent ou une erreur lors de la transcription qui est résolu par le clivage du mauvais transcrit synthétisé (Marr and Roberts, 2000). La nouvelle extrémité 3' formée peut être fixée à nouveau par l'ARNp à la position appropriée pour continuer la synthèse d'ARN.

2.3.1.4. Diversité au sein des pauses transcriptionnelles

Les deux classes de sites de pause décrites précédemment constituent les pauses les plus classiquement retrouvées. Cependant, divers exemples de pauses n'appartenant pas à ces classes ont été identifiés. Par exemple, il a été montré que le facteur σ régule également l'élongation de la transcription en induisant des pauses (Brodolin et al., 2004). Ces pauses s'effectuent grâce à la reconnaissance de régions ressemblant à la région -10 du promoteur dans la région transcrite de l'ADN (Perdue and Roberts, 2010, 2011).

Le ribosome est également impliqué dans la transcription chez les procaryotes et plus précisément dans les sites de pauses. En effet, il a été démontré que celui-ci peut littéralement pousser l'ARNp et empêcher le phénomène de «backtrack» (Kohler et al., 2017; Proshkin et al., 2010).

Récemment, une séquence consensus de pause a été identifiée au niveau de la région du codon d'initiation de la traduction qui est déterminante pour la pause de l'ARNp (Larson et al., 2014). Ce consensus de 16 nucléotides est retrouvé au sein de la majorité des unités de transcription et déterminerait une pause universellement retrouvée au niveau du codon d'initiation de la traduction. Cette pause universelle, a été proposée pour jouer un rôle primordial dans le couplage entre la transcription et la traduction chez les procaryotes en favorisant l'accessibilité de la séquence Shine-Dalgarno (SD) pour une fixation efficace du ribosome.

2.3.2. Principaux facteurs de transcription

Le complexe transcriptionnel ne comporte pas uniquement l'ARNp, l'ARN en cours de synthèse et l'ADN matrice. Des protéines accessoires ou facteurs de transcription peuvent s'associer avec les différentes composantes du complexe et réguler la transcription. Dans cette section nous aborderons les principaux facteurs régulant la transcription chez les procaryotes.

2.3.2.1. NusA et NusG

À l'origine, les facteurs Nus (NusA, B, E et G) ont été identifiés comme faisant partie du système d'anti-terminaison dirigé par la protéine N chez le phage lambda (Friedman and Baron, 1974) et régulent la transcription à différents niveaux.

Le facteur NusA est un facteur d'élongation de la transcription conservé chez les bactéries et les archées (Ingham et al., 1999) et joue des rôles contradictoires lors de la transcription. Son rôle le plus connu est de favoriser les pauses de l'ARNp et donc de ralentir sa progression (Burns et al., 1998). Cette augmentation de l'efficacité de pause de l'ARNp est particulièrement retrouvée dans un contexte de pause de classe I. En effet, NusA stimule l'interaction de la tige-boucle avec l'ARNp ce qui altère la conformation de l'enzyme et favorise la pause (Toulokhonov et al., 2001). NusA participe également à la terminaison de la transcription chez *B. subtilis* (Yakhnin and Babitzke, 2002), inhibe la terminaison dépendante de la protéine Rho (Burns et al., 1998) et est une composante essentielle de l'anti-terminaison au niveau de l'ARNr et des opérons du phage lambda (Shankar et al., 2007). Le domaine N-terminal de NusA interagit avec l'ARNp près du canal de sortie de l'ARN (le domaine C-terminal de la sous-unité α) (Kainz and Gourse, 1998) et pourrait déloger le facteur σ de l'ARNp lors de l'entrée en phase d'élongation (Mooney et al., 2009).

Le facteur NusG quant à lui augmente la vitesse de la transcription en interagissant avec l'ARNp via sa partie N-terminale (Belogurov et al., 2009). NusG fait en sorte que l'ARNp passe moins de temps au niveau des sites de pauses en inhibant le «backtrack» et en favorisant la post-translocation entre les pauses (Herbert et al., 2010). De par son interaction avec d'autres facteurs de transcription via sa partie C-terminale, NusG est impliqué dans deux processus très importants pour l'expression génique. Il est connu pour augmenter l'efficacité de la terminaison dépendante de la protéine Rho (Li et al., 1993) en établissant un lien entre l'ARNp et le facteur de terminaison. Il permet également le couplage entre la transcription et la traduction en reliant l'ARNp et le ribosome via NusE (également appelé S10) qui fait partie de ce dernier (Burmann et al., 2010).

2.3.2.2. RfaH

Le facteur de transcription RfaH est un paralogue de NusG, mais contrairement à ce dernier, n'est pas un facteur général de transcription et n'est pas essentiel pour la bactérie (Bailey et al., 1996). Il influence positivement l'élongation de la transcription en inhibant la terminaison dépendante de la protéine Rho et le temps des pauses de l'ARNp (Belogurov et al., 2009). RfaH régule spécifiquement certains gènes en étant recruté au niveau des sites *ops* ou sites suppresseurs de la polarité de l'opéron (Artsimovitch and Landick, 2000). Les facteurs NusG et RfaH partagent le même site de liaison au niveau de l'ARNp expliquant ainsi les similarités observées lors du contrôle l'élongation de la transcription par ces deux facteurs (Sevostyanova et al., 2008, 2011)

2.3.2.3. GreA et GreB

Les facteurs GreA et GreB favorisent l'activité endo-nucléolytique de l'ARNp lors du mécanisme de «backtrack». GreA et GreB sont respectivement responsables du clivage de petits (deux ou trois nucléotides) et de longs fragments d'ARN exposés par le «backtrack» (Shaevitz et al., 2003). En libérant plus rapidement l'ARNp de sont état pré-transloqué, GreA et GreB favorisent l'élongation de la transcription et sont également responsables du mécanisme de correction de l'enzyme en permettant le retrait des mauvaises incorporations de nucléotides pendant la transcription (Laptenko and Borukhov, 2003).

2.4. Terminaison de la transcription

Malgré la grande stabilité de l'ARNp au cours de l'élongation, des événements de terminaison peuvent se produire et conduire au décrochage de l'enzyme et le relargage de l'ARN. La terminaison marque généralement la fin d'une unité de transcription mais peut aussi se produire prématurément suite à un signal externe pour réguler l'expression génique (Ray-Soni et al., 2016). Par exemple, le découplage entre la traduction et la

transcription peut conduire à l'arrêt de la synthèse du transcrit qui n'est pas traduit par le ribosome (Richardson, 1991).

2.4.1. Terminaison intrinsèque

La terminaison dite intrinsèque est le processus de terminaison le plus largement retrouvé chez les bactéries et s'effectue suite à la formation d'une structure d'ARN, appelée terminateur.

Un terminateur canonique de transcription est caractérisé par une tige-boucle stable suivie d'une séquence poly-uracile d'environ 9 nucléotides de long (Reynolds et al., 1992). Le contexte transcriptionnel au niveau du terminateur conduit à une pause de l'ARNp suivie d'un décrochage de l'enzyme à cause de la proximité de la tige. Actuellement deux modèles mécanistiques sont proposés pour expliquer l'effet de cette tige terminatrice sur le processus de terminaison.

Dans le modèle de translocation, la formation co-transcriptionnelle de la tige tendrait à tirer sur l'extrémité 5' de l'ARN, conduisant à un déplacement de l'ARNp en aval sans transcription subséquente d'ARN (Santangelo and Roberts, 2004).

Dans un autre modèle, dit allostérique, la tige conduirait à un changement de conformation de l'ARNp en envahissant le site actif ce qui déstabiliserait le complexe transcriptionnel (Epshtein et al., 2007)

2.4.2. Terminaison par la protéine Rho

Ce type de terminaison de la transcription implique le facteur Rho, une protéine hexamérique qui possède une activité ATPase dépendante de l'ARN (Galluppi and Richardson, 1980). Rho se fixe sur des séquences spécifiques de l'ARN : les séquences *rut* (pour *Rho-utilization* site) qui sont riches en cytosines et dépourvues de structures secondaires et de ribosomes (Ciampi, 2006). Lorsque que la protéine Rho est fixée sur l'ARNm, celle-ci effectue un mouvement de 5' en 3' sur l'ARN naissant en utilisant

l'énergie de l'hydrolyse de l'ATP pour terminer la transcription (Richardson, 2003).
Il a été démontré que la terminaison dépendante de la protéine Rho se déroulait au niveau des sites de pause de l'ARNp (Lau et al., 1983; Morgan et al., 1983). En effet dans un contexte transcriptionnel, le ralentissement de l'ARNp permet à la protéine de rattraper l'enzyme et de la déloger de la matrice d'ADN.
Aujourd'hui le débat reste très actif concernant la mécanistique de terminaison par cette protéine avec trois modèles proposés (Figure 3).

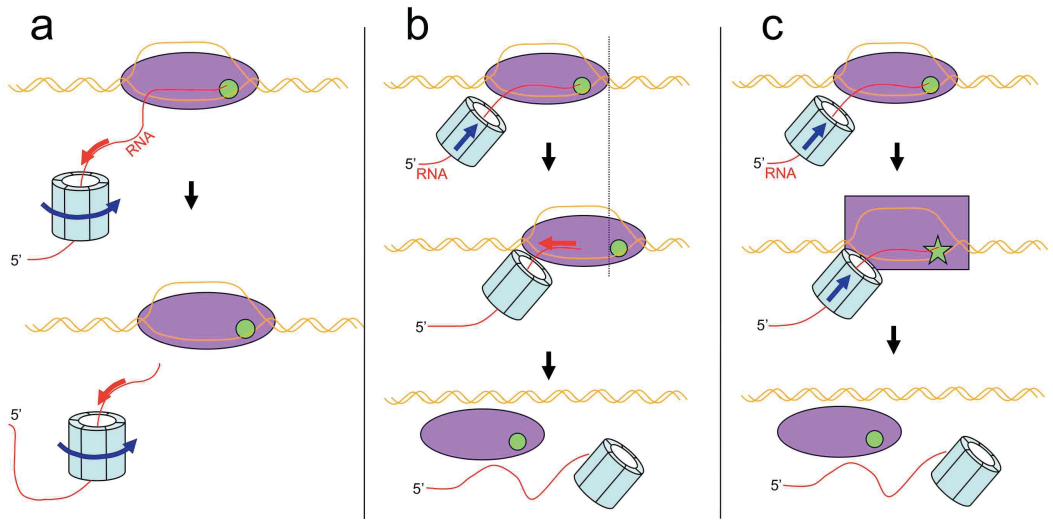


Figure 3 : Modèles proposés de terminaison dépendante de la protéine Rho.
a) Modèle de la «bobine» où Rho tire sur l'ARN pour rompre l'hybride ADN-ARN. b) Modèle d'hyper-translocation où Rho pousse l'ARNp en avant pour provoquer la terminaison. c) Modèle allostérique où Rho provoque un changement de conformation de l'ARNp. Adaptée de (Epshtein et al., 2010).

Plus récemment des études à l'échelle du génome ont établis que la protéine Rho peut interagir avec l'ARNp dès l'initiation de la transcription (Epshtein et al., 2010; Mooney et al., 2009). Ces études suggèrent que le facteur serait toujours associé à l'ARNp et attendrait un signal comme l'accessibilité d'une séquence *rut* pour terminer efficacement la transcription. La protéine Rho est essentielle chez *E. coli* car à l'échelle du génome elle

est responsable de la terminaison d'environ 70 % des transcrits (Peters et al., 2012). En effet, Rho termine la transcription au niveau de différents types de transcrits comme les petits ARN ou les ARNt et même à la fin d'un gène d'un opéron afin de prévenir la transcription de l'opéron au complet (Peters et al., 2012; Sedlyarova et al., 2016)

3. Les riborégulateurs

3.1. Généralités

Les riborégulateurs sont des structures d'ARN majoritairement retrouvées dans la région 5' non traduite (5'-UTR) des ARNm bactériens. Ces ARN sont capables de lier spécifiquement un métabolite particulier appelé ligand, entraînant un changement de conformation de l'ARN pour réguler l'expression génique. Un riborégulateur est définit par deux domaines distincts : l'aptamère et la plateforme d'expression (Figure 4).

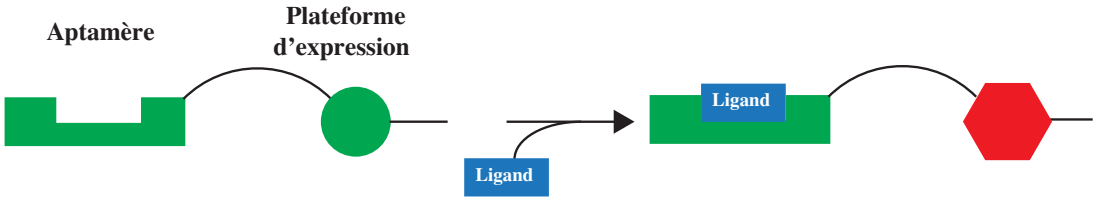


Figure 4 : Architecture simplifiée d'un riborégulateur.

L'aptamère est le domaine de liaison spécifique du ligand et la plateforme d'expression est la région responsable de la régulation génique. En effet, la liaison du ligand au niveau de l'aptamère entraîne un changement de conformation de la plateforme d'expression modulant ainsi l'expression du gène. En général la liaison du ligand conduit à une répression de l'expression génique avec certaines exceptions où la forme liée du riborégulateur est permissive à l'activation du gène (Mandal and Breaker, 2004).

3.2. Familles

Contrairement aux ARN artificiels développés en laboratoire, la présence d'une interaction ARN-métabolite n'était pas prédite pour exister et réguler l'expression génique *in vivo* (Lauhon & Szostak 1995). Par la suite, on commençait à considérer des structures d'ARN comme participant directement à la régulation de l'expression génique mais aucun facteur protéique expliquant ce mécanisme n'avait été identifié (Miranda-Rios *et al*. 2001). Face à cette absence de protéine responsable de la régulation, certains laboratoires ont ainsi émis l'hypothèse que la présence d'un métabolite cellulaire participait directement à l'inactivation de certains gènes.

Depuis la découverte et la caractérisation des premiers riborégulateurs liant la flavine mononucléotide (FMN) et la thiamine pyrophosphate (TPP) (Mironov et al., 2002; Winkler et al., 2002), une grande variété de tels éléments régulateurs ont été identifiés et sont habituellement nommés et classés en fonction du ligand reconnu. Aujourd'hui, les riborégulateurs sont connus pour lier une multitude de ligands naturels tels que le co-enzyme B12 (Nahvi et al., 2002), la lysine (Sudarsan et al., 2003a), la guanine (Mandal et al., 2003), la S-adénosyl-méthionine (SAM) (Epshtein et al., 2003a; Winkler et al., 2003), la glycine (Mandal et al., 2004), l'adénine (Mandal and Breaker, 2004), la glucosamine-6-phosphate (GlcN6P) (Winkler et al., 2004), le magnésium (Groisman et al., 2006), la 7-aminométhyl-7-déazaguanine (préQ1) (Roth et al., 2007), la 2'-désoxy-guanosine (Kim et al., 2007), la di-guanosine monophosphate cyclique (di-GMPc) (Sudarsan et al., 2008), la S-adénosyl-monocystéine (SAH) (Wang et al., 2008), le molybdène (Moco) (Regulski et al., 2008), le tétrahydrofolate (Ames et al., 2010), la glutamine (Ames and Breaker, 2011), le fluor (Baker et al., 2012), la di-adénosine monophosphate cyclique (di-AMPc) (Nelson et al., 2013), des ARNt (Zhang and Ferré-D'Amaré, 2015), le manganèse (Dambach et al., 2015; Price et al., 2015), le nickel et le cobalt (Furukawa et al., 2015), la guanidine (Nelson et al., 2017).

Il existe également des ARN régulateurs capables de changer de conformation suite à l'action d'un facteur externe comme le pH (Nechooshtan et al., 2014), cependant le débat reste présent concernant leur classification en riborégulateur puisqu'ils ne font pas intervenir la liaison d'un métabolite.

Chacune de ces familles de riborégulateurs se distingue par leur aptamère qui est généralement conservé au sein d'une même famille (Serganov, 2009). L'architecture de ces aptamères va de la simple tige-boucle pouvant lier le préQ1 (Roth et al., 2007), à une structure beaucoup plus complexe comme les jonctions multiples de l'aptamère liant le co-enzyme B12 (Johnson et al., 2012). Chaque riborégulateur lie son ligand respectif avec spécificité, c'est-à-dire qu'il est capable de discriminer son ligand d'une molécule analogue. Par exemple, le riborégulateur TPP est aussi capable de lier la thiamine, mais avec une affinité mille fois moindre (Winkler et al., 2002).

3.3. Les riborégulateurs TPP

Comme mentionné précédemment, les riborégulateurs sont capables de lier une multitude de ligands de natures différentes. Durant mes travaux, j'ai utilisé comme modèle d'étude les riborégulateurs liant le TPP chez la bactérie *E. coli*. Le TPP (Figure 5) est la forme active de la thiamine (vitamine B1) et est la plus retrouvée dans la cellule.

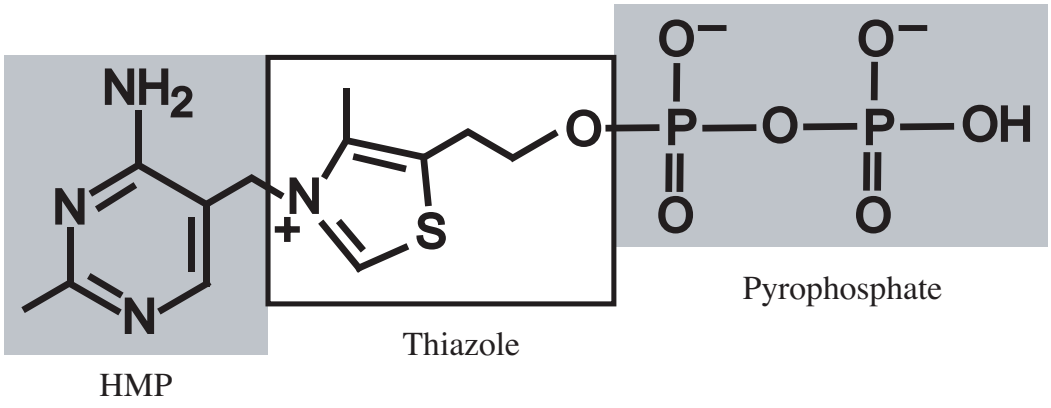


Figure 5 : Structure chimique du TPP. Adaptée de (Serganov et al., 2006)

La thiamine est synthétisée chez les plantes, les champignons et les bactéries mais pas chez les vertébrés, qui la prélèvent dans leur alimentation. Le TPP est essentiel pour la bactérie car il est impliqué dans de nombreux processus biologiques en temps que cofacteur protéique dans la production énergétique de la cellule (Ciszak et al., 2003).

3.3.1. Diversité

Les riborégulateurs liant le TPP sont les seuls exemples de riborégulateurs retrouvés dans le domaine eucaryote puisqu'ils ont été identifiés chez l'algue verte, les champignons et les plantes (Sudarsan et al., 2003b). Chez *E.* coli, trois riborégulateurs TPP sont présents et contrôlent l'expression des gènes de synthèse ou de transport de la thiamine (Rodionov et al., 2002). Le riborégulateur situé en amont de l'opéron *tbpAPQ* régule la synthèse d'un transporteur de la thiamine. Il a été observé que l'aptamère de ce riborégulateur est également un petit ARN (*sroA*) qui peut réguler l'expression de gènes en *trans* et conduire à la résistance aux antibiotiques chez *Salmonella typhimurium* (Vogel et al., 2003; Yu and Schneiders, 2012). Le riborégulateur *thiM* est situé en amont de l'opéron *thiMD*. Cet opéron est impliqué dans la glycolyse et le métabolisme des purines avec d'un côté ThiM qui est une thiazole kinase et ThiD une phospho-méthyl-pyrimidine kinase (HMP-P kinase) (Petersen and Downs, 1997). Le riborégulateur *thiC*, contrôle quant à lui l'expression de l'opéron *thiCEFSGH*, impliqué dans le métabolisme des purines, des stéroïdes, de la cystéine et directement dans la synthèse de la thiamine monophosphate, un intermédiaire du TPP (Vander Horn et al., 1993).

3.3.2. Structure de l'aptamère

L'aptamère des riborégulateurs TPP (ou thi-box) correspond à 39 nucléotides conservés au niveau de leur séquence et de la structure globale de l'ARN (Miranda-Ríos et al., 2001; Rodionov et al., 2002). Aujourd'hui les cristaux de différents aptamères des riborégulateurs TPP ont été résolus chez *E. coli*, (riborégulateur *thiM*) (Serganov et al., 2006) et chez les plantes (riborégulateur *thiC*) (Thore et al., 2006) et nous permettent de mieux comprendre le mode de liaison du ligand au riborégulateur (Figure 6).

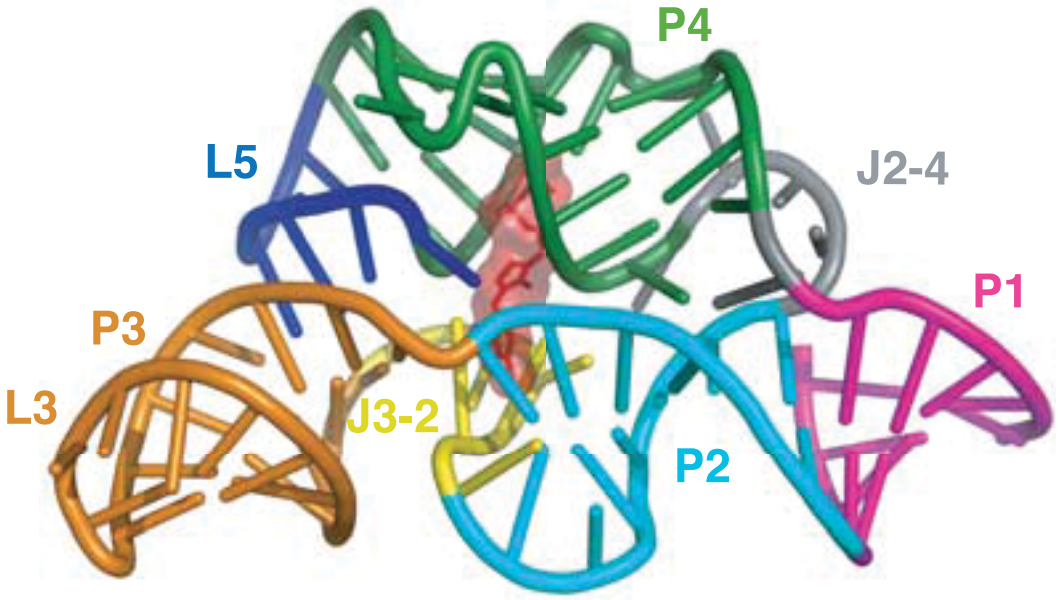


Figure 6 : Structure de l'aptamère du riborégulateur *thiM* lié au TPP

Le cristal de l'aptamère du riborégulateur *thiM* d'*E. coli* est présenté avec le TPP placé au centre de celui-ci (en rouge). J : jonction, P : tige. Adaptée de (Serganov et al., 2006).

Les groupements Hydroxy-méthyl pyrimidine (HMP) et pyrophosphate du TPP sont directement responsables de sa liaison avec l'aptamère. Cette interaction s'effectue au niveau de la jonction entre la tige trois et la tige deux (J2-3) et à l'interface entre les tiges cinq et quatre. Une interaction entre la boucle de la tige cinq (L5) et la tige trois (P3) est également visible sur le cristal et a été montrée pour être déterminante pour la liaison du ligand (interaction L5-P3) (Kulshina et al., 2010). En effet, la formation de cette «poche de liaison du TPP» permet à l'ARN d'adopter une conformation qui capture le TPP une fois que celui-ci est lié à l'aptamère (Haller et al., 2013).

3.4. Modes de régulation

Chez les procaryotes, la production d'ARNm et l'initiation de la traduction sont les processus les plus ciblés par les riborégulateurs (Figure 7). Cependant, les plateformes d'expression étant peu conservées, il existe quelques cas où le changement de structure de l'ARN conduit à d'autres processus de régulation génique décrits dans cette section.

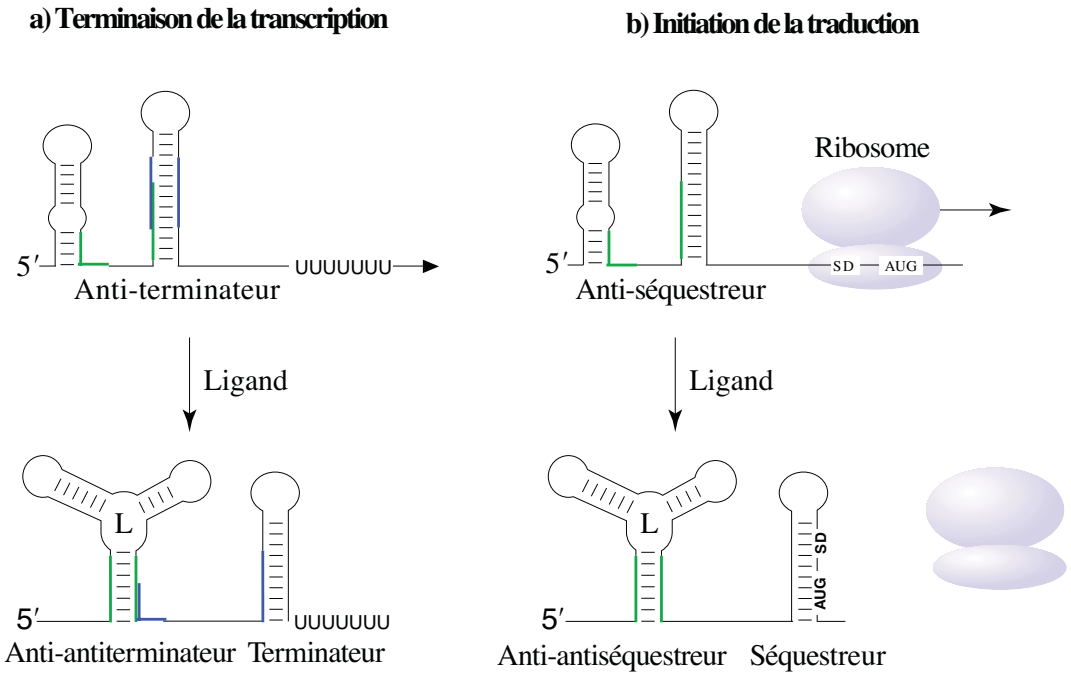


Figure 7 : Principaux mécanismes de régulation employés par les riborégulateurs.
Les mécanismes de terminaison de la transcription (a) et d'inhibition de la traduction (b) sont présentés dans le cadre d'une répression de l'expression génique. Les structures terminateur et anti-terminateur ainsi que séquestreur et anti-séquestreur sont mutuellement exclusives. SD : séquence Shine-Dalgarno, L : ligand, AUG : codon d'initiation de la traduction. Adaptée de (Nudler and Mironov, 2004).

3.4.1. Terminaison intrinsèque de la transcription

La formation de terminateurs intrinsèques est l'un des moyens les plus répandus des plateformes d'expression pour réguler l'expression génique (Grundy and Henkin, 2003). En effet, la liaison du ligand induit la formation d'un terminateur de transcription canonique qui déstabilise l'ARNp avant la synthèse de la région codante du gène. A l'inverse, lorsque le ligand n'est pas lié au riborégulateur celui-ci est dit dans une forme «permissive». Dans cette conformation, la présence d'une structure anti-terminatrice entre en compétition avec la formation de la tige terminatrice favorisant l'élongation de la transcription. Ce mode de régulation ne nécessite pas la présence d'autres facteurs même si la cinétique de transcription est un paramètre déterminant pour la formation de la tige terminatrice ou anti-terminatrice (Lemay et al., 2011; Wickiser et al., 2005).

3.4.2. Initiation de la traduction

La liaison du ligand à un riborégulateur peut également moduler l'initiation de la traduction du gène situé en aval. D'un point de vue structural, le mécanisme est très similaire à celui de la terminaison intrinsèque de la transcription. En effet, la liaison du ligand conduit à la formation d'une tige dans la plateforme d'expression qui séquestre la séquence SD ou RBS (pour *Ribosome binding site*) et/ou le codon d'initiation de la traduction. Lors de ce processus, le gène au complet peut être transcrit mais la production en protéine est inhibée. Bien que cette règle ne soit pas exclusive, il semble il y avoir une prévalence pour une régulation traductionnelle chez les bactéries Gram négatives alors que les bactéries Gram positives utiliseraient plutôt la terminaison de la transcription (Grundy and Henkin, 2003). Cette observation pourrait s'expliquer par le fait que le génome des Gram positives est plus riche en polycistrons et elles auraient un avantage notoire à induire une terminaison de la transcription (Nudler and Mironov, 2004).

3.4.3. Terminaison par la protéine Rho

Dans certains cas, la terminaison de la transcription peut impliquer la présence de la protéine Rho qui va terminer la transcription lorsque le ligand est lié au riborégulateur (Hollands et al., 2012). Ce mécanisme implique que la protéine Rho peut se lier sur l'ARN seulement en présence du ligand, c'est-à-dire que le réarrangement structural rend accessible le site *rut*. Il a été montré que le riborégulateur magnésium chez *Salmonella enterica* (*S.* enterica) et le riborégulateur FMN chez *E. coli* sont tous deux régulés par la protéine Rho en présence de leur ligand (Hollands et al., 2012). Pour le riborégulateur magnésium, un site de pause dont la prépondérance est dépendante de la présence du ligand est le lieu où la terminaison s'effectue (Hollands et al., 2014).
Ce mécanisme peut être qualifié de direct puisque le changement de conformation permet à lui seul de recruter la protéine Rho. Dans un mécanisme dit de polarité, la liaison du ligand a un effet indirect sur la terminaison. En effet, dans le cas où l'initiation de la traduction est inhibée, la protéine Rho peut se lier sur l'ARNm car il n'est pas protégé par la présence du ribosome en cours de traduction (Bastet *et al,* 2017). Ainsi, il est possible que la transcription des riborégulateurs traductionnels soit également terminée par la protéine Rho pour empêcher la synthèse d'un ARNm qui ne sera pas traduit en protéine.

3.4.4. Stabilité de l'ARNm

Dans une cellule, la stabilité de l'ARNm et donc sa durée de vie est déterminée par de nombreux facteurs comme par exemple l'accessibilité aux RNase. Cette régulation de la quantité d'ARNm est particulièrement accentuée dans le cas où celui-ci n'est pas protégé par le ribosome en cours de traduction (Deana and Belasco, 2005). Récemment, il a été démontré que les riborégulateurs peuvent directement influer sur la stabilité de l'ARNm en modulant le recrutement de la machinerie de dégradation : le dégradosome chez *E. coli*. Dans le cas du riborégulateur lysine, la liaison du ligand va conduire à l'accessibilité de deux sites de coupure par la RNase E dans la plateforme d'expression et provoquer la dégradation de l'ARNm (Caron et al., 2012). Dans cet exemple, le recrutement du dégradosome est indépendant de la traduction du gène situé en aval, établissant ainsi un

autre mécanisme direct impliquant le changement de conformation du riborégulateur.

3.4.5. Epissage alternatif

Comme mentionné précédemment, les riborégulateurs liant le TPP sont les seuls riborégulateurs identifiés à ce jour dans le domaine eucaryote. Ces riborégulateurs vont réguler un processus déterminant pour l'expression génique : l'épissage alternatif. Deux exemples ont été caractérisés à ce jour impliquant directement le riborégulateur et la liaison du ligand. Dans un premier cas, le riborégulateur module l'accessibilité à deux sites d'épissage où un seul de ces sites est accessible en fonction de la présence du ligand (Cheah et al., 2007). Dans un second exemple, le riborégulateur est localisé dans le 3'-UTR du gène et la présence du ligand résulte en la formation d'une extrémité 3' plus longue qui favorise sa dégradation (Wachter et al., 2007).

3.4.6. Combinaison de plusieurs mécanismes

Les différents mécanismes décrits dans cette section montrent la capacité des riborégulateurs à réguler directement l'expression génique grâce à des mécanismes relativement simples. Aussi, il est important de noter que ces différents modes de régulation ne semblent pas exclusifs. Un exemple qui illustre ce phénomène est le cas du riborégulateur FMN chez *Corynebacterium glutamicum,* où le facteur Rho et une ribonucléase régulent tous deux la quantité d'ARNm dépendamment de la présence du ligand (Takemoto et al., 2014). Ces deux mécanismes ne sont pas dépendant l'un de l'autre suggérant ainsi que les riborégulateurs sont capables d'utiliser différents mécanismes combinés pour une régulation plus fine de l'expression génique.

3.5. Dynamique de liaison du ligand

L'étude des riborégulateurs nécessite de connaître leur structure ainsi que les différentes étapes de leur repliement menant à l'obtention de la structure finale. Dans ce contexte, la

fenêtre de liaison du ligand, c'est-à-dire la période pendant laquelle celui-ci peut se lier à l'aptamère, détermine le patron de repliement du riborégulateur.

Lors d'un contrôle thermodynamique, la liaison du ligand stabilise une structure du riborégulateur par rapport à d'autres qui sont possibles. D'un autre côté, le contrôle peut s'effectuer de manière cinétique où la vitesse de repliement du riborégulateur est directement en compétition avec le temps requis pour que l'aptamère et son ligand atteignent un équilibre. Dans un cadre cinétique, il faut que la concentration locale du ligand excède sa constante de dissociation (Kd) pour que le riborégulateur soit en conformation liée au ligand (Coppins et al., 2007). Dans ce contexte, où la vitesse de la transcription joue un rôle déterminant, il est possible de citer également des facteurs extérieurs à la transcription, comme la participation des ions pour le repliement de l'ARN. Un exemple concret est illustré par l'influence du magnésium sur le repliement des riborégulateurs au TPP. En effet, le magnésium favorise l'interaction L5-P3 pour une liaison efficace du ligand et fait donc partie intégrante du mode de repliement de l'aptamère (Kulshina et al., 2010).

4. Etude du repliement de l'ARN

Lorsque l'on parle d'ARN particuliers comme les riborégulateurs, c'est avant tout la structure qui dirige leur fonction de régulation. La formation alternative de structures d'ARN, soit le repliement de l'ARN, va ainsi déterminer la structure finale obtenue et donc la capacité pour le riborégulateur de réguler l'expression génique. En effet, au cours de sa synthèse, l'ARN se replie et forme des structures qui peuvent aller de la simple paire de base à une structure tertiaire plus complexe comme un aptamère. Ce repliement dit co-transcriptionnel, s'effectue dès qu'une longueur suffisante d'ARN est synthétisée et aura une importance notoire pour l'adoption de la structure finale de la chaîne nucléotidique ainsi que pour la régulation de certains processus qui se produisent pendant la transcription (décrits dans cette section).

4.1. Fonctionnalité du repliement co-transcriptionnel

Le repliement co-transcriptionnel de l'ARN fait aujourd'hui l'objet de nombreuses études et les travaux effectués sur la RNase P seront cités à titre d'exemple. La RNase P est un complexe nucléoprotéique qui génère les extrémités 5' matures de tous les ARNt grâce à une réaction endo-nucléolytique spécifique à certains sites (Pace and Brown, 1995). L'ARN constitue la sous-unité catalytique de cette enzyme avec deux domaines : le domaine de spécificité et le domaine catalytique. Le processus de la transcription influence directement le repliement de la RNase P (Pan et al., 1999), mais également la réponse au facteur de transcription NusA qui conduit à un repliement plus efficace du domaine de spécificité. NusA permet ainsi d'établir une pause plus longue qui conduit à un repliement fonctionnel de l'ARN (Wong et al., 2005).
Cet exemple démontre l'importance du contexte transcriptionnel pour étudier le repliement d'un ARN mais également son activité régulatrice. En effet, la structure adoptée par un ARN particulier peut déterminer le résultat final de l'expression génique en favorisant la terminaison de la transcription par exemple.

4.2. La transcription en conditions natives

Pour étudier la structure de l'ARN ainsi que les différentes étapes de son repliement, il est nécessaire de se placer dans des conditions minimales et contrôlées. La transcription *in vitro* par l'ARNp bactérienne est un processus connu et maîtrisé depuis de nombreuses années (Levin et al., 1987), et a grandement participé à mieux comprendre les mécanismes régulant la production d'ARN comme les pauses de l'ARNp (Zhang and Landick, 2016), la contribution des facteurs de transcription (Burns et al., 1998; Herbert et al., 2010; Sevostyanova et al., 2011; Yakhnin et al., 2008) ou bien encore la terminaison de la transcription (Epshtein et al., 2007). Des variantes de la transcription *in vitro* ont été depuis développées afin de cibler certains processus ou de changer les conditions expérimentales en fonction de la question posée. Il est ainsi possible de citer la transcription à cycle unique (ou *single round* en anglais), où l'ARNp ne peut transcrire la matrice d'ADN qu'une seule fois grâce à l'ajout d'une protéine comme l'héparine qui

empêche la ré-initiation de la transcription (Walter et al., 1967). Ce type d'expérience est particulièrement utilisé pour étudier la terminaison de la transcription, notamment par les riborégulateurs, afin de quantifier l'efficacité de terminaison. En effet, la méthode du cycle unique a l'avantage de permettre le contrôle au sein de la quantité d'ARN produit et d'étudier directement l'impact des conditions expérimentales sur le processus de transcription.

Enfin, l'utilisation de la transcription «en phase solide» permet de pouvoir changer les conditions d'analyse comme le milieu réactionnel ou la position de l'ARNp sur la matrice d'ADN (Nudler et al., 2003). Cette méthode initialement développée par le groupe du Dr. Goldfarb repose sur la liaison de l'une des composantes du complexe transcriptionnel, à une bille d'affinité (Kashlev et al., 1996). Le complexe transcriptionnel étant solidement attaché à la surface, il est possible de modifier le contexte d'analyse en changeant la composition du milieu ou en ajoutant un facteur protéique à un moment donné de la transcription par exemple.

4.3. Barrage transcriptionnel

Au cours de sa transcription, l'ARN peut adopter une ou plusieurs conformations différentes dépendamment du contexte transcriptionnel et les structures d'ARN peuvent se former à l'échelle de la milliseconde (Al-Hashimi and Walter, 2008; Bevilacqua et al., 2016). Ces structures d'ARN sont transitoires et nécessitent de pouvoir isoler un complexe transcriptionnel à un moment donné. Le principe du barrage transcriptionnel permet de bloquer la progression de l'ARNp sans que celle-ci ne décroche de la matrice d'ADN (Nudler et al., 2003). Grâce à une protéine barrière comme par exemple une version mutée de l'enzyme EcoR1 (EcoR1Q111) (Pavco and Steege, 1990), l'ARNp rencontre littéralement un obstacle arrêtant la transcription jusqu'à ce que l'obstacle soit retiré. Cette approche a été utilisée à plusieurs reprises pour étudier différents processus transcriptionnels comme la terminaison par la protéine Rho (Shankar et al., 2007), la coopérativité des ARNp lors de la transcription (Epshtein et al., 2003b) et plus récemment la structure d'un riborégulateur au cours de sa transcription (Watters et al., 2016).

5. Objectifs du projet de recherche

Au début de mon projet de recherche, deux publications faisaient état d'une analyse d'un riborégulateur dans un contexte transcriptionnel (Perdrizet et al., 2012; Wickiser et al., 2005). Dans la première conduite par le groupe du Dr. Breaker, il a été montré que la cinétique de transcription était déterminante pour qu'un riborégulateur transcriptionnel atteigne son équilibre thermodynamique. Dans cet exemple, la fenêtre de liaison du ligand est délimitée par la position de l'ARNp au cours de la transcription. Lorsque la structure anti-terminatrice est formée, il n'est plus possible pour le riborégulateur de lier son ligand et de conduire à une répression de l'expression génique.

Par la suite, l'équipe du Dr. Pan s'est intéressée à l'impact des pauses transcriptionnelles sur le repliement d'un riborégulateur traductionnel. Leur étude démontre que le patron de repliement de la plateforme d'expression est déterminé par les pauses de l'ARNp. En absence des pauses identifiées, le riborégulateur n'est plus capable de moduler l'expression génique, suggérant que le repliement de l'ARN est coordonné par le processus de transcription pour obtenir la structure finale fonctionnelle du riborégulateur.

Ces travaux ont constitué la base de mon projet de recherche qui était de mieux comprendre l'impact de la transcription sur le repliement des riborégulateurs. Nous avions pour hypothèses que les pauses de l'ARNp coordonnaient effectivement le repliement des riborégulateurs en général et que la formation de structures transitoires détermine ce processus. Notre objectif était d'identifier ces structures transitoires qui mènent à l'obtention d'un riborégulateur fonctionnel et de caractériser leur implication pour la régulation génique. Pour atteindre cet objectif général nous avons développé de nouvelles méthodes d'analyse de l'ARN qui reposent sur l'obtention d'un complexe transcriptionnel arrêté aux pauses naturelles de l'ARNp.

Nous avons identifié une structure particulière appelée «Anti-P1» qui stabilise le riborégulateur dans sa conformation non liée au ligand. Cette structure se forme pendant la transcription et est favorisée au niveau d'un site de pause de l'ARNp. Ce point de contrôle transcriptionnel délimite ainsi une fenêtre de liaison du ligand. Ces travaux ont été publiés dans la revue Nature Communications en Janvier 2017 et constituent le premier chapitre de ma thèse (Chauvier et al., 2017).

Dans un deuxième chapitre notre objectif était de pouvoir analyser la structure de l'ARN en complexe transcriptionnel à l'échelle de la molécule unique. Pour répondre à cet objectif, nous avons interrogé la dynamique de structure de l'ARN en complexe transcriptionnel en utilisant le principe du FRET (pour *Förster Resonance Energy Transfer* en anglais). Le FRET repose sur l'échange d'énergie qui se produit entre deux fluorophores : un donneur et un accepteur (Roy et al., 2008). Cet échange d'énergie est dépendant de la distance entre les deux fluorophores permettant ainsi d'observer les changements structuraux de l'ARN en microscopie. Ces travaux sont présentés sous forme d'article et sont en processus de soumission dans un article scientifique.

CHAPITRE 1

DÉCRYPTAGE DES MÉCANISMES MOLÉCULAIRES DU RIBORÉGULATEUR *THIC* :

1.1 Présentation de l'article et contribution

Depuis sa découverte en 2002 par le groupe du Dr. Breaker, le riborégulateur *thiC* présent chez *E. coli* a été peu caractérisé. Ce riborégulateur TPP a été prédit pour réguler l'expression génique par le mécanisme d'inhibition de la traduction (Rodionov et al., 2002) mais aucune preuve directe n'était disponible au début de mon projet.

Dans cette première publication, nous avons démontré que le riborégulateur *thiC* utilise deux mécanismes distincts pour réguler l'expression du gène situé en aval. La liaison du ligand au riborégulateur conduit à une terminaison de la transcription par la protéine Rho d'une part, et à une inhibition de l'initiation de la traduction. La cinétique de transcription définit une fenêtre de liaison du ligand qui est délimitée par la présence d'une pause de l'ARNp dans la région d'initiation de la traduction. Cette pause agit à titre de point de contrôle pour la régulation du riborégulateur. En effet, dans un contexte où le ligand s'est déjà lié, elle favorise la terminaison dépendante de la protéine Rho. Cependant, en absence du TPP, le complexe pausé adopte une conformation réfractaire à la liaison du ligand grâce à la formation d'une structure particulière : l'Anti-P1. Considérant que les pauses de l'ARNp sont particulièrement présentes dans la région d'initiation de la traduction (Larson et al., 2014), cet article soulève l'hypothèse que le mécanisme d'Anti-P1 serait universellement retrouvé pour la régulation génique des riborégulateurs chez *E. coli*. En effet, nous avons identifié une pause proche de la région d'initiation de la traduction pour tous les riborégulateurs testés qui pourrait également favoriser l'adoption d'une conformation réfractaire à la liaison du ligand.

Frédéric Picard-Jean a effectué et analysé les expériences de SHAPE. Jean-Christophe Berger-Dancause a réalisé et analysé les expériences de traduction *in vitro*. Laurène Bastet a déterminé l'implication de la protéine Rho *in vivo* par essai de gène rapporteur. Mohammad Reza Naghdi a effectué l'analyse bio-informatique. Audrey Dubé a réalisé les expériences de buvardage de type Northern et l'extension d'amorce. Pierre Turcotte a identifié et quantifié les pauses du riborégulateur *lysC*. J'ai réalisé et analysé toutes les autres expériences en plus de rédiger le manuscrit et de faire les figures avec le Pr. Daniel Lafontaine, Frédéric Picard-Jean et le Pr. Jonathan Perreault.

1.2 Manuscrit

Adrien Chauvier, Frédéric Picard-Jean, Jean-Christophe Berger-Dancause, Laurène Bastet, Mohammad Reza Naghdi, Audrey Dubé, Pierre Turcotte, Jonathan Perreault & Daniel A. Lafontaine (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. Nature Communications.

Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation

Adrien Chauvier[1], Frédéric Picard-Jean[1], Jean-Christophe Berger-Dancause[1], Laurène Bastet[1], Mohammad Reza Naghdi[2], Audrey Dubé[1], Pierre Turcotte[1], Jonathan Perreault[2] and Daniel A. Lafontaine[1]*

[1]Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Quebec, Canada, J1K 2R1.
[2]INRS-Institut Armand-Frappier, Laval, Quebec, Canada, H7V 1B7.

*To whom correspondence should be addressed.

Address correspondence to:
Daniel Lafontaine,
2500, Blvd Université, Sherbrooke, Quebec, Canada, J1K 2R1.
E-mail: Daniel.Lafontaine@USherbrooke.ca

ABSTRACT

On the basis of nascent transcript sequencing, it has been postulated but never demonstrated that transcriptional pausing at translation start sites is important for gene regulation. Here, we show that the *Escherichia coli* thiamin pyrophosphate (TPP) *thiC* riboswitch contains a regulatory pause site in the translation initiation region that acts as a checkpoint for *thiC* expression. By biochemically probing nascent transcription complexes halted at defined positions, we find a narrow transcriptional window for metabolite binding in which the downstream boundary is delimited by the checkpoint. We show that transcription complexes at the regulatory pause site favor the formation of a riboswitch intramolecular lock that strongly prevents TPP binding. In contrast, cotranscriptional metabolite binding increases RNA polymerase pausing and induces Rho-dependent transcription termination at the checkpoint. Early transcriptional pausing may provide a general mechanism whereby transient transcriptional windows directly coordinate the sensing of environmental cues and bacterial mRNA regulation.

Introduction

Gene expression is a highly regulated process that is vital for cellular homeostasis. Numerous cellular factors such as RNA molecules regulate gene expression at the transcriptional level and contribute to an ever expanding repertoire of mechanisms governing cellular processes[1]. Riboswitches are non-coding RNA elements that directly monitor the level of cellular metabolites and control gene expression by modulating either mRNA levels or the efficiency of translation initiation[2]. These RNA switches consist of two modular domains: an aptamer and an expression platform. While the aptamer is implicated in specific metabolite recognition, the expression platform regulates gene expression by controlling the formation of structures that modulate transcription terminators, ribosome binding sites (RBS), RNase E cleavage sites or mRNA splicing[2]. Recently identified RNA motifs with unconventional expression platforms[3,4] suggests the existence of novel riboswitch regulation mechanisms not yet discovered.

Although only a few riboswitch regulation mechanisms have been characterized *in vivo*, growing evidence indicates that multiple biological processes can be controlled by riboswitch conformational changes. For example, it was recently demonstrated that the lysine-dependent *lysC* riboswitch controls both translation initiation and RNase E cleavage activity upon ligand binding[5]. Furthermore, in addition to modulating translation initiation, the *ribB* and *mgtA* riboswitches were shown to regulate transcription termination by using the transcription factor Rho[6]. The *mgtA* riboswitch was also found to contain a critical RNA polymerase pause site important for Rho transcription termination[7]. These results suggest that riboswitch sequences embed all the information required for metabolite sensing and genetic regulation. However, the lack of clear consensus sequences for RNase E or Rho-dependent regulation makes it difficult to predict whether these control mechanisms are widely used by riboswitches[8,9]. Nevertheless, the large substrate recognition spectrum of such RNA binding proteins indicates that they may be involved as key players in the regulation of several riboswitch-controlled genes.

Sequencing of nascent elongating transcripts (NET-seq) recently revealed a 16-nt consensus pause sequence that is enriched at translation start sites in both *Escherichia coli* and *Bacillus subtilis*[10], suggesting that transcriptional pausing may be important for the control of translation initiation. Here, we investigate the *in vivo* regulation mechanism of the thiamin pyrophosphate (TPP)-sensing *thiC* riboswitch in *E. coli*. We identify a transcriptional pause site occurring within the vicinity of the translation start codon. Using *in vitro* assays, we find that pause efficiency increases upon TPP binding and is required to trigger Rho-dependent transcription termination. Mutational analysis shows transcriptional pausing to be crucial for mRNA control *in vivo*. Furthermore, we perform biochemical probing on transcription elongation complexes to directly monitor TPP binding on nascent riboswitch transcripts. We find that while TPP binding is efficient for transcription complexes located upstream of the pause site, it is severely reduced for complexes stalled at the pause site. Our results suggest that the nascent RNA forms an intramolecular lock that both prevents TPP binding to the aptamer and exposes the RBS/AUG start codon, promoting translation initiation. The pause site acts as an early transcriptional checkpoint where a decision is made to either repress or allow transcription elongation and translation initiation of the *thiC* operon. Here we describe a riboswitch-based molecular mechanism in which transcriptional pause site is involved in both sensing environmental cues and controlling gene expression.

Results

The *thiC* Riboswitch Controls Transcription and Translation

The *E. coli thiC* riboswitch is located upstream of the *thiCEFSGH* operon involved in TPP biosynthesis[11]. Despite the metabolic importance of TPP, no promoter regulation has been reported to modulate *thiC* expression. However, the *thiC* riboswitch was shown to negatively regulate gene expression in TPP-rich media[12], ensuring TPP homeostasis. The riboswitch consists of a conserved aptamer and a large expression platform containing the RBS and AUG start codon (**Fig. 1a**)[13]. In the TPP-bound conformation (OFF state), the RBS and start codon are sequestered within a stem-loop structure predicted to inhibit translation initiation (**Fig. 1a**). Currently, no regulation mechanism has been established to explain how the riboswitch regulates gene expression[12], mainly because the complete

structure of the ligand-free riboswitch is still unknown. To predict the secondary structure of the *thiC* riboswitch in the absence of TPP, we retrieved intergenic sequences located upstream of *thiC* in proteobacteria (see **Supplementary Methods**). By aligning 77 sequences, we identified a highly conserved secondary structure describing the TPP-free ON state (**Fig. 1b** and **Supplementary Fig. 1a,b**) in which an alternative pairing (anti-P1 stem) prevents both P1 and RBS sequestering stems (**Fig. 1a**), suggesting efficient translation initiation. In this model, the ON state structure of the *thiC* riboswitch is mutually exclusive to the TPP-bound conformer, thereby providing a molecular mechanism explaining how the *thiC* riboswitch modulates translation initiation on TPP binding.

Figure 1

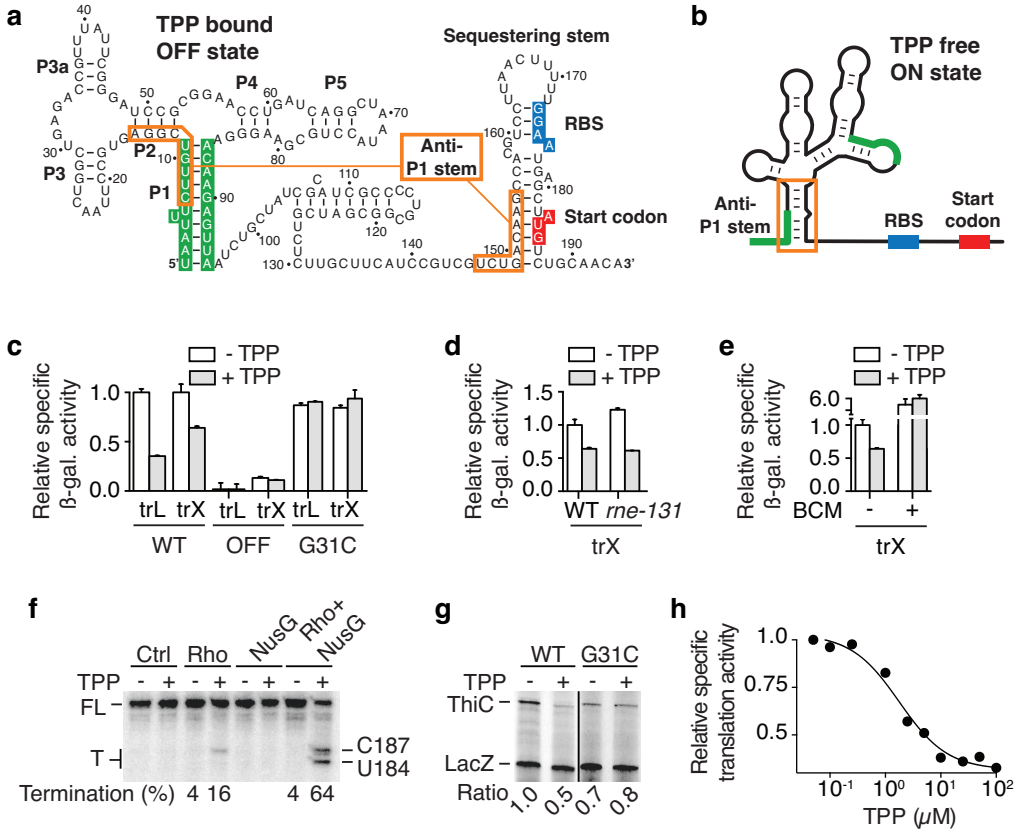


Figure 1. The thiC riboswitch controls gene expression at the levels of transcription and translation.

(a) Secondary structure of the *thiC* riboswitch representing the TPP bound state. The P1 stem, the RBS and the start codon are represented in green, blue and red, respectively. The anti-P1 stem is shown in orange rectangles. The nomenclature of the helical domains is based on a previous study[12].

(b) Schematic representing the TPP-free state determined in this study. The color scheme is the same as in **Fig. 1a**.

(c) ß-Galactosidase assays of translational ThiC-lacZ (trL) and transcriptional *thiC-lacZ* (trX) fusions for the wild type, ON, OFF and G31C mutants. Values were normalized to the activity obtained for the wild type in the absence of TPP. The average values of three independent experiments with standard deviations (SDs) are shown.

(d) ß-Galactosidase assays of transcriptional *thiC-lacZ* (trX) performed in the context of the wild type and *rne-131* strains. Values were normalized to the activity obtained for the wild type in absence of TPP. The average values of three independent experiments with SDs are shown.
(e) ß-Galactosidase assays of transcriptional *thiC-lacZ* (trX) performed in the wild type in absence or presence of TPP or bicyclomycin (BCM). Values were normalized to the activity obtained for the wild type in absence of TPP. The average values of three independent experiments with SDs are shown.
(f) *In vitro* Rho-dependent transcription performed using the *thiC* riboswitch. Full-length (FL) and termination (T) products are indicated on the left. Termination efficiencies are indicated below for reactions done with Rho. Termination products are indicated on the right.
(g) Transcription-translation-coupled *in vitro* assays using the wild type and G31C mutant. Reactions were performed in absence (-) or presence (+) of TPP. Translation products are shown on the left and ratios of ThiC expression are normalized to that of LacZ.
(h) Transcription-translation-coupled *in vitro* assays using the wild type riboswitch. The half transition of translation repression corresponds to a value of 1.7 μM ± 0.3 μM.

We investigated the *thiC* riboswitch regulatory mechanism using *in vivo* reporter gene assays. However, since the transcription start site was unknown, we first determined the transcriptional +1 residue by performing primer extension analysis on *thiC* mRNA (**Supplementary Fig. 2a**). The deduced mRNA sequence differed from previously used transcript versions[12] and the newly determined transcription start site allows formation of an elongated P1 stem including residues 1-6 and 91-95 (**Fig. 1a**). We engineered transcriptional and translational *lacZ* chromosomal *E. coli* constructs comprised of the *thiC* riboswitch sequence and the first 10 codons of *thiC* open reading frame (ORF). While translational fusions report on protein and mRNA levels, transcriptional *thiC-lacZ* fusions have two independent RBS (*thiC* and *lacZ*) allowing us to monitor only *thiC* mRNA levels (**Supplementary Fig. 2b**). The ß-galactosidase activity was repressed by 65% in the presence of TPP in the context of the translational fusion (**Fig. 1c**, WT, trL

construct). TPP-induced gene repression was also observed for the transcriptional fusion (**Fig. 1c**, WT, trX construct), in agreement with the riboswitch modulating *thiC* mRNA levels[12]. To confirm that the genetic repression is controlled through riboswitch conformational changes, we tested constructs in which the riboswitch P1 stem was stabilized in the OFF conformation to negatively regulate *thiC* expression (**Supplementary Fig. 2c**). As expected, ß-galactosidase activities were significantly reduced and were not affected by the presence of TPP (**Fig. 1c**). The introduction of a single-point mutation (G31C) preventing ligand binding[14] resulted in the loss of TPP-dependent gene repression (**Fig. 1c**, G31C construct), indicating that riboswitch metabolite binding is required for genetic regulation.

The *thiC* Riboswitch Modulates Rho-dependent Termination

As previously observed for *E. coli* riboswitches, we speculated that TPP-dependent modulation of *thiC* mRNA levels could result from either RNase E-dependent mRNA decay[5] or Rho-dependent transcription termination[6]. To establish whether RNase E regulates *thiC* mRNA levels, we used the bacterial strain *rne-131* that prevents RNA degradosome assembly[15]. An efficient TPP-mediated regulation was maintained using *thiC-lacZ* transcriptional fusions in the *rne-131* strain (**Fig. 1d**), suggesting that RNase E is not involved in the control of *thiC* mRNA levels. However, we observed a complete loss of TPP-mediated gene regulation in the presence of the Rho inhibitor bicyclomycin[16] (**Fig. 1e**), suggesting that Rho regulates *thiC* mRNA levels. We then determined the position of Rho-dependent transcription termination using single-round *in vitro* transcription assays[6]. Control transcription reactions performed in the absence of Rho did not produce any prematurely terminated products, consistent with the absence of intrinsic transcription termination (**Fig. 1f**, see Ctrl lanes). However, TPP-induced transcription termination (~16%) was detected in presence of Rho (**Fig. 1f**, see Rho lanes). Transcriptional sequencing revealed that the termination position corresponds to C187 (**Supplementary Fig. 3a**). Further addition of the Rho-interacting factor NusG[17,18] yielded significantly stronger TPP-dependent termination (~64%) at both positions U184 and C187 (**Fig. 1f**). Together, these results indicate that TPP binding to the *thiC*

riboswitch results in Rho-dependent transcription termination at positions U184 and C187.

We next sought to identify the Rho binding site in the riboswitch expression platform. Rho binding sequences (*rut* sites) are usually rich in cytosine but poor in guanine, and are relatively unstructured[19]. Close inspection of *thiC* nucleotide composition indicated that the 100-140 nt region is a good match for a typical *rut* site (**Supplementary Fig. 3b**). Partial S1 nuclease digestion on nascent *thiC* RNA showed that positions 102-104 and 137-138 are more accessible in the presence of TPP (**Supplementary Fig. 3c**), suggesting that they are targeted by Rho in the ligand-bound state. Furthermore, we observed an increase in nuclease digestion of these regions in presence of Rho (**Supplementary Fig. 3d**), indicating that Rho binding reorganizes the riboswitch structure. Our data are consistent with Rho crystal structures in which RNA binding sites are proposed to allow the RNA to wrap around the Rho hexamer[20,21]. Interestingly, Rho binding to the *mgtCBR* leader also increases nuclease digestion of the *rut* site[22], suggesting that both *thiC* and *mgtCBR* sequences share similar Rho binding mechanisms. Thus, these results indicate that TPP binding modulates access to a *rut* site located within the *thiC* riboswitch expression platform.

The *thiC* Riboswitch Directly Controls Translation Initiation

Based on the secondary structure model, it is predicted that the *thiC* riboswitch modulates the initiation of translation upon TPP binding (**Fig. 1a**). To directly monitor TPP-dependent control of *thiC* translation initiation, we carried out transcription-translation-coupled *in vitro* assays using the cell-free PURExpress system[23]. As an internal control, we included a DNA template corresponding to a *lacZ* fragment in the same reaction (see **Supplementary Methods**). In the absence of TPP, ThiC and LacZ translation products were detected (**Fig. 1g**, WT, - lane). However, we observed a 2-fold reduction of ThiC products in the presence of TPP, consistent with the riboswitch repressing translation initiation on ligand binding (**Fig. 1g**, WT, + lane). No such TPP-dependent repression was observed when using a G31C mutant riboswitch (**Fig. 1g**, G31C), in agreement with *in vivo* data (**Fig. 1c**). Using the wild type construct, we next performed *in vitro* translation assays while varying the TPP concentration. Fitting analysis revealed a half

transition value of 1.7 μM ± 0.3 μM for translation repression (**Fig. 1h**). Given that the *in vivo* TPP concentration is in the low micromolar range[24], it suggests that the *thiC* riboswitch is well tuned to detect relevant cellular TPP variations in *E. coli*.

TPP Binding Induces Riboswitch Conformational Rearrangements

To investigate how TPP modulates the structure of the *thiC* riboswitch, we used the well-established RNase H cleavage assay that was previously used to investigate the conformation of nascent *E. coli btuB* riboswitch mRNAs transcribed by *E. coli* RNA polymerase[25,26]. We used a similar approach in which we designed 10-mer DNA oligonucleotide probes to target the P1 stem (positions 10-19) and the RBS region (positions 172-181) (**Fig. 2a**). Nascent RNAs were first studied using the P1 probe in the absence or presence of TPP. While a strong cleavage product was observed in the absence of TPP (**Fig. 2a**), a complete loss of RNase H cleavage was obtained in presence of the ligand (**Fig. 2a**), consistent with TPP stabilizing the formation of the P1 stem (**Fig. 1a**). We obtained similar results using a DNA oligonucleotide targeting the RBS domain (**Fig. 2a**), indicating that the RBS sequestering stem is also stabilized upon TPP binding.

Next, we performed RNase H assays to study the TPP-induced *thiC* riboswitch conformational change as a function of TPP concentration. The cleavage efficiency of RNase H targeting the P1 stem progressively decreased as the TPP concentration increased (**Fig. 2b** and **Supplementary Fig. 4a**). Fitting analysis yielded a half TPP concentration (K_{switch}) value of 422 nM ± 27 nM for the structural modulation. A similar K_{switch} value of 453 nM ± 13 nM was obtained using the RBS probe (**Fig. 2b** and **Supplementary Fig. 4b**). Given that both K_{switch} values obtained using the P1 and RBS probes are ~4-fold higher than the 100 nM dissociation constant of the TPP-*thiC* aptamer complex[12], it suggests that the *thiC* riboswitch operates under a kinetic regime where the transcription rate is important for ligand binding, as reported for *B. subtilis* flavin mononucleotide (FMN)[27] and adenine[28] riboswitches. As expected for a kinetic regime, we found that K_{switch} values varied with decreasing NTP concentrations (**Supplementary Table 1**), consistent with slower transcription rates allowing riboswitch regulation at lower ligand concentrations[27,28].

Figure 2

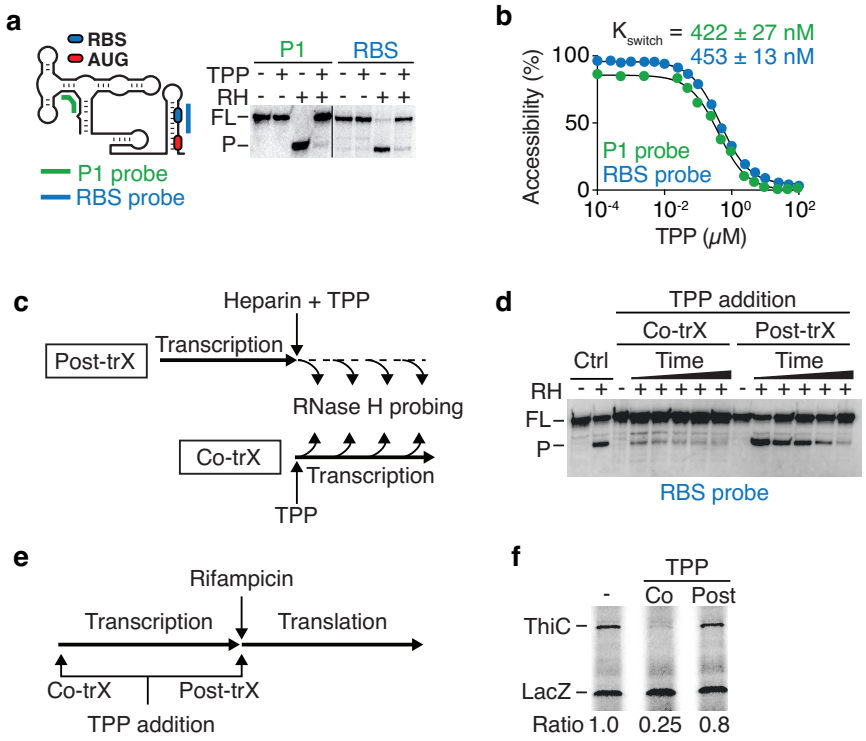


Figure 2. The thiC riboswitch performs TPP sensing cotranscriptionally.

(a) Schematic of the construct used for RNase H probing assays and the locations of P1 (green) and RBS (blue) probes. The right panel shows RNase H assays performed in the absence (-) or presence (+) of TPP or RNase H (RH), using the P1 or RBS probe. The full-length (FL) and cleaved product (P) are shown on the left.

(b) K_{switch} determination for the wild type riboswitch. *In vitro* transcription reactions were done in presence of increasing TPP concentrations ranging from 100 pM to 100 μM. Experiments were done in triplicate and the data shown are a representative result.

(c) Experimental setup monitoring TPP binding to the *thiC* riboswitch. Post-transcriptional TPP binding (Post-trX) was assessed by performing transcription in absence of TPP. After stopping transcription, TPP is added and RNase H assays are performed at several time points. Cotranscriptional TPP binding (Co-trX) was assessed by adding TPP before transcription initiation and by performing RNase H assays during transcription.

(d) RNase H probing assays monitoring TPP binding to the *thiC* riboswitch. Control reactions (Ctrl) done in presence of RNase H (RH) show a cleaved product (P) without TPP. Post-transcriptional and cotranscriptional reactions were performed as indicated in **Fig. 2c**. The full-length (FL) and cleaved product (P) are indicated on the left.

(e) Experimental setup monitoring TPP binding using transcription-translation *in vitro* assays. The effect of post-transcriptional (Post-trX) TPP binding was assessed by initiating the transcription reaction in absence of TPP. After the addition of rifampicin, amino acids and TPP were added to allow TPP binding to the riboswitch during translation. The cotranscriptional (Co-trX) TPP binding was assessed by adding TPP at the beginning of transcription.

(f) Transcription-translation *in vitro* assays assessing TPP binding to the riboswitch. Reactions were performed in absence (-) or presence of TPP added either cotranscriptionally (Co) or post-transcriptionally (Post) as indicated in **Fig. 2e**. Ratios of ThiC expression normalized to that of LacZ.

The *thiC* Riboswitch Performs TPP Sensing Cotranscriptionally

To investigate the kinetics of TPP binding, we assessed the efficiency of metabolite sensing in a cotranscriptional or post-transcriptional context using RNase H assays. In these experiments (**Fig. 2c**), TPP was added either cotranscriptionally or post-transcriptionally (after the addition of heparin). As expected, control experiments showed that the RBS region is accessible to RNase H cleavage in the absence of the ligand (**Fig. 2d**, Ctrl, + lane). Cotranscriptional binding of TPP strongly protected the RBS, even at the first time point (**Fig. 2d**, Co-trX, 15 s). However, post-transcriptional TPP addition yielded weaker RNase H protection for early time points (15 s, 45 s and 90 s) (**Fig. 2d**, Post-trX), suggesting a slower rate of TPP binding. Fitting analysis yielded a fast binding rate having a lower bound of 0.20 s^{-1} ± 0.04 s^{-1} for cotranscriptional TPP binding and an apparent binding rate of 0.05 s^{-1} ± 0.01 s^{-1} for post-transcriptional binding (**Supplementary Fig. 5a**), consistent with cotranscriptional ligand binding being ~4-fold more efficient. Similar results were obtained using a P1 probe (**Supplementary Fig. 5b,c**). Given that rifampicin treatments showed that *thiC* mRNA half-life is extremely

short *in vivo* (**Supplementary Fig. 5d**, <60 s), our results support the idea that the *thiC* riboswitch regulates gene expression in bacteria by binding TPP cotranscriptionally.

We next tested cotranscriptional TPP binding using transcription-translation *in vitro* assays. Here, transcription was uncoupled from translation by omitting the addition of amino acids during the transcription step (**Fig. 2e**). In the second step, rifampicin and amino acids were added to stop transcription and initiate translation, respectively (**Fig. 2e**). As shown in **Fig. 2f**, while the cotranscriptional binding of TPP reduced ThiC expression by ~4-fold, no effect was observed when TPP was added at the beginning of the translation step (post-transcriptionally). These results indicate that the transcription process is a major determinant for both TPP sensing and regulation of translation initiation.

Pausing in the Start Codon Vicinity is Important for Control

The high efficiency of TPP cotranscriptional recognition by the *thiC* riboswitch suggests that metabolite sensing is performed within a defined transcriptional window. Of the key factors involved in the transcription process, RNAP transcriptional pausing has been shown to influence RNA folding in numerous systems[7,25,27–29] and to be important for Rho-dependent transcription termination[7,30]. To determine RNAP pausing within the *thiC* riboswitch, we performed single-round *in vitro* transcription assays and analyzed transcripts at various time points (**Fig. 3a**). Three pause regions were observed in the riboswitch expression platform that encompassed positions A138, C158 and C187 (**Supplementary Fig. 6a**). Interestingly, not only is RNAP pausing at C187 remarkably efficient compared to both A138 and C158, but also the half-life of the pause is increased by ~2.8-fold in presence of TPP (**Fig. 3b** and **Supplementary Fig. 6b**). That the C187 pause half-life is TPP-dependent suggests that the *thiC* riboswitch structure is involved in RNAP pausing. Accordingly, we found that destabilizing the U160-A175 hairpin by introducing a U160A mutation abolished the effect of TPP on the C187 pause site (**Fig. 3b** and **Supplementary Fig. 6c**).

Figure 3

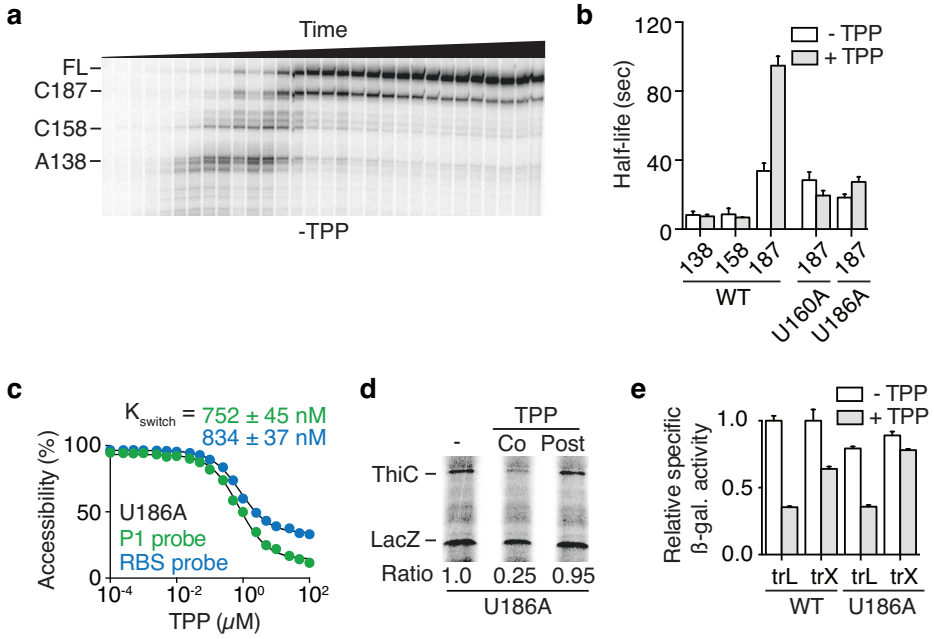


Figure 3. The *thiC* riboswitch exhibits transcriptional pausing at the translation start site important for transcriptional control.

(a) Transcriptional pausing of the *thiC* riboswitch in the absence of TPP using 25 μM NTP. This nucleotide concentration is used to allow easier detection of transcriptional pausing. Pause sites A138, C158 and C187 are indicated on the left. The mapping of the pauses sites is shown in **Supplementary Fig. 6a**.

(b) Quantification of the half-life of transcriptional pause sites in the wild type and the U160A and U186A mutants. Experiments were performed using 100 μM NTP in absence and presence of 10 μM TPP. Quantification data for pauses A138 and C158 are found in the **Supplementary Table 2**. The average values of three independent experiments with SDs are shown.

(c) K_{switch} determination using the P1 or RBS probe for the U186A riboswitch mutant. *In vitro* transcription reactions were done with increasing TPP concentrations ranging from 100 pM to 100 μM. K_{switch} values of 752 nM ± 45 nM and 834 nM ± 37 nM were obtained using the P1 and RBS probes, respectively. Experiments were done in triplicate and the data shown are a representative result.

(d) Transcription-translation *in vitro* assays assessing TPP binding to the U186A riboswitch mutant. Reactions were performed in absence (-) or presence of 25 μM TPP added either cotranscriptionally (Co) or post-transcriptionally (Post) as indicated in **Fig. 2e**. Ratios of ThiC expression are indicated below.
(e) ß-Galactosidase assays of translational ThiC-lacZ (trL) and transcriptional *thiC-lacZ* (trX) fusions for the wild type and U186A mutant. Enzymatic activities were determined in absence and presence of 0.5 mg per mL TPP. Values were normalized to the activity obtained for the wild type in the absence of TPP. The average values of three independent experiments with SDs are shown.

Given that Rho-dependent transcription termination occurs at C187 (**Fig. 1f**), we speculated that RNAP pausing at C187 could be important for transcription termination. To verify this hypothesis, we identified a mutation (U186A) in the *thiC* riboswitch that strongly inhibits TPP-dependent pausing at C187 (**Fig. 3b** and **Supplementary Fig. 7a**). We found that the U186A mutation severely impaired Rho transcription termination *in vitro* (**Supplementary Fig. 7b**), suggesting a critical role for the C187 pause site in Rho termination activity. The U186A mutation did not alter TPP-mediated riboswitch conformational changes when using RNase H (**Fig. 3c** and **Supplementary Fig. 7c,d**) or in transcription-translation assays (**Fig. 3d**). Accordingly, *in vivo* assays revealed that while a U186A translational *lacZ* fusion exhibits regulation similar to the wild type construct (**Fig. 3e**, U186A, trL), the efficiency of TPP-dependent mRNA regulation is reduced when using transcriptional lacZ fusions (**Fig. 3e**, U186A, trX). These results suggest that the *thiC* riboswitch does not rely on Rho transcription termination to regulate at the translational level. Together, *in vitro* and *in vivo* results clearly indicate the critical role of the C187 pause site in controlling Rho transcription termination on TPP binding.

Monitoring TPP Binding in Transcription Elongation Complexes

To determine how cotranscriptional TPP binding modulates the structure of the *thiC* riboswitch, we developed a structural probing approach using RNAP transcription elongation complexes (EC) halted at specific positions. In these experiments, we halted various EC using a biotin-streptavidin roadblock located downstream of the stalled

position[31]. For example, RNAP stalling at the first pause site (EC-138) was obtained by introducing a biotin-streptavidin complex at position C147 (**Fig. 4a**). As expected, the presence of the 138-mer transcript was found to be biotin and streptavidin-dependent (**Supplementary Fig. 8a**). Furthermore, the 138-mer nascent transcript was still detected after a wash step in which transcription complexes were attached to cobalt beads via the RNAP His tag (**Supplementary Fig. 8b**), consistent with the presence of intact transcription elongation complexes.

We first monitored the binding of TPP to EC-138 using RNase H assays. In the absence of TPP, we obtained 75% cleavage efficiency using the P1 probe (**Fig. 4b**, EC-138), indicating that the P1 stem is mostly accessible under this conditions. However, adding TPP significantly decreased cleavage efficiency (down to 8%), giving a cleavage ratio of 9.4 (**Fig. 4b**, EC-138). These results suggest that TPP sensing is efficiently performed in the context of EC-138. The efficiency of TPP binding was reduced when using upstream complexes EC-100 and EC-108, giving cleavage ratios of 1.3 and 6.7, respectively (**Fig. 4b**, EC-100 and EC-108). The strongly impaired TPP sensing ability of the EC-100 complex could be due to RNAP hindering P1 stem formation. However, we observed efficient TPP sensing when RNAP was stalled at the second pause site (**Fig. 4b**, EC-158).

Figure 4

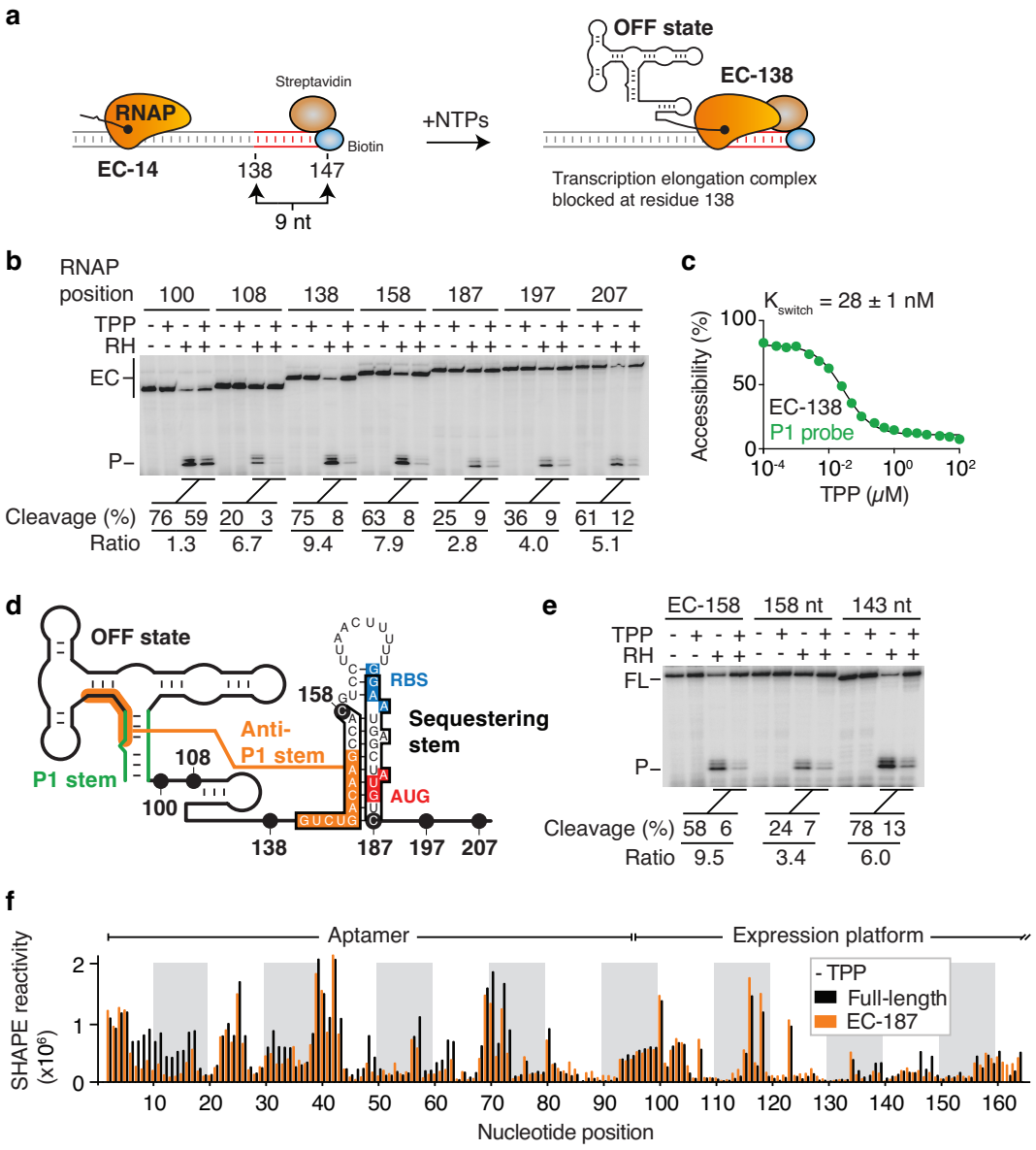


Figure 4. The TPP binding efficiency is modulated along the *thiC* riboswitch transcriptional pathway.

(a) Schematic of the experimental system for the formation of elongation complexes at position 138 (EC-138). Transcription initiation performed using a subset of NTP leads RNAP to stall at position 14 (EC-14). Using a DNA template containing a biotin-streptavidin complex, the addition of all four NTP results in RNAP to block at nine residues upstream from the end of the template, thus producing stable EC-138 complexes.

(b) RNase H probing assays performed on *thiC* riboswitch elongation complexes. Reactions were done in absence (-) or presence (+) of 10 μM TPP and in the absence (-) or presence (+) of RNase H (RH). Uncleaved elongation complexes (EC) and cleaved products (P) are indicated on the left. Cleavage efficiencies and ratios are indicated below.

(c) K_{switch} determination using the P1 probe for the EC-138 complex. Cleavage reactions were done with increasing TPP concentrations ranging from 100 pM to 100 μM. A K_{switch} value of 28 nM ± 1 nM was obtained using the P1 probe. The corresponding gel is shown in **Supplementary Fig. 9a**. Experiments were done in triplicate and the data shown are a representative result.

(d) Schematic of the *thiC* riboswitch showing the different stalled RNAP positions. For pause sites C158 and C187, the predicted sequence contained in the RNAP exit channel is boxed. The localization of C158 and C187 pause sites allows RNAP to either inhibit or allow the formation of the anti-P1 stem, respectively.

(e) RNase H probing assays done on the EC-158 elongation complex and on 158 nt and 143 nt *thiC* transcripts. Reactions were done in absence (-) or presence (+) of 10 μM TPP and in absence (-) or presence (+) of RNase H (RH). Uncleaved (FL) and cleaved products (P) are indicated on the left. Cleavage efficiencies and ratios are indicated below.

(f) SHAPE modification of the *thiC* riboswitch performed on the full-length transcript and EC-187 elongation complex in absence of TPP. The histogram represents the relative SHAPE reactivity for positions 1 to 164.

Strikingly, we obtained a markedly reduced cleavage ratio of 2.8 at the third pause site (EC-187), suggesting that TPP binding is not efficient for this transcription complex (**Fig. 4b**, EC-187). Nevertheless, TPP sensing was partially recovered in complexes located at downstream ORF positions (EC-197 and EC-207) (**Fig. 4b**). These results indicate that although metabolite binding is efficient when RNAP is located at the first or second pause site (EC-138 and EC-158), TPP modulation decreases significantly at the third pause site (EC-187).

We measured the TPP binding affinity of transcription complexes by performing RNase H assays using a range of TPP concentrations. We found that TPP sensing is more potent when RNAP is stalled at pause sites 138 (28 nM ± 1 nM) or 158 (27 nM ± 2 nM) (**Fig. 4c** and **Supplementary Fig. 9a,b**) compared to the full-length transcript (**Fig. 2b**). However, we obtained a markedly increased K_{switch} value of 2.1 μM ± 0.3 μM for EC-187 (~100-fold higher than EC-138 or EC-158; **Supplementary Fig. 9c**), indicating that TPP binding is severely impaired at this pause site. The large difference in TPP affinity suggests that the *thiC* nascent transcript in the context of EC-187 folds into a structure different from those adopted in EC-138 and EC-158. Indeed, given that the last 14 transcribed residues (positions 176-187) are protected within RNAP[27,3231], we expect that the RNAP stalled at position 187 prevents the formation of the RBS sequestering stem. As a result, the folding of the long-range anti-P1 stem is favored, locking the aptamer domain into a non-competent TPP binding structure. This structure cannot be formed in EC-138 or EC-158 since the 3' moiety of the anti-P1 stem is either not yet transcribed (EC-138) or contained in the RNAP exit channel (EC-158) (**Fig. 4d**). We confirmed the negative influence of the anti-P1 stem on TPP binding by introducing destabilizing anti-P1 stem mutations at positions 147-149 in EC-187 (**Supplementary Fig. 9d**), where increased TPP sensing affinity was measured for the mutant (410 nM ± 53 nM) in comparison to the wild type EC-187 complex (2.1 μM ± 0.3 μM).

In contrast to EC-187, we speculated that the presence of RNAP in EC-158 would prevent the formation of the anti-P1 stem, thereby positively influencing TPP binding (**Fig. 4d**). As expected, RNase H probing of an identical 158 nt transcript in the absence of RNAP resulted in less efficient TPP-mediated structural change compared to EC-158 (**Fig. 4e**, 158 nt, ratio = 3.4), consistent with the anti-P1 stem perturbing TPP sensing. In agreement with this, using a shorter 143 nt transcript that did not allow the anti-P1 stem to form showed an increased cleavage ratio (**Fig. 4e**, 143 nt). Together, our results support a model in which selective exposure of the 146-154 nt region dictates anti-P1 stem formation, which acts as an intramolecular lock preventing TPP sensing.

SHAPE Analysis of Transcription Elongation Complexes

RNA structure can be studied at the nucleotide level using selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE)[28,33–40]. We adapted this highly informative technique to interrogate the RNA structure in the context of EC-187. We produced nascent *thiC* transcripts using *E. coli* RNAP and either halted them at position 187 (EC-187) or allowed RNAP to transcribe the full riboswitch sequence. Immediately after transcription we probed the RNA structure using 2-methylnicotinic acid imidazolide (NAI). Interestingly, SHAPE reactivity was similar for the EC-187 nascent RNA and full-length transcript, with the noteworthy exception of the C7-G16 region (**Fig. 4f** and **Supplementary Fig. 10a**). The decrease in reactivity for residues C7-G16 in EC-187 suggests that this region is involved in a stable structure, such as the long-range anti-P1 stem interaction (**Fig. 1a**). However, in absence of TPP, the higher SHAPE reactivity of C7-G16 in the full-length transcript suggests that the anti-P1 stem is not the predominant form. This is presumably due to an equilibrium shift toward the formation of the RBS sequestering stem that can form in the context of the full-length riboswitch, but not in the context of EC-187.

We next repeated similar experiments but added TPP cotranscriptionally. For both the EC-187 complex and the full-length transcript, adding NAI resulted in reactivity changes in both the aptamer and expression platform (**Supplementary Fig. 10b-d**). For each species, most TPP-induced changes observed in the aptamer domain were similar to those reported for the *thiM* riboswitch[39,41]. For example, NAI reactivity increased for positions C53, G54, A56, U63 and A80. The expression platform produced a TPP-dependent reactivity profile consistent with the predicted structure in which residues U131, U132, C141, G142 and U146 were more reactive in the presence of TPP (**Supplementary Fig. 10b-d**). Taken together, SHAPE analyses show that, in the context of EC-187, the *thiC* riboswitch adopts a TPP-free structure in which the anti-P1 stem is stabilized.

Riboswitches Exhibit Pausing in Start Codon Vicinity

To determine if transcriptional pausing in the translation start region is widespread among *E. coli* riboswitches, we performed transcription kinetics for the *thiB*, *thiM*, *ribB*, *btuB* and *lysC* riboswitches (**Fig. 5** and **Supplementary Fig. 11-15**). In all cases, we found a clear pausing region in the vicinity of the RBS and start codon (**Fig. 5a**). We did not detect a significant metabolite-dependent effect on half-life pausing for *btuB, thiM* and *lysC* riboswitches (**Fig. 5b,c,d**). However, while adding TPP resulted in a marked reduction in half-life pausing in the region of position 117 of the *thiB* riboswitch (**Fig. 5e**), the *ribB* riboswitch showed an increased pause half-life in the start region with FMN (**Fig. 5f**), in agreement with Rho transcription termination[6]. In all cases, the localization of the pause site in the vicinity of the RBS/start codon suggests the formation of an anti-P1 stem by allowing RNAP either to sequester a competing sequence (*thiB*, *thiM*, *btuB* and *lysC*) or by delaying the transcription of the anti-P1 competing sequence (*ribB*) (**Supplementary Fig. 11-15**). The predicted anti-P1 stem of the *lysC* riboswitch is particularly large (**Supplementary Fig. 13**), suggesting very efficient translation initiation. This is in agreement with *lysC* exhibiting the highest protein synthesis rate in minimal media among *E. coli* riboswitches[42]. Detailed biochemical characterization will be required to elucidate the role of riboswitch pause sites. The presence of multiple pause sites within riboswitch expression platforms is intriguing and strongly suggests that the transcription process is a key element in riboswitch ligand sensing and gene regulation.

Figure 5

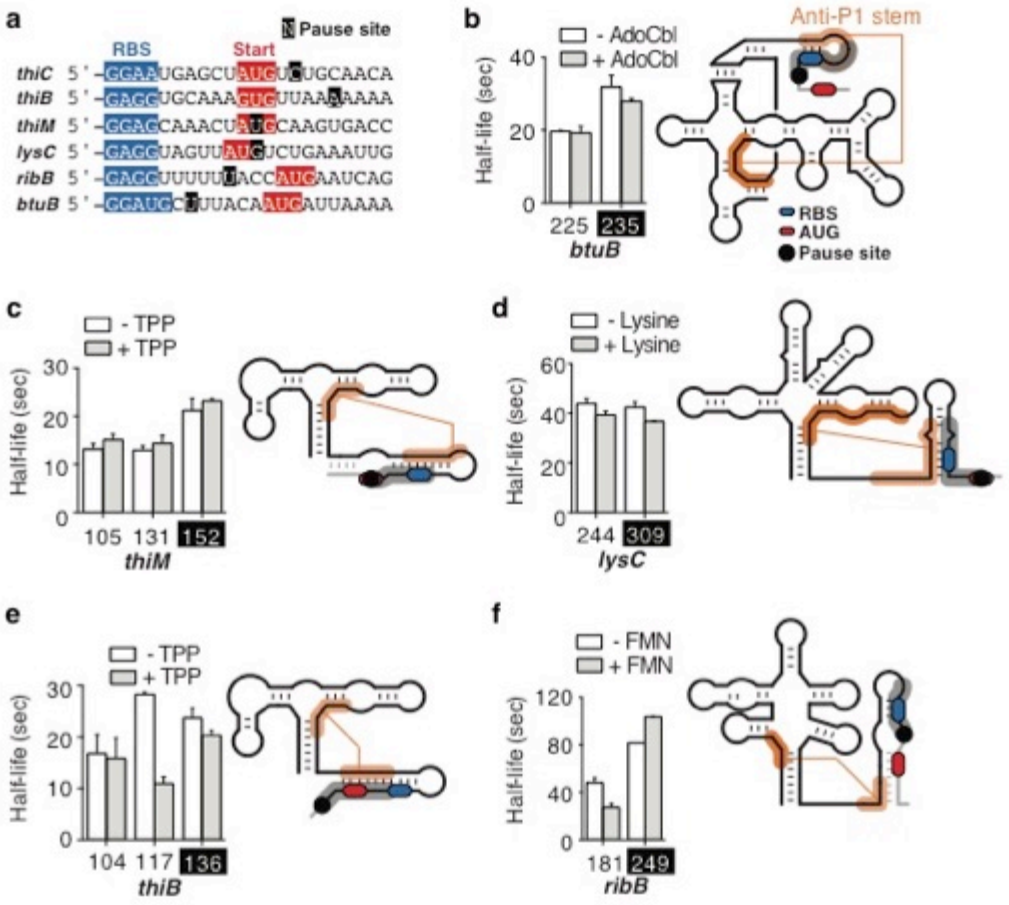


Figure 5. Transcriptional pause regions in *E. coli* riboswitches.

(a) Sequences of *E. coli* riboswitch start codons. A single position designates each pausing region, centered around positions 187 (*thiC*), 136 (*thiB*), 152 (*thiM*), 309 (*lysC*), 249 (*ribB*) and 235 (*btuB*). Pause sites, RBS and AUG start codon are indicated in black, blue and red, respectively.

(b) Transcriptional pausing in the *btuB* riboswitch. A schematic representation of the *btuB* riboswitch is shown on the right and the 235 pause site is indicated. The gray region represents the section of the riboswitch transcript sequestered by the RNA polymerase when stalled at the pause site, thus allowing anti-P1 formation (shown in orange lines). The RBS and AUG start codon are represented in blue and red, respectively. The pause regions and riboswitch sequence are shown in **Supplementary Fig. 11**.

(c) Transcriptional pausing in the *thiM* riboswitch. A schematic representation of the *thiM* riboswitch is shown on the right and the pause site at residue 152 is indicated. The color scheme is identical to the one used in **Fig. 5b**. The pause regions and riboswitch sequence are shown in **Supplementary Fig. 12**.
(d) Transcriptional pausing in the *lysC* riboswitch. A schematic representation of the *lysC* riboswitch is shown on the right and the pause site at residue 309 is indicated. The color scheme is identical to the one used in **Fig. 5b**. The pause regions and riboswitch sequence are shown in **Supplementary Fig. 13**.
(e) Transcriptional pausing in the *thiB* riboswitch. A schematic representation of the *thiB* riboswitch is shown on the right and the pause site at residue 136 is indicated. The color scheme is identical to the one used in **Fig. 5b**. The pause regions and riboswitch sequence are shown in **Supplementary Fig. 14**.
(f) Transcriptional pausing in the *ribB* riboswitch. A schematic representation of the *ribB* riboswitch is shown on the right and the pause site at residue 249 is indicated. The color scheme is identical to the one used in **Fig. 5b**. The pause regions and riboswitch sequence are shown in **Supplementary Fig. 15**.

Discussion

Since their discovery, riboswitches have been extensively studied for their importance in bacterial gene regulation and more recently as potential targets for novel antimicrobial agents[43,44]. Despite the large amount of structural information obtained for aptamer structures[2], only a few studies have addressed riboswitch regulation mechanisms and how metabolite binding coordinates gene expression. Here, we elucidate the regulation mechanism of the *E. coli thiC* riboswitch and show that TPP binding mediates both Rho-dependent transcription termination and repression of translation initiation. We developed an approach in which cotranscriptional TPP-mediated conformational changes were monitored in the context of transcription elongation complexes stalled at specific positions. Our results demonstrate that the *thiC* riboswitch regulates in a narrow transcriptional window in which the downstream boundary is delimited by a critical RNAP pausing site found in the vicinity of the translation initiation codon. Strikingly, while TPP binding affinity is high for transcription complexes located upstream of the

pause site, ligand binding is impaired by ~100-fold when transcription complexes are stalled at this regulatory checkpoint. In contrast, TPP-bound transcription complexes reaching the checkpoint undergo transcription termination or translation repression. Such a regulatory checkpoint provides a strategic gate where the outcome of the genetic decision is sealed. Given that transcriptional pausing occurs widely in *E. coli* and *B. subtilis*[10], our study provides a working model describing how riboswitches—or any structured RNA element—may rely on site-specific transcriptional pausing to regulate bacterial gene expression.

Transient RNAP pausing throughout transcription has been reported to be centrally involved in sequential RNA folding and in the recruitment of key regulatory factors[45]. RNAP pausing was also previously shown to be involved both in sensing environmental cues and controlling gene expression, as previously reported for the pyrBI leader[46], where UTP-specific transcriptional pausing controls translational coupling and intrinsic termination. Similarly, our results show that *thiC* transcriptional pausing is intimately linked to riboswitch TPP sensing and gene regulation, consistent with a recently predicted kinetic model[47]. Based on our study, we propose a mechanism in which TPP sensing is efficiently performed at pause sites A138 and C158, which are expected to provide more time for metabolite sensing (**Fig. 6**, left panel). Importantly, the formation of the anti-P1 stem at C187 suggests that transcriptional complexes located within a narrow transcriptional window between positions G100 and C187 efficiently perform TPP sensing. This corresponds to ~1.6% of the *thiC* operon length, explaining the strong preference of the *thiC* riboswitch to bind ligand cotranscriptionally (**Fig. 2d,f**). The importance of pausing in the *thiC* RBS/AUG region is consistent with recent NET-seq data showing a major pause site at position U188 *in vivo*[10]. Our study indicates that in the absence of TPP, formation of the anti-P1 stem at C187 locks the riboswitch in the ON state, which markedly reduces the metabolite binding affinity of EC-187 (**Fig. 6**, middle panel). However, the majority of incoming TPP-bound transcription complexes reaching position 187 undergo Rho-dependent transcription termination (**Fig. 6**, middle panel). As long as transcription complexes clear the pause site, translation initiation is modulated accordingly (**Fig. 6**, right panel), providing a fail-safe mechanism at the translational

level for TPP-bound complexes not terminated by Rho. Taken together, our results show that the C187 pause site acts as a central checkpoint where gene expression is either secured or repressed.

Figure 6

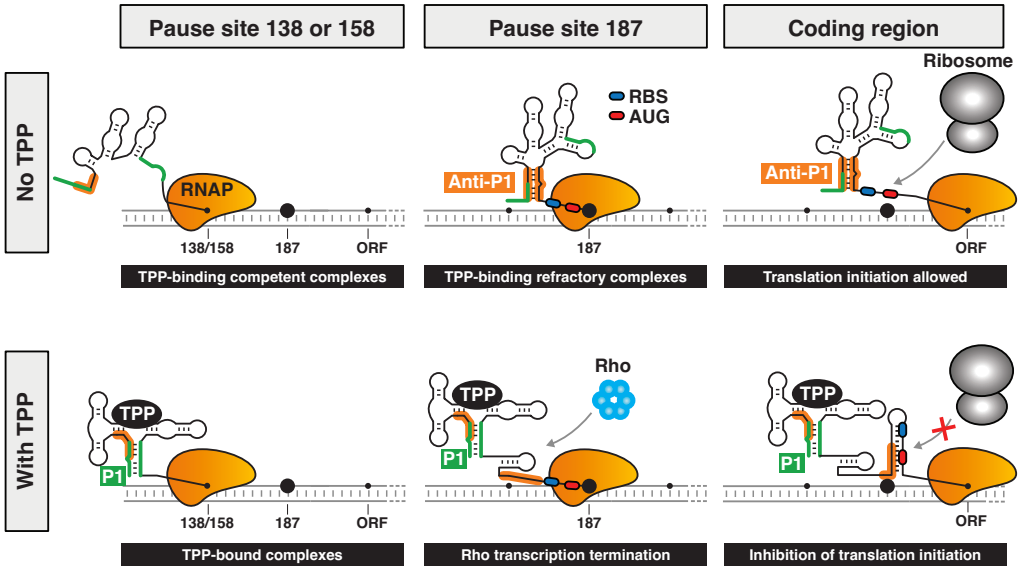


Figure 6. Transcriptional and translational regulatory mechanisms of the *thiC* riboswitch.

Upper panel. In absence of TPP, transcription elongation complexes located at pause sites 138 or 158 exhibit TPP-binding competent riboswitch structures. However, elongation complexes reaching pause site 187 adopt the anti-P1 structure that strongly reduces TPP binding affinity. Elongation complexes reaching the coding region allow efficient 30S ribosomal subunit binding and translation initiation. *Lower panel*. In presence of TPP, elongation complexes located at pause sites 138 or 158 exhibit *thiC* riboswitch structures in which the P1 stem is formed. Upon reaching the 187 pause site, a significant fraction of elongation complexes undergo Rho transcription termination. However, elongation complexes carrying on transcription into the coding region allow the formation of the sequestering stem, inhibiting translation initiation.

In *E. coli*, Rho-dependent termination is involved in many processes, including resolving RNA-DNA hybrids and coupling transcription and translation through polarity[48,49]. Although polarity can result from inefficient ORF translation allowing Rho to bind to RNA, an additional polarity model has been described in which Rho and the ribosome act as alternative partners for the RNAP-bound NusG factor[48,50]. In this model, the formation of ribosome-NusG or Rho-NusG complex ensures either ribosome translation or transcription termination, respectively. The impact of NusG on *thiC* Rho-dependent termination (**Fig. 1f**) strongly suggests that a NusG-based mechanism, rather than a translation inhibition mechanism, directly controls *thiC* mRNA levels, as found in constitutive Rho-dependent terminators. Furthermore, in contrast to *E. coli ribB* and *Salmonella enterica mgtA* riboswitches[6], the *thiC* riboswitch modulates both *rut* site accessibility and pause half-life on metabolite binding, making the *thiC* riboswitch the first example of a mechanism that controls Rho transcription termination through both processes. Our results indicate that the *thiC* riboswitch uses a novel mechanism that selectively modulates *rut* site accessibility through metabolite-induced RNA conformational changes to regulate transcription elongation (**Supplementary Fig. 3c**). This is in contrast to other models in which transcription is allowed to proceed by inactivating Rho either through Hfq[51] or Rho-Antagonizing RNA Elements[22]. Furthermore, our data indicate that Rho termination requires the TPP-mediated increase in C187 pause time, which is abolished by destabilizing the 160-175 nt hairpin structure (**Fig. 3b**, U160A mutant). Such a regulatory mechanism ensures direct control of transcription elongation through TPP-mediated *thiC* riboswitch conformational changes.

The importance of pause sites for riboswitch regulation was first established for the kinetically controlled *B. subtilis ribD* riboswitch[27], where RNAP pausing was found to prevent antiterminator formation. The *thiC* riboswitch represents a more sophisticated level of gene regulation since it selectively prevents and promotes anti-P1 stem formation at pause sites 158 and 187, respectively. Such a stepwise transcription process enables potent TPP sensing (EC-158) and accurate signal transduction (EC-187) to produce the required genetic outcome. Additional studies implied that transcriptional pausing could

coordinate aptamer and expression platform folding[25,28]. Although the precise role of RNAP pausing on riboswitch folding is still to be determined in those cases, a discontinuous transcription mechanism is likely important for RNA folding and metabolite sensing. Even though additional experiments are required to ascertain the single nucleotide resolution mapping of transcriptional pause sites, our findings nevertheless clearly demonstrate the presence of RNAP pausing within the RBS/AUG region of *E. coli* riboswitches. Interestingly, such pauses were also detected *in vivo* for the *lysC* and *thiM* riboswitches[10], consistent with transcriptional pausing being important for gene regulation. Further work will be required to ascertain the influence of transcriptional pausing within *E. coli* riboswitch sequences. The presence of multiple transcriptional pause regions might indicate strong evolutionary pressure on riboswitches to maintain specific metabolite-binding capabilities and transcriptional pausing activities. From an evolutionary standpoint, the presence of already existing pause sites allowing efficient translation initiation probably facilitated the establishment of Rho-dependent transcription termination mechanisms. The identification of pause sites in a wide range of bacterial transcripts strongly suggests that transcriptional dynamics is important for the regulation of bacterial mRNAs.

Methods

DNA Oligonucleotides and Bacterial Strains

DNA oligonucleotides were purchased from Integrated DNA Technologies. Strains used in this study were derived from *E. coli* MG1665 and are described in **Supplementary Table 3**. The strain DH5α was used for routine cloning procedures[58]. The strain BL21 (DE3) was used for overproduction of Rho, NusG and RNA polymerase proteins. The *thiC* transcriptional and translational fusions were constructed using the PM1205 strain[5]. The PCR strategy to obtain *thiC* riboswitch translational constructs is described in **Supplementary Table 4** and the oligonucleotides used are described in **Supplementary Table 5**. Transcriptional fusions required an additional PCR step (AD17 and oligo-stop oligonucleotides) to insert a stop codon. Mutations performed in *thiC* were made using a 3 steps PCR strategy (**Supplementary Table 4**). Briefly, PCR1 and PCR2 were performed using genomic DNA and PCR3 was performed using PCR1 and PCR2

products. The procedure to insert pfr-delta *thiC* in the *rne-131* strain was previously described[5]. All obtained *lacZ* fusions were sequenced to ensure the integrity of constructs.

ß-Galactosidase Assays

Kinetic assays for ß-galactosidase experiments were performed as described previously[5]. Briefly, an overnight bacterial culture grown in M63 minimal medium containing 0.2% glycerol was diluted to an OD_{600} of 0.02 in 50 mL of fresh medium. The culture was incubated at 37°C until an OD_{600} of 0.1 was obtained. Arabinose (0.1%) was added to induce the expression of *lacZ* constructs. ß-galactosidase experiments assessing the involvement of Rho were performed in 3 mL of culture media as described above. TPP (0.5 mg per mL) and bicyclomycin (25 μg per mL) were added as indicated. The experiments were performed in triplicate and the average values with standard deviations (SD) are reported. Descriptions of the mutants are in **Supplementary Table 4**.

RNase H Probing Assays

The PCR strategy to obtain DNA templates used for transcription reactions is described in **Supplementary Table 6**. Aliquots of transcription reactions (described in **Methods**) were mixed with 20 μM DNA oligonucleotides for 5 min. RNase H cleavage was initiated by adding a solution of RNase H in 5 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$, 100 mM KCl, 50 μM EDTA and 10 mM 2-mercaptoethanol at 37°C for 5 min. Reactions were stopped by adding an equal volume of a stop solution (95% formamide, 20 mM EDTA and 0.4 % SDS). The experiments were performed in triplicate and the average values with standard deviations (SD) are reported. To monitor co- or post-transcriptional TPP binding, the transcription reaction was mixed with 200 μM DNA and RNase H for 15 s at 15 s, 45 s, 90 s, 2 min and 3 min after TPP addition. The reported errors for the TPP binding rates are the standard error in the fitting, which is assumed to be approximated by the standard deviation[52]. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Native SHAPE Probing

The SHAPE reaction was essentially prepared as previously described[35]. Native transcription elongation complexes were obtained by incubating 20 mM Tris-HCl, pH 8.0, 100 mM KCl, 20 mM $MgCl_2$, DNA template, *E. coli* RNA polymerase and sigma70 factor at 37°C for 3 min. Reactions were performed in the absence or presence of 10 μM TPP. Transcription reactions were completed by adding NTP at 37°C for 10 min. The EC-187 complex was obtained by transcribing a DNA template containing a biotin on the antisense strand 5' end. Nascent RNA molecules were subjected either to 10% DMSO or 150 mM NAI for 15 min at 37°C. Transcription reactions were quenched by passing complexes through G50 gel filtration columns and purified RNAs were collected in 0.5X TE buffer. Selectively acylated RNA was incubated with different radioactively labeled DNA oligonucleotides to cover the complete riboswitch (1494FPJ, 1495FPJ, 1789FPJ and 1790FPJ) at 65°C for 3 min, then cooled to 22°C for 5 min. Reverse transcription reactions were performed as previously described[35] and products were resolved by denaturing PAGE. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Structural Bioinformatics

To predict the secondary structure of the *thiC* riboswitch in the absence of TPP, intergenic sequences located upstream of the *thiC* were retrieved in proteobacteria using the RiboGap database (http://ribogap.iaf.inrs.ca). Using these sequences, RNA secondary structure prediction was performed using both LocaRNA[53] and CMfinder[54] from which a potential candidate exhibiting an unpaired ribosome binding site (RBS) region was obtained. The validity of the prediction was further tested by analyzing the stability of local optimal structures using RNAConSLOpt[55]. Most predictions corresponded to OFF models, but the single predicted ON conformation was close to the LocaRNA and CMfinder models. The predicted secondary structure model for an alignment of 77 gammaproteobacteria (**Supplementary Data**) is represented in **Supplementary Fig. 1a** and shows the presence of an alternative pairing (anti-P1 stem) preventing both P1 and sequestrator helices (**Supplementary Fig. 1b**), consistent with a TPP-free ON state structure allowing efficient translation initiation. Using the ON state structure, an

additional homology search was performed using INFERNAL on all intergenic regions with predicted TPP riboswitches in sequenced genomes[56] and resulting hits were processed as previously described[57]. A total of 127 additional sequences were found to contain the anti-P1 stem domain but were not included due to additive minor divergences of the ON state riboswitch structure. In our model, the ON state structure of the *thiC* riboswitch (**Supplementary Fig. 1b**) is mutually exclusive to the TPP-bound conformer, thereby providing a molecular mechanism explaining how the *thiC* riboswitch modulates translation initiation upon TPP binding.

In Vitro Transcription

The PCR strategy to obtain DNA templates used for transcription reactions is described in **Supplementary Table 6**. *In vitro* transcriptions were performed in 20 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 20 mM NaCl, 14 mM 2-mercaptoethanol and 0.1 mM EDTA. The DNA template, sigma70 factor and RNAP were incubated at 37°C for 5 min. The UAA trinucleotide, CTP/GTP nucleotides and [α-^{32}P] UTP were then added and the reaction incubated at 37°C for 8 min, thus yielding an elongation complexes stalled at position 14 (EC-14). The sample was passed through G50 columns to remove any free nucleotides. Transcription reactions were completed by adding all four nucleotides with heparin to allow only one round of transcription. TPP was added at 10 μM when indicated. Time pausing experiments were either performed using 25 or 100 μM NTPs. For reactions done at 25 μM NTP, time aliquots were taken from 10 s to 25 min. Time aliquots of 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min were used for reactions done at 100 μM NTP. A chase reaction was done by adding 1 mM NTP. The determination of the half-life of transcriptional pausing was done as previously described[59]. Transcriptional sequencing was done as previously described[28]. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Preparation of Transcription Elongation Complexes

The description of PCR constructs used for the preparation of elongation complexes can be found in **Supplementary Table 6.** Transcription were performed as described above using DNA templates containing a biotin at the 5' end of the antisense strand.

Streptavidin was added to a ratio of 2.5:1 with DNA and incubated for 5 min before adding NTP to complete the transcription reaction. Control experiments using cobalt beads were performed as described before[60]. Beads were resuspended in a solution of 95% formamide, 20 mM EDTA and 0.4% SDS after the wash step. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Rho Transcription Termination Assays

In vitro transcriptions were performed as described above using a transcription buffer containing 40 mM Tris-HCl, pH 8.0, 50 mM KCl, 5 mM $MgCl_2$ and 1.5 mM DTT. Rho (50 nM) was added in the first step of the transcription reaction. Transcriptions were completed by adding 50 μM NTPs and rifampicin (20 μg per mL). The NusG factor (50 nM) was added where indicated. Resulting reactions were treated with a phenol/chloroform extraction and mixed with an equal volume of a stop solution containing 95% formamide, 10 mM EDTA and 0.1% SDS. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

***In Vitro* Translation**

The description of the PCR strategy to obtain DNA templates used for transcription-translation assays can be found in **Supplementary Table 6**. *In vitro* translations were performed with the PURExpress Kit from New England Biolabs. Reactions were performed in the presence of transcription and ribosome solutions, 0.1 μM *thiC* and *lacZ* DNA template, 0.16 μM *E. coli* RNA polymerase, 0.32 μM sigma70 factor and [^{35}S]-methionine at 37°C for 2 h. Reactions were stopped by placing samples on ice for 10 min. Acetone 100% was added to a 4:1 ratio and the resulting solution incubated for 15 min at 4°C. Precipitation was performed by centrifugation at 13,000 rpm for 10 min. Pellets were resuspended in denaturing buffer. Samples were resolved on SDS-PAGE and gels were visualized using a Typhoon FLA 9500 (GE Healthcare). For each experiment, the expression of ThiC was normalized to that of LacZ, which acts as an internal control not modulated by the presence of 25 μM TPP. Titrations were performed using TPP concentrations corresponding to 50 nM, 100 nM, 250 nM, 1 μM, 2.5 μM, 5 μM, 10 μM, 25 μM, 50 μM and 100 μM. For uncoupled transcription-translation assays, the reaction

was initiated by performing a transcription step for 15 min at 37°C using a transcription solution without amino acids/tRNA. Rifampicin (250 μg per mL) was next added to the reaction and incubated for 1 min to prevent transcription re-initiation. Translation was initiated by adding a solution containing amino acids and tRNA and was incubated for 15 min at 37°C. Reactions were stopped as described for transcription-translation coupled assays. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Northern Blot Analysis

The Northern blot analysis was performed as previously described[5]. Briefly, bacteria were grown at 37°C in M63 0.2% glucose minimal medium to midlog phase. Total RNA was extracted using using hot phenol[61]. Rifampicin (250 μg per mL) was added to block transcription. The DNA probe used for the detection is AD113. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Primer Extension Analysis

The determination of the transcription start site (TSS) was done as previously reported[28]. Briefly, 40 μg of total RNA was incubated using a radioactively labeled DNA oligonucleotide. Reverse transcription reactions were performed according to the manufacturer protocol and reactions were precipitated and resolved on denaturing gels. PCR reactions were used as sequencing markers. The DNA oligonucleotide used for the primer extension is AD5. Images have been cropped for presentation. Full size images are presented in Supplementary Fig. 16.

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary Information, or from the corresponding author on request.

References

1. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136,** 615–628 (2009).

2. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152,** 17–24 (2013).

3. Irnov, I., Sharma, C. M., Vogel, J. & Winkler, W. C. Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res* **38,** 6637–6651 (2010).

4. Weinberg, Z. *et al*. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11,** R31 (2010).

5. Caron, M. P. *et al*. Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A* **109,** E3444-53 (2012).

6. Hollands, K. *et al*. Riboswitch control of Rho-dependent transcription termination. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 5376–5381 (2012).

7. Hollands, K., Sevostiyanova, A. & Groisman, E. A. Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. *Proc. Natl. Acad. Sci.* **111,** E1999–E2007 (2014).

8. Bernstein, J. A., Lin, P. H., Cohen, S. N. & Lin-Chao, S. Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci U S A* **101,** 2758–2763 (2004).

9. Peters, J. M. *et al*. Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev.* **26,** 2621–2633 (2012).

10. Larson, M. H. *et al*. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science (80-.).* **344,** 1042–1047 (2014).

11. Begley, T. P. *et al*. Thiamin biosynthesis in prokaryotes. *Arch Microbiol* **171,** 293–300 (1999).

12. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).

13. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S.

Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem* **277,** 48949–48959 (2002).

14. Ontiveros-Palacios, N. *et al*. Molecular basis of gene regulation by the THI-box riboswitch. *Mol Microbiol* **67,** 793–803 (2008).

15. Kido, M. *et al*. RNase E polypeptides lacking a carboxyl-terminal half suppress a mukB mutation in Escherichia coli. *J Bacteriol* **178,** 3917–3925 (1996).

16. Kohn, H. & Widger, W. The molecular basis for the mode of action of bicyclomycin. *Curr Drug Targets Infect Disord* **5,** 273–295 (2005).

17. Shashni, R., Qayyum, M. Z., Vishalini, V., Dey, D. & Sen, R. Redundancy of primary RNA-binding functions of the bacterial transcription terminator Rho. *Nucleic Acids Res* **42,** 9677–9690 (2014).

18. Li, J., Mason, S. W. & Greenblatt, J. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev* **7,** 161–172 (1993).

19. Alifano, P., Rivellini, F., Limauro, D., Bruni, C. B. & Carlomagno, M. S. A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* **64,** 553–563 (1991).

20. Allison, T. J. *et al*. Crystal structure of the RNA-binding domain from transcription termination factor rho. *Nat. Struct. Biol.* **5,** 352–6 (1998).

21. Bogden, C. E., Fass, D., Bergman, N., Nichols, M. D. & Berger, J. M. The structural basis for terminator recognition by the Rho transcription termination factor. *Mol. Cell* **3,** 487–93 (1999).

22. Sevostyanova, A. & Groisman, E. A. An RNA motif advances transcription by preventing Rho-dependent termination. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E6835-6843 (2015).

23. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* **36,** 299–304 (2005).

24. Schyns, G. *et al*. Isolation and characterization of new thiamine-deregulated mutants of Bacillus subtilis. *J. Bacteriol.* **187,** 8127–8136 (2005).

25. Perdrizet 2nd, 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,** 3323–3328 (2012).

26. Lussier, A., Bastet, L., Chauvier, A. & Lafontaine, D. A. A Kissing Loop Is Important for btuB Riboswitch Ligand Sensing and Regulatory Control. *J Biol Chem* **290,** 26739–26751 (2015).

27. 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).

28. Lemay, J. F. *et al*. Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms. *PLoS Genet* **7,** e1001278 (2011).

29. Wong, T. N., Sosnick, T. R. & Pan, T. Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci U S A* **104,** 17995–18000 (2007).

30. Morgan, W. D., Bear, D. G. & von Hippel, P. H. Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. *J. Biol. Chem.* **258,** 9565–9574 (1983).

31. Frieda, K. L. & Block, S. M. Direct observation of cotranscriptional folding in an adenine riboswitch. *Science (80-.).* **338,** 397–400 (2012).

32. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci U S A* **95,** 14699–14704 (1998).

33. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and

primer extension (SHAPE). *J Am Chem Soc* **127,** 4223–4231 (2005).

34. Blouin, S., Chinnappan, R. & Lafontaine, D. A. Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation. *Nucleic Acids Res* **39,** 3373–3387 (2010).

35. Heppell, B. *et al*. Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem Biol* **7,** 384–392 (2011).

36. Stoddard, C. D., Gilbert, S. D. & Batey, R. T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14,** 675–684 (2008).

37. Soulière, M. F. *et al*. Tuning a riboswitch response through structural extension of a pseudoknot. *Proc. Natl. Acad. Sci. U. S. A.* **110,** E3256-64 (2013).

38. Johnson Jr., J. E., Reyes, F. E., Polaski, J. T. & Batey, R. T. B12 cofactors directly stabilize an mRNA regulatory switch. *Nature* **492,** 133–137 (2012).

39. Steen, K.-A., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) for direct analysis of covalent adducts and of nucleotide flexibility in RNA. *Nat. Protoc.* **6,** 1683–94 (2011).

40. Vicens, Q., Mondragón, E. & Batey, R. T. Molecular sensing by the aptamer domain of the FMN riboswitch: A general model for ligand binding by conformational selection. *Nucleic Acids Res.* **39,** 8586–8598 (2011).

41. Steen, K. A., Malhotra, A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by protection from exoribonuclease. *J Am Chem Soc* **132,** 9940–9943 (2010).

42. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157,** 624–635 (2014).

43. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24,** 1558–1564 (2006).

44. Howe, J. A. *et al*. Selective small-molecule inhibition of an RNA structural element. *Nature* **526,** 672–677 (2015).

45. Belogurov, G. A. & Artsimovitch, I. Regulation of Transcript Elongation. *Annu. Rev. Microbiol.* **69,** 49–69 (2015).

46. Donahue, J. P. & Turnbough, C. L. Nucleotide-specific transcriptional pausing in the pyrBI leader region of Escherichia coli K-12. *J. Biol. Chem.* **269,** 18185–91 (1994).

47. Guedich, S. *et al*. Quantitative and predictive model of kinetic regulation by E. coli TPP riboswitches. *RNA Biol.* **13,** 373–90 (2016).

48. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: The RNA 3???-end chronicles. *J. Mol. Biol.* **412,** 793–813 (2011).

49. de Crombrugghe, B., Adhya, S., Gottesman, M. & Pastan, I. Effect of Rho on transcription of bacterial operons. *Nat New Biol* **241,** 260–264 (1973).

50. Burmann, B. M. *et al*. A NusE:NusG complex links transcription and translation. *Science (80-.).* **328,** 501–504 (2010).

51. Rabhi, M. *et al*. The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *EMBO J* **30,** 2805–2816 (2011).

52. Lemay, J. F., Penedo, J. C., Tremblay, R., Lilley, D. M. & Lafontaine, D. A. Folding of the adenine riboswitch. *Chem Biol* **13,** 857–868 (2006).

53. Will, S., Reiche, K., Hofacker, I. L., Stadler, P. F. & Backofen, R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput. Biol.* **3,** 680–691 (2007).

54. Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder--a covariance model based RNA motif finding algorithm. *Bioinformatics* **22,** 445–52 (2006).

55. Li, Y., Zhong, C. & Zhang, S. Finding consensus stable local optimal structures for aligned RNA sequences and its application to discovering riboswitch elements. *Int. J. Bioinform. Res. Appl.* **10,** 498–518 (2014).

56. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29,** 2933–2935 (2013).

57. El Korbi, A., Ouellet, J., Naghdi, M. R. & Perreault, J. Finding instances of riboswitches and ribozymes by homology search of structured RNA with Infernal. *Methods Mol. Biol.* **1103,** 113–26 (2014).

58. Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97,** 5978–5983 (2000).

59. Landick, R., Wang, D. & Chan, C. L. Quantitative analysis of transcriptional pausing by Escherichia coli RNA polymerase: his leader pause site as paradigm. *Methods Enzym.* **274,** 334–353 (1996).

60. Nudler, E., Gusarov, I. & Bar-Nahum, G. Methods of walking with the RNA polymerase. *Methods Enzymol.* **371,** 160–169 (2003).

61. Aiba, H., Adhya, S. & de Crombrugghe, B. Evidence for two functional gal promoters in intact Escherichia coli cells. *J Biol Chem* **256,** 11905–11910 (1981).

Acknowledgments

We thank members of the Lafontaine laboratory and Dr Carlos Penedo and Dr Alain Lavigueur for discussion. JP and D.A.L were supported by the Fonds de Recherche du Québec - Santé as Junior 1 and Senior Scholar, respectively. This work was supported by the Natural Sciences and Engineering Council of Canada (J.P) and Canadian Institutes of Health Research (D.A.L.).

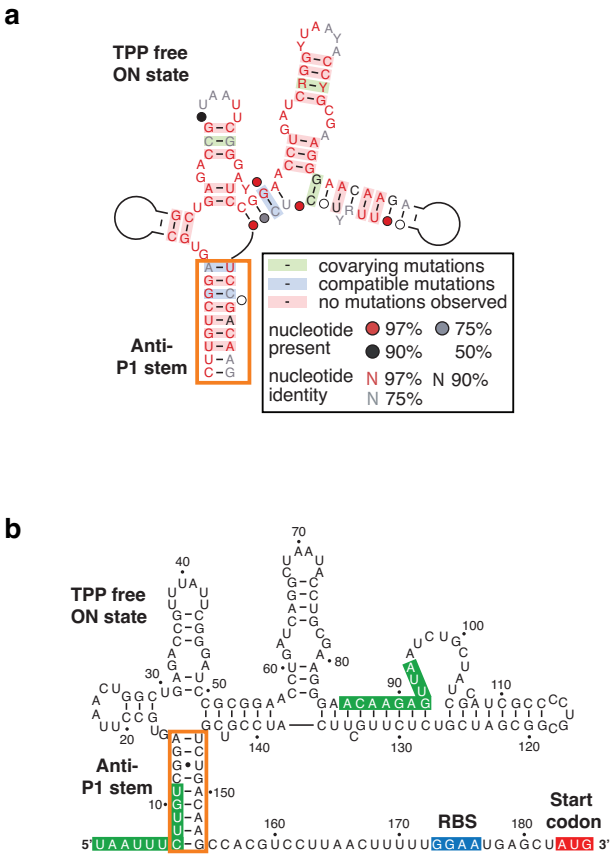
Author contributions

A.C., F.P.J., J.P. and D.A.L. planned and analyzed the experiments. A.C., F.P.J., J.C.B.D., L.B., M.R.N., A.D., P.T. performed the experiments. A.C., F.P.J., J.P. and D.A.L. wrote the paper.

Competing financial interests

The authors declare no competing financial interests.

Supplementary Figure 1

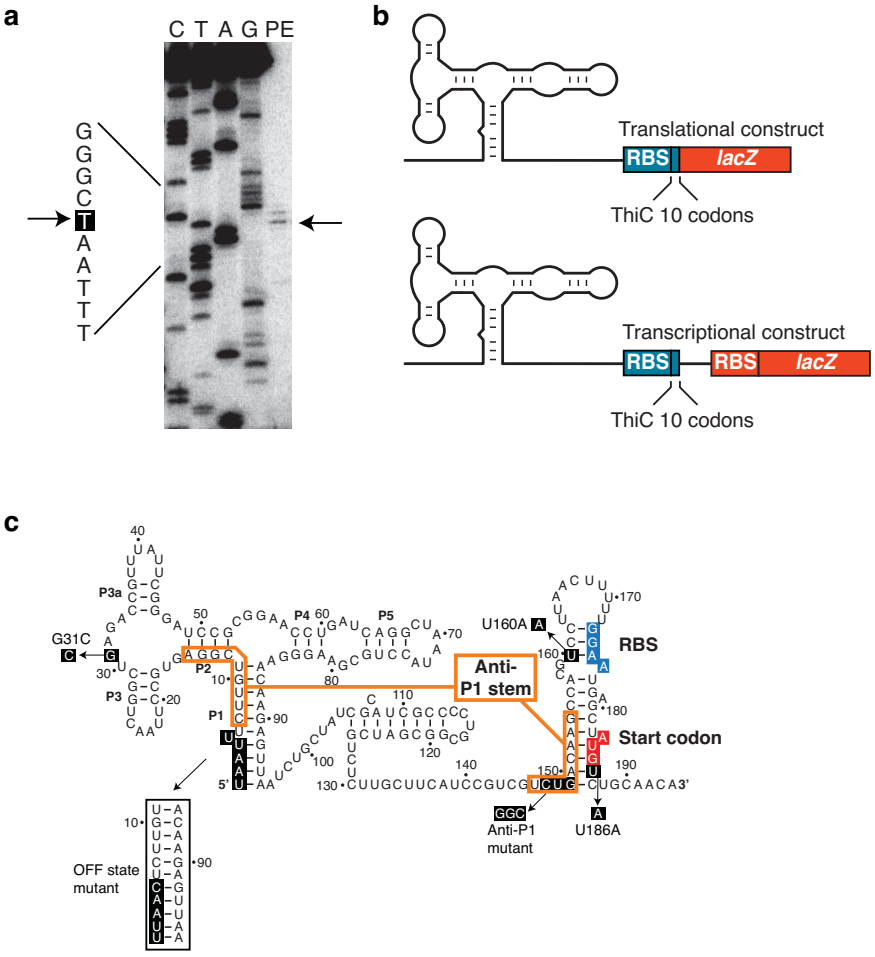


Supplementary Figure 1. Bioinformatic analysis of the *thiC* riboswitch.

(a) Consensus secondary structure obtained from an alignment of 77 different sequences of *thiC* riboswitches. The legend in the figure describes the color coding and the schematic loops represent variable-length stem loops. Nucleotides conserved at least at 97%, 90% or 75% are represented in red, black or gray, respectively. Bases for which the identity is not conserved but that are present in 97%, 90%, 75% or 50% of candidates are indicated by circles in red, black, gray or white, respectively. Base pairs that show conservation, covariation or compatibility (e.g., G-C to G-U) are shown in red, green or blue, respectively. The figure was drawn with R2R[1] from the Stockholm alignment in **Supplementary Data** and was refined to keep as many sequences as possible to best describe the structure corresponding to the *E. coli thiC* riboswitch.

(b) Secondary structure of the *thiC* riboswitch representing the TPP-free state. Sequences involved in the P1 stem, RBS and start codon are shown in green, blue and red, respectively. The anti-P1 stem is shown in an orange rectangle.

Supplementary Figure 2



Supplementary Figure 2. Constructs of the *thiC* riboswitch used in this study.

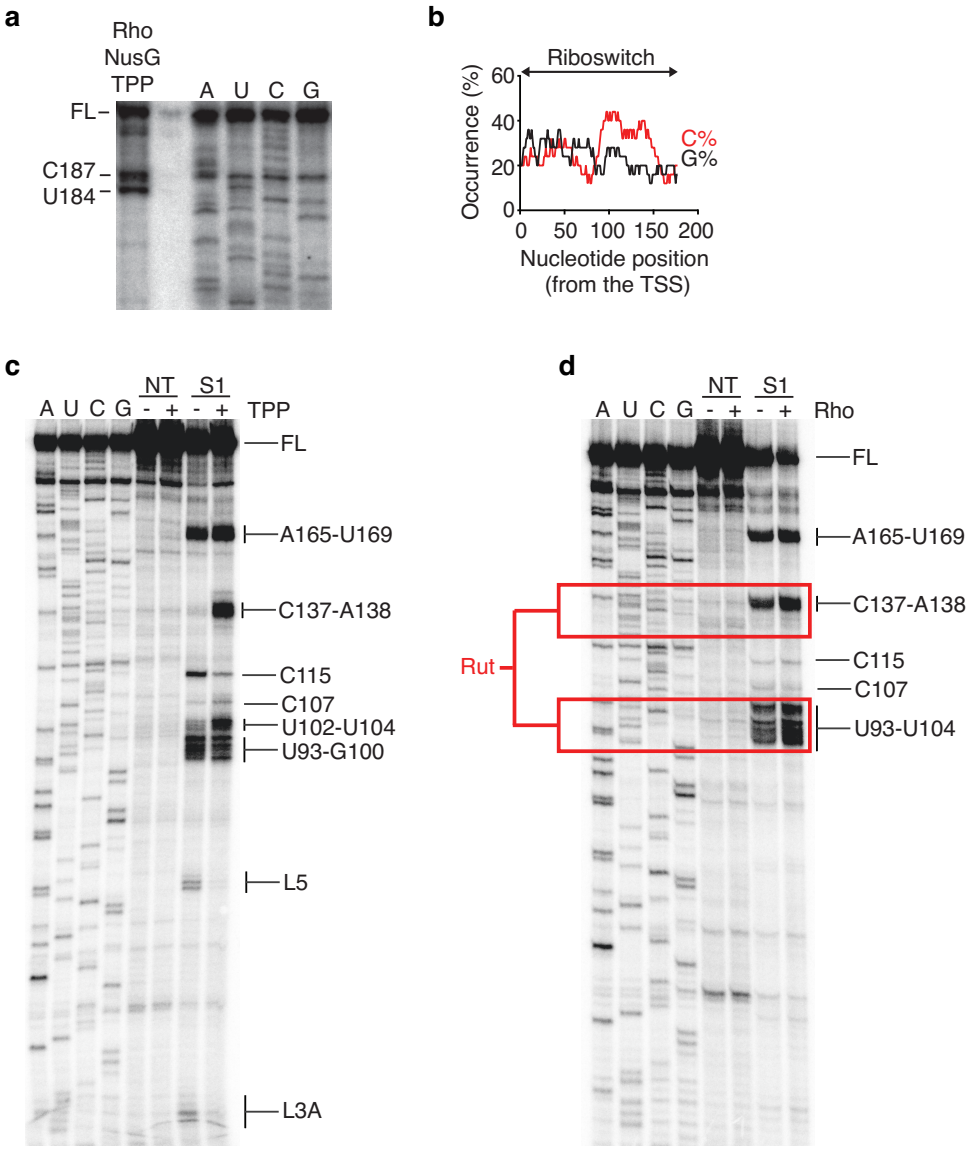
(a) Determination of *thiC* transcription start site by primer extension. Total RNA extracted from *E. coli* was used with oligonucleotide AD5. The arrow represents the +1 transcription start site.

(b) Schematic of the *thiC* translational (top) and transcriptional (bottom) fusions. Both fusions contain an arabinose inducible promoter fused to the *thiC* riboswitch domain with the first 10 *thiC* codons. While the translational fusion reports on both the mRNA and protein levels, the presence of an additional RBS site in the *lacZ* transcriptional fusion allows only to report the level of mRNA.

(c) Secondary structure of the *thiC* riboswitch and mutants used in this study. Mutations are shown in black. The OFF state mutant is expected to stabilize the P1 stem by allowing perfect base pairing between residues 1-5 and 92-96. The anti-P1 stem is

designed to disrupt the formation of the anti-P1 stem when the RNA polymerase is stalled at the C187 pause site. The anti-P1 stem, the RBS and the start codon are shown in orange, blue and red, respectively.

Supplementary Figure 3



Supplementary Figure 3. Rho transcription termination of the *thiC* riboswitch.

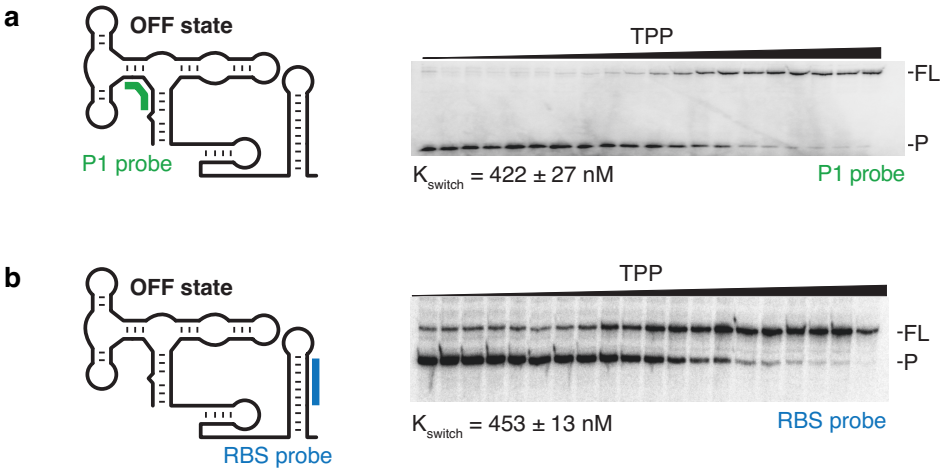
(a) *In vitro* Rho-dependent transcription performed using the *thiC* riboswitch in the presence of 50 nM Rho, 50 nM NusG and 10 μM TPP. The determination of transcription termination sites were determined using 3'-O-Methyl NTPs in the transcription elongation mix.

(b) Sequence analysis of cytosine (C%) and guanine (G%) distribution of the *thiC* riboswitch sequence. A scanning window of 25 nt was used to find the occurrences of G and C as a function or the transcription start site.

(c) Nuclease S1 probing of the *thiC* riboswitch in the absence (-) or presence (+) of 10 μM TPP. Reactions were done in the absence (NT) or presence (S1) of nuclease S1. Regions showing significant cleavage changes are indicated on the right. The determination of S1 cleavage sites sites was performed using 3'-O-Methyl NTPs in the transcription elongation mix.

(d) Nuclease S1 probing of the *thiC* riboswitch performed in the presence of 10 μM TPP. Reactions were done in the absence (-) or presence (+) of 1 μM Rho and in the absence (NT) or presence (S1) of nuclease S1. Regions showing significant cleavage changes are indicated on the right. The predicted rut site is indicated by red boxes. The determination of S1 cleavage sites sites was performed using 3'-O-Methyl NTPs in the transcription elongation mix.

Supplementary Figure 4

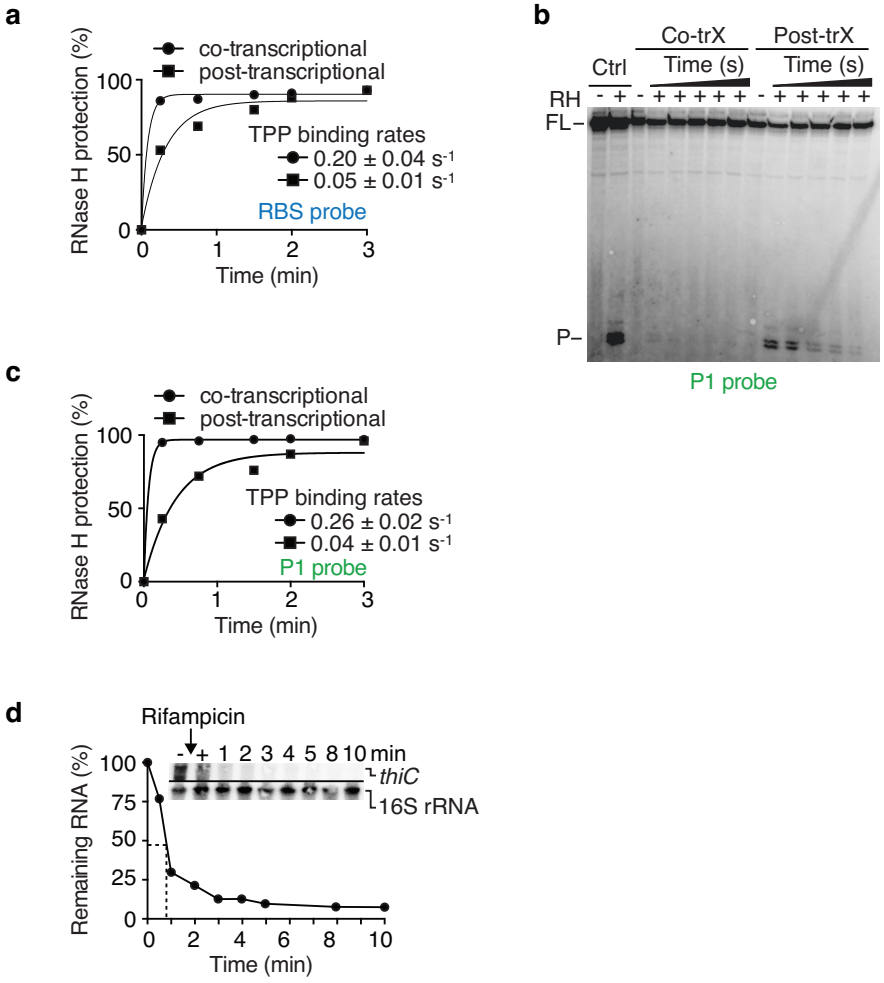


Supplementary Figure 4. K_{switch} determination of the *thiC* riboswitch using P1 or RBS probe.

(a) K_{switch} determination using the P1 probe done in the presence of increasing TPP concentrations ranging from 100 pM to 100 μM. A K_{switch} value of 422 ± 27 nM was obtained. The targeted region by the P1 probe is indicated on the cartoon to the left. Full length (FL) and cleaved products (P) are indicated on the right.

(b) K_{switch} determination using the RBS probe done in the presence of increasing TPP concentrations ranging from 100 pM to 100 μM. A K_{switch} value of 453 nM ± 13 nM was obtained. The targeted region by the RBS probe is indicated on the cartoon to the left. Full length (FL) and cleaved products (P) are indicated on the right.

Supplementary Figure 5



Supplementary Figure 5. Kinetics of TPP binding to the *thiC* riboswitch and half-life of the *thiC* transcript *in vivo*.

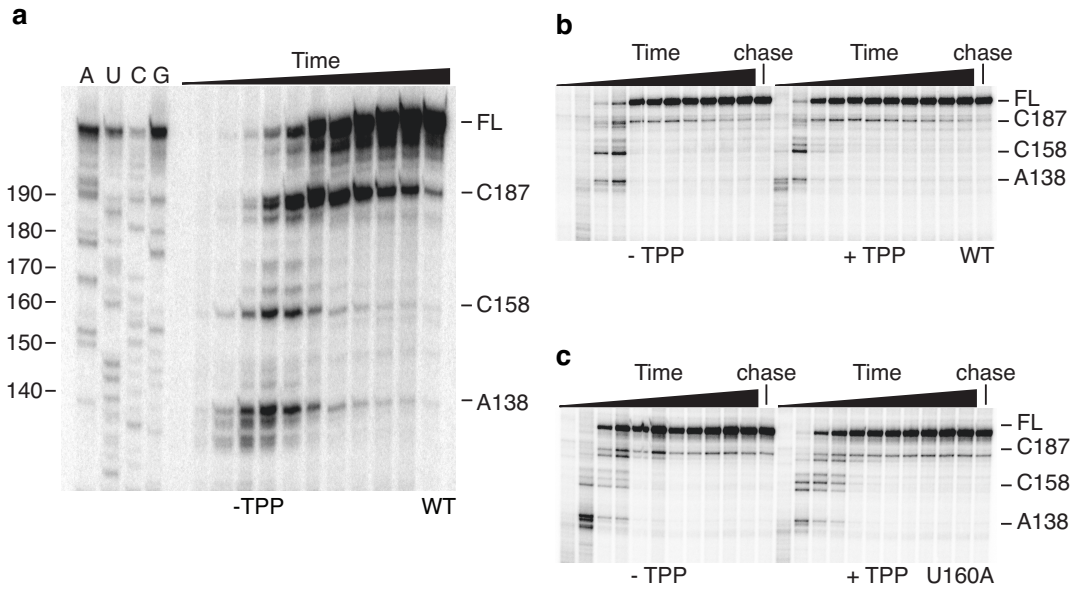
(a) Quantification analysis of RNase H experiments using the RBS probe. Data were fitted to a single-exponential model and yielded apparent TPP binding rates of 0.20 s^{-1} ± 0.04 s^{-1} and 0.05 s^{-1} ± 0.01 s^{-1} for cotranscriptional and post-transcriptional binding, respectively.

(b) RNase H probing assays monitoring TPP binding using the P1 probe. Control reactions (Ctrl) done in the absence (-) or presence (+) of RNase H (RH) indicate the presence of cleaved products (P) in the absence of TPP. Post-transcriptional and cotranscriptional reactions were done as described in **Fig. 2c**. The full length (FL) and cleaved product (P) are indicated on the left.

(c) Quantification analysis of RNase H experiments using the P1 probe. Data were fitted to a single-exponential model and yielded apparent TPP binding rates of 0.26 s^{-1} ± 0.02 s^{-1} and 0.04 s^{-1} ± 0.01 s^{-1} for cotranscriptional and post-transcriptional binding, respectively.

(d) Northern blot assays performed as a function of time using the *E. coli* MG1665 strain[2] at various time points before (-) and after (+) rifampicin (250 μg per mL) addition. The half-life of the mRNA was deduced from the densitometry analysis, which was estimated to be < 60 s.

Supplementary Figure 6



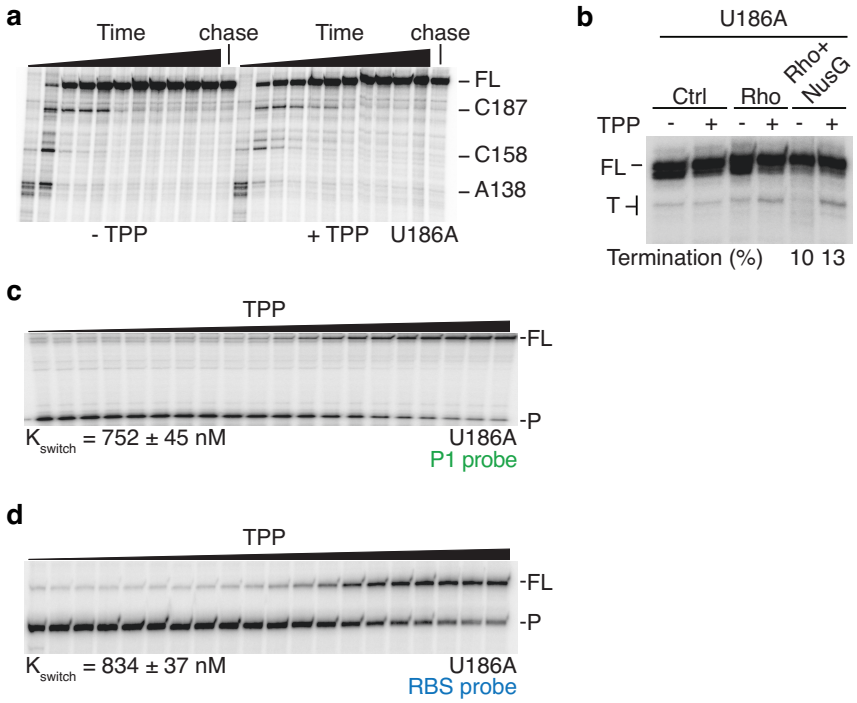
Supplementary Figure 6. Transcriptional pausing of the wild type *thiC* riboswitch and selected mutants.

(a) Mapping of *thiC* riboswitch transcriptional intermediates done in the absence of TPP. Transcription reactions were performed using 25 μM NTPs. Pause sites were mapped by using 3'-O-Methyl NTPs in the transcription elongation mix. Transcriptional intermediates are indicated on the right.

(b) Transcriptional pausing of the wild type *thiC* riboswitch in the absence or presence of 10 μM TPP. Reactions were done in the presence of 100 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs.

(c) Transcriptional pausing of the U160A riboswitch mutant in the absence or presence of 10 μM TPP. Reactions were done in the presence of 100 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs.

Supplementary Figure 7



Supplementary Figure 7. The U186A riboswitch mutant strongly inhibits Rho transcription termination.

(a) Transcriptional pausing of the U186A riboswitch mutant in the absence or presence of 10 μM TPP. Reactions were done in the presence of 100 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs.

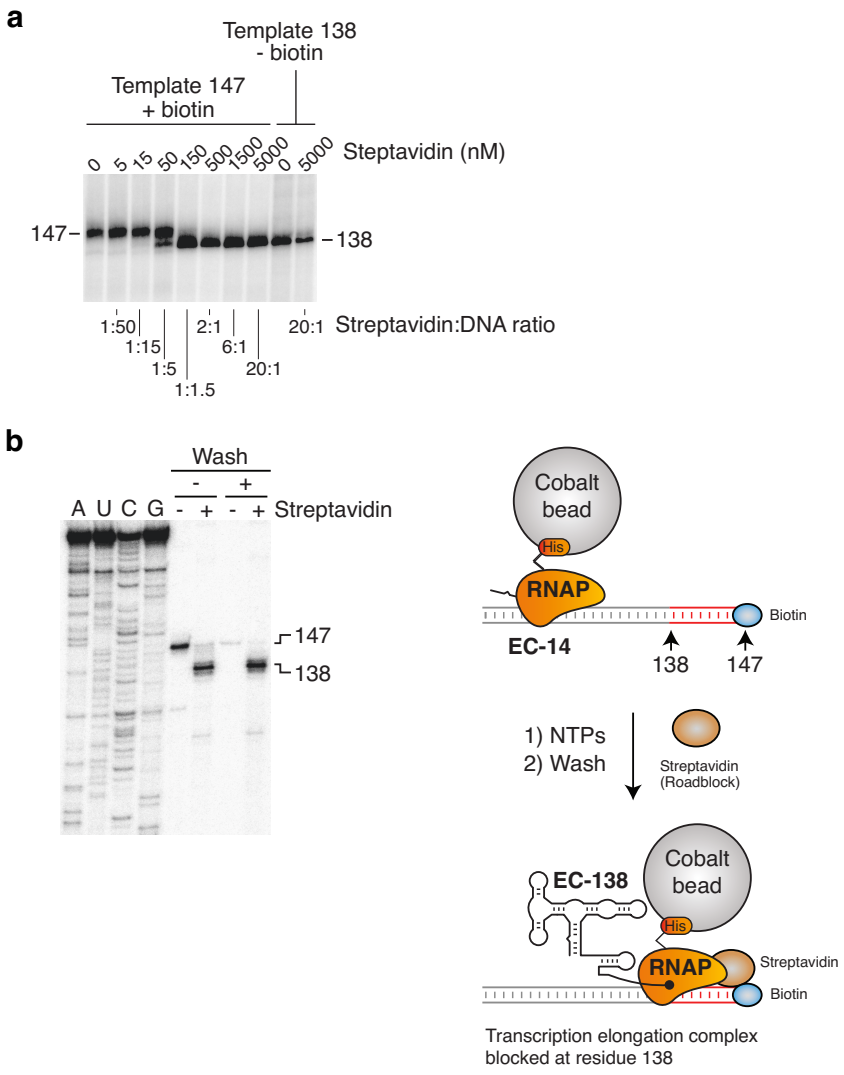
(b) *In vitro* Rho-dependent transcription performed using the U186A riboswitch mutant in the absence (-) or presence (+) of 10 μM TPP. Reactions were done in the absence (Ctrl) or presence of 50 nM Rho or 50 nM NusG

(c) K_{switch} determination of the U186A mutant using the P1 probe done in the presence of increasing TPP concentrations ranging from 100 pM to 100 μM. A K_{switch} value of 752 nM ± 45 nM was obtained. Full length (FL) and cleaved product (P) are indicated on the right.

(d) K_{switch} determination of the U186A mutant using the RBS probe done in the presence of increasing TPP concentrations ranging from 100 pM to 100 μM. A K_{switch} value of 834

nM ± 37 nM was obtained. Full length (FL) and cleaved product (P) are indicated on the right.

Supplementary Figure 8

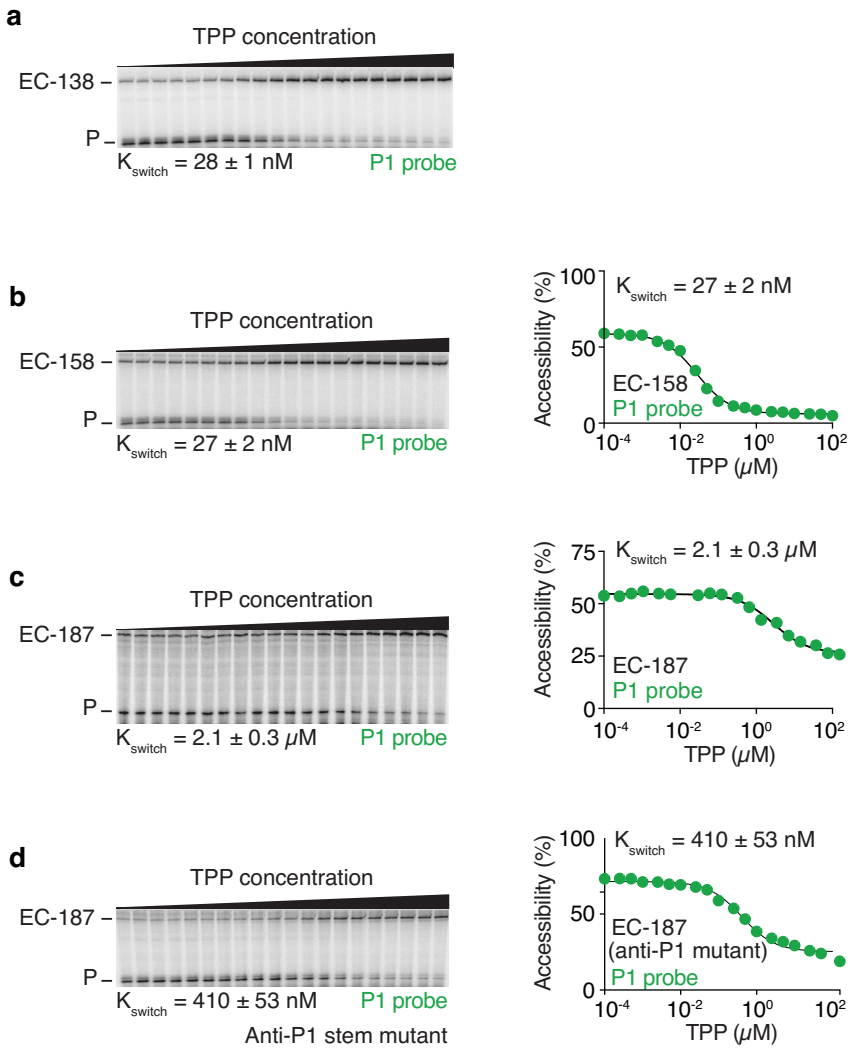


Supplementary Figure 8. Formation of stable transcription elongation complexes using a biotin-streptavidin roadblock.

(a) Transcription reactions performed as a function of streptavidin concentration. In the absence of streptavidin, a transcript of 147 nt is observed. However, the addition of streptavidin results in the appearance of a 138 nt transcript due to the formation of the biotin-streptavidin roadblock. A template allowing the production of a 138 nt transcript is used as a molecular weight marker. The ratio of streptavidin:DNA is shown at the bottom.

(b) Control experiments showing the integrity of transcription elongation complexes stalled at a biotin-streptavidin roadblock. In these experiments, cobalt beads were used as solid support to pull down transcription complexes. The schematic describes the experimental procedure in which the addition of NTPs followed by a washing step yield transcription complexes stalled at position 138. A representative gel shows that the 138 nt transcript is retained after a washing step only in the presence of streptavidin, consistent with the presence of intact elongation complexes. In the absence of a washing step, both 147 nt and 138 nt transcripts are obtained in the absence or presence of streptavidin, respectively.

Supplementary Figure 9



Supplementary Figure 9. Probing *thiC* riboswitch elongation complexes stalled at transcriptional pause sites.

(a) K_{switch} determination of EC-138. Reactions were done as a function of TPP concentration ranging from 100 pM to 100 μM. A K_{switch} value of 28 nM ± 1 nM was obtained using the P1 probe. The quantification is shown in **Fig 4c**.

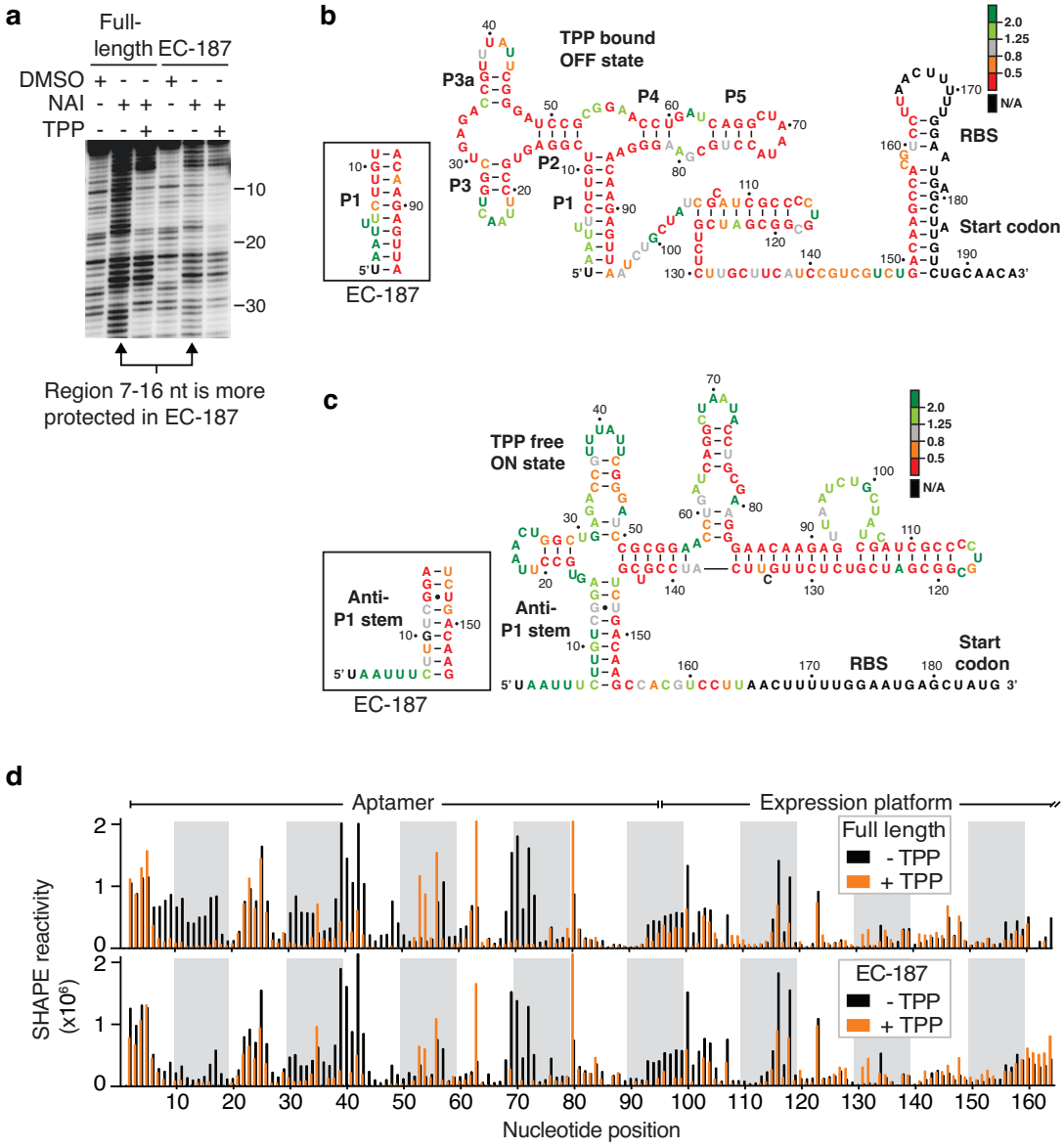
(b) K_{switch} determination of EC-158. Reactions were done as a function of TPP concentration ranging from 100 pM to 100 μM using the P1 probe. A quantification of the data is shown on the right in which a K_{switch} value of 27 nM ± 2 nM was obtained.

(c) K_{switch} determination of EC-187. Reactions were done as a function of TPP concentration ranging from 100 pM to 100 μM using the P1 probe. A quantification of

the data is shown on the right in which a K_{switch} value of 2.1 μM ± 0.43 μM was obtained.

(d) K_{switch} determination of EC-187 in the context of the anti-P1 stem mutant (C147G/U148G/G149C). Reactions were done as a function of TPP concentration ranging from 100 pM to 100 μM using the P1 probe. A quantification of the data is shown on the right in which a K_{switch} value of 410 nM ± 53 nM was obtained.

Supplementary Figure 10



Supplementary Figure 10. SHAPE probing of *thiC* riboswitch elongation complexes stalled at transcriptional pause sites.

(a) SHAPE modifications of the *thiC* riboswitch full-length sequence or in the context of the EC-187 elongation complex performed in the absence (-) or presence (+) of 10 μM TPP and in the absence (-) or presence of 150 mM NAI. Nucleotide positions are indicated on the right. Control experiments were done in the presence of DMSO.

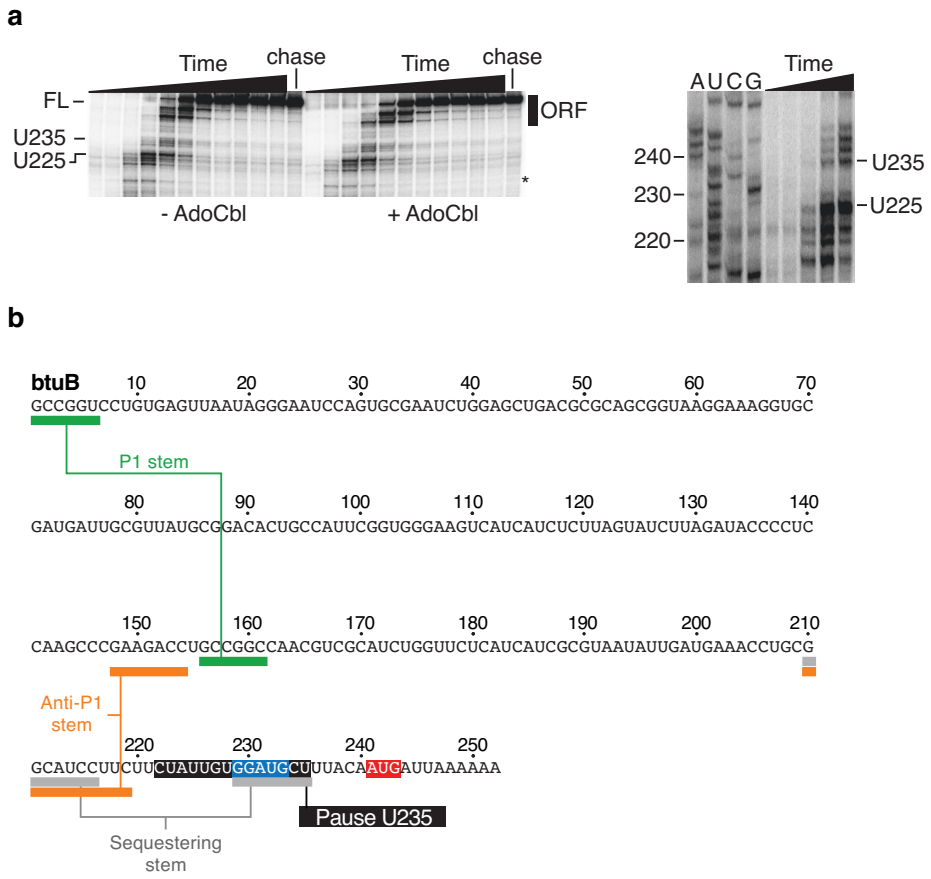
(b) Secondary structure of the *thiC* riboswitch OFF state showing the summary of NAI protection obtained in the presence of 10 μM TPP. SHAPE results for the EC-187

construct are shown in insert. Nucleotides for which no information could be obtained are indicated in black (N/A).

(c) Secondary structure of the *thiC* riboswitch ON state showing the summary of NAI protection obtained in the absence of TPP. SHAPE results for the EC-187 construct are shown in insert. Nucleotides for which no information could be obtained are indicated in black (N/A).

(d) SHAPE modification of the *thiC* riboswitch performed on the full-length transcript and EC-187 elongation complex in the absence and presence of 10 μM TPP. The histogram represents the relative SHAPE reactivity for positions 1 to 164.

Supplementary Figure 11

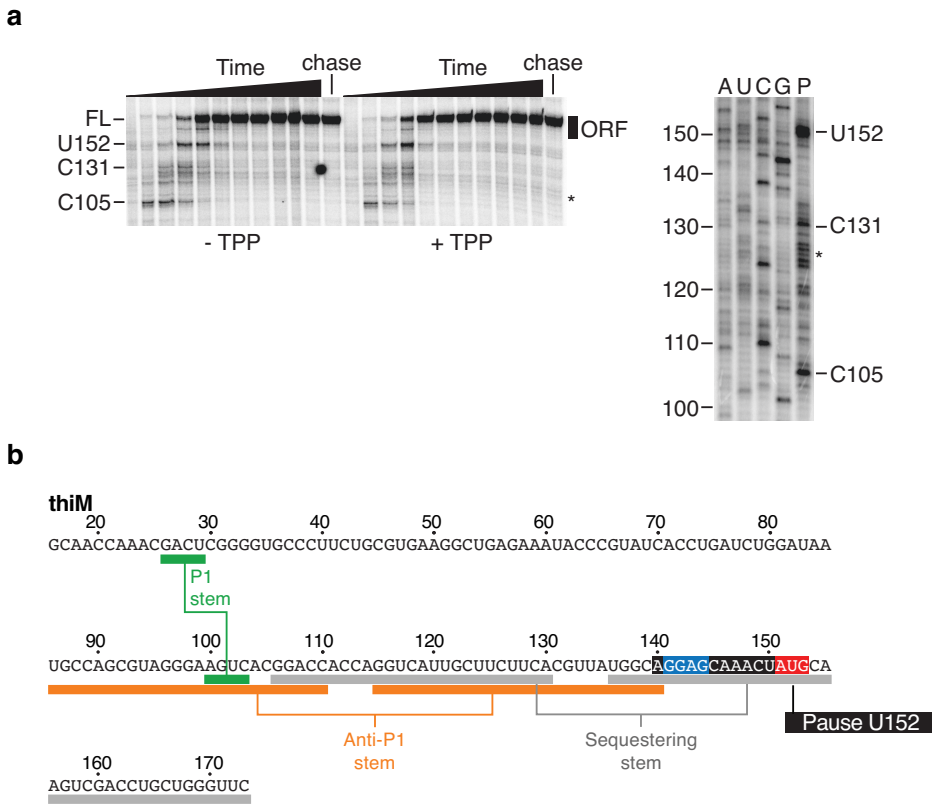


Supplementary Figure 11. Transcriptional pausing of the *btuB* riboswitch.

(a) Transcriptional pausing of the *btuB* riboswitch in the absence or presence of 100 μM AdoCbl. Reactions were done in the presence of 25 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs. Transcriptional intermediates are indicated on the right. Transcriptional pause sites found downstream in the ORF or exhibiting low efficiency (shown by a star) were not taken into account. The right panel shows the pause site mapping of *btuB* transcript intermediates. Reaction times were performed for 15 s, 30 s, 45 s, 60 s and 90 s.

(b) Sequence of the *btuB* riboswitch showing the predicted P1, anti-P1 and sequestering stems are indicated in green, orange and gray, respectively. The expected region sequestered by RNAP at the U235 pause site is shown in black. The prediction of paired regions is based on a previous study[3].

Supplementary Figure 12

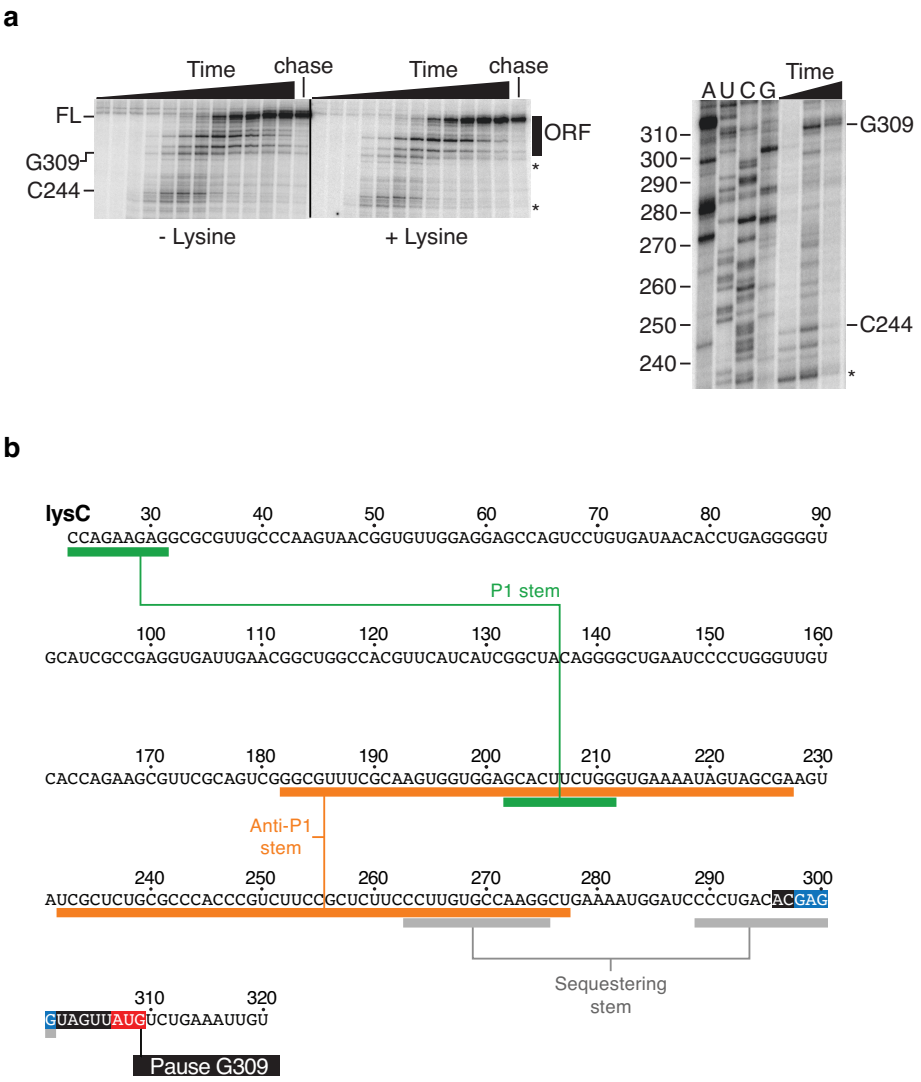


Supplementary Figure 12. Transcriptional pausing of the *thiM* riboswitch.

(a) Transcriptional pausing of the *thiM* riboswitch in the absence or presence of 10 μM TPP. Reactions were done in the presence of 25 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs. Transcriptional intermediates are indicated on the right. Transcriptional pause sites found downstream in the ORF or exhibiting low efficiency (shown by a star) were not taken into account. The right panel shows the pause site mapping of *thiM* transcript intermediates. A reaction time (P) was performed at 45 s.

(b) Sequence of the *thiM* riboswitch showing the predicted P1, anti-P1 and sequestering stems are indicated in green, orange and gray, respectively. The expected region sequestered by RNAP at the U152 pause site is shown in black. The prediction of paired regions is based on a previous study[4].

Supplementary Figure 13

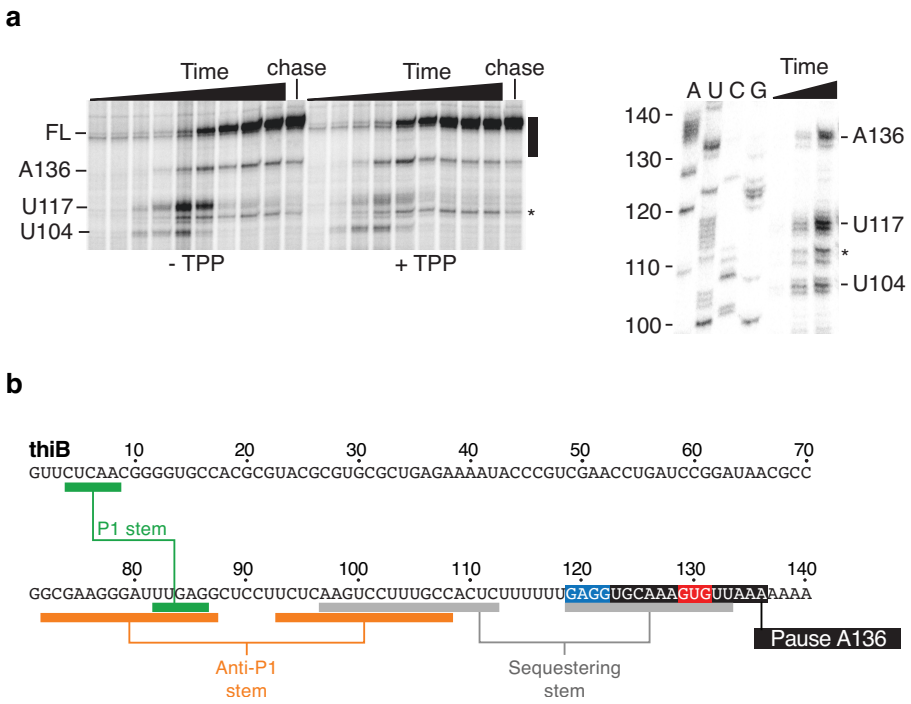


Supplementary Figure 13. Transcriptional pausing of the *lysC* riboswitch.

(a) Transcriptional pausing of the *lysC* riboswitch in the absence or presence of 1 mM lysine. Reactions were done in the presence of 50 μM NTPs and were incubated for 10 s, 30 s, 45 s, 60 s, 90 s, 120 s, 150 s, 3 min, 4 min, 5 min, 6 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs. Transcriptional intermediates are indicated on the right. Transcriptional pause sites found downstream in the ORF or exhibiting low efficiency (shown by a star) were not taken into account. The right panel shows the pause site mapping of *lysC* transcript intermediates. Reaction times were performed for 60 s, 90 s and 120 s.

(b) Sequence of the *lysC* riboswitch showing the predicted P1, anti-P1 and sequestering stems are indicated in green, orange and gray, respectively. The expected region sequestered by RNAP at the G304 pause site is shown in black. The prediction of paired regions is based on a previous study[5].

Supplementary Figure 14

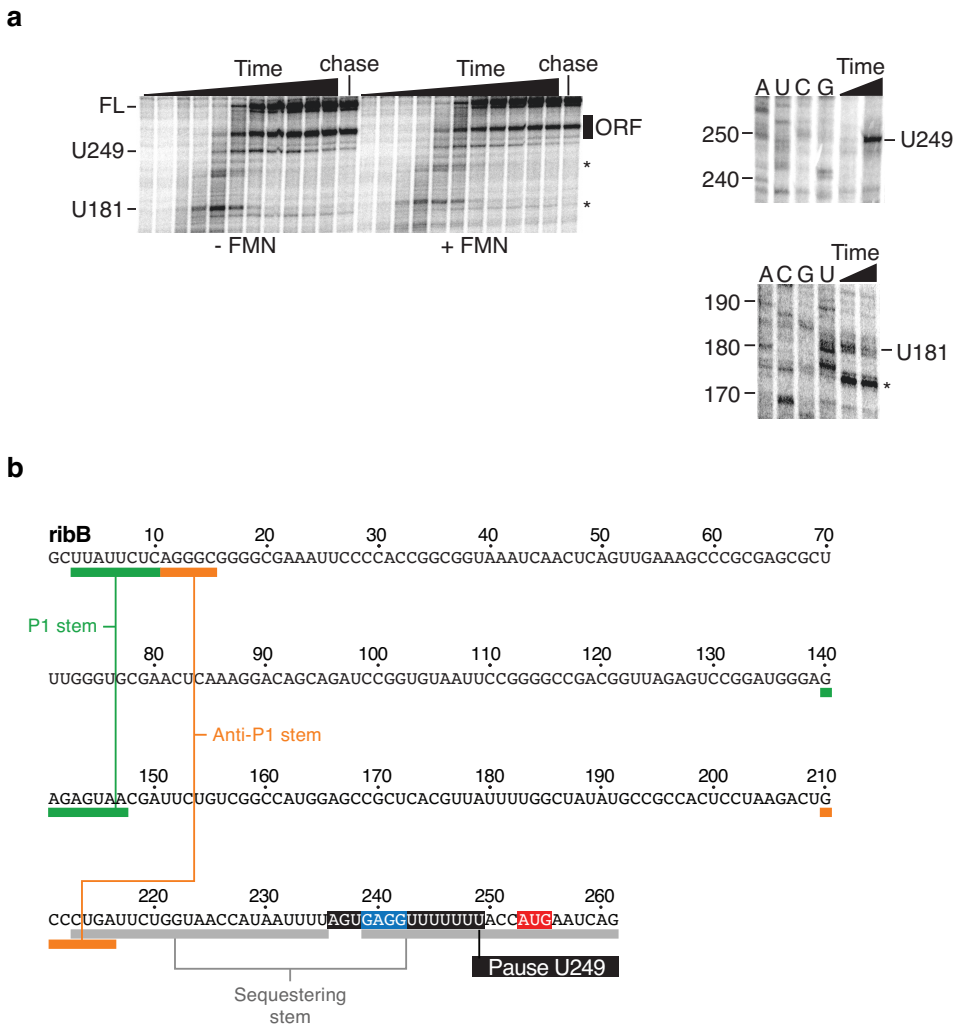


Supplementary Figure 14. Transcriptional pausing of the *thiB* riboswitch.

(a) Transcriptional pausing of the *thiB* riboswitch in the absence or presence of 10 μM TPP. Reactions were done in the presence of 25 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs. Transcriptional intermediates are indicated on the right. The transcriptional pause site located between 104 and 117 (shown by a star) was not taken into account due to the very low fraction escaping the pause site. The right panel shows the pause site mapping of *thiB* transcript intermediates. Reaction times were performed for 15 s, 30 s and 45 s.

(b) Sequence of the *thiB* riboswitch showing the predicted P1, anti-P1 and sequestering stems are indicated in green, orange and gray, respectively. The expected region sequestered by RNAP at the A136 pause site is shown in black. The prediction of paired regions is based on a previous study[4].

Supplementary Figure 15



Supplementary Figure 15. Transcriptional pausing of the *ribB* riboswitch.

(a) Transcriptional pausing of the *ribB* riboswitch in the absence or presence of 10 μM FMN. Reactions were done in the presence of 25 μM NTPs and were incubated for 15 s, 30 s, 45 s, 60 s, 90 s, 2 min, 3 min, 4 min, 5 min, 8 min and 10 min. Chase reactions were done by adding 1 mM NTPs. Transcriptional intermediates are indicated on the right. Transcriptional pause sites found downstream in the ORF or exhibiting low efficiency (shown by a star) were not taken into account. The upper and lower right panels show the pause site mapping of *ribB* transcript intermediates. Due to the large difference in molecular sizes, two different migration times were conducted. Reaction times were performed for 60 s and 90 s (top panel) and 45 s and 60 s (bottom panel).

(b) Sequence of the *ribB* riboswitch showing the predicted P1, anti-P1 and sequestering stems are indicated in green, orange and gray, respectively. The expected region sequestered by RNAP at the U247 pause site is shown in black. The prediction of paired regions is based on a previous study[6].

Supplementary Table 1. Kswitch values determined at various NTP concentrations.

Template	Probes	[NTP]	K_{switch}
Wild type	RBS	1 mM	453 nM ± 13 nM
	RBS	500 μM	222 nM ± 22 nM
	RBS	50 μM	190 nM ± 19 nM
	P1	1 mM	422 nM ± 27 nM
	P1	500 μM	310 nM ± 16 nM
	P1	50 μM	170 nM ± 15 nM

Supplementary Table 2. Half-life values determined in this study.

Templates	TPP	Pause sites	Half-lives
Wild type	-	A138	8 s ± 2 s
	-	C158	9 s ± 3 s
	-	C187	34 s ± 5 s
	+	A138	7 s ± 1 s
	+	C158	7 s ± 1 s
	+	C187	95 s ± 5 s
U160A	-	A138	12 s ± 2 s
	-	C158	9 s ± 5 s
	-	C187	28 s ± 5 s
	+	A138	10 s ± 3 s
	+	C158	8 s ± 3 s
	+	C187	20 s ± 3 s
U186A	-	A138	9 s ± 2 s
	-	C158	12 s ± 1 s
	-	C187	18 s ± 2 s
	+	A138	11 s ± 3 s
	+	C158	12 s ± 2 s
	+	C187	27 s ± 3 s

Supplementary Table 3. Summary of strains or plasmids used in this study.

Strains	Relevant marker	References
EM1055	MG1655 *ΔlacZ* X174	Masse and Gottesman 2002[7]
EM1377	EM1055 *rne131 zce-726*::Tn*10*	Masse *et al*. 2003[8]
PM1205	*lacI*'::P_{BAD}-*cat-sacB-lacZ*, mini I tet^R	Mandin and Gottesman 200[9]
DAL1	BL21(DE3) F–ompT hsdS(rB- mB-) gal dcm λ (DE3)	Laboratory collection
EM1047	DH5a + pACYC184	Laboratory collection
EM1237	DY330 [W3110 delta-lacU169 gal490 lambda-cI857 delta-(cro-bioA)]	Yu *et al*. 2000[10]
	BL21(DE3) pIA247 (nusG)	Laboratory collection
	BL21(DE3) pET28b-rho	Laboratory collection
TPP7	EM1055::pfr-delta-$ThiC_{11cd}$	This study
TPP8	EM1377::pfr-delta-$ThiC_{11cd}$	This study
TPP1	PM1205 lacI′::P_{BAD}-$ThiC_{10cd}$-LacZ	This study
TPP2	PM1205 lacI′::P_{BAD}-*thiC*$_{10cd}$-*lacZ*	This study
TPP3	PM1205 lacI′::P_{BAD}-$ThiC_{10cd}$-G31C-LacZ	This study
TPP4	PM1205 lacI′::P_{BAD}-$ThiC_{10cd}$-G31C-LacZ	This study
TPP5	PM1205 lacI′::P_{BAD}-$ThiC_{10cd}$-mutOFF-LacZ	This study
TPP6	PM1205 lacI′::P_{BAD}-*thiC*$_{10cd}$-mutOFF-*lacZ*	This study
TPP9	PM1205 lacI′::P_{BAD}-$ThiC_{10cd}$-U186A-LacZ	This study
TPP10	PM1205 lacI′::P_{BAD}-*thiC*$_{10cd}$-U186A-*lacZ*	This study

Supplementary Table 4. Summary of lacZ fusions used in this study.

Strains	Constructs	Oligonucleotides
TPP1	$ThiC_{10cd}$	AD17-AD18 (genomic DNA)
TPP2	$thiC_{10cd}$	AD17-AD13 (genomic DNA)
TPP3	$ThiC_{10cd}$-G31C	PCR1: AD17-SR-G31C(2) (genomic DNA) PCR2: SR-G31C(3)-AD18 (genomic DNA) PCR3: AD17-AD18 (PCR1-2)
TPP4	$thiC_{10cd}$-G31C	PCR1: AD17-SR-G31C(2) (genomic DNA) PCR2: SR-G31C(3)-AD13 (genomic DNA) PCR3: AD17-AD13 (PCR1-2)
TPP7	$ThiC_{10cd}$-mutOFF	2174AC-AD18(genomic DNA)
TPP8	$thiC_{10cd}$-mutOFF	2174AC-AD13 (genomic DNA)
TPP11	$ThiC_{10cd}$-U186A	AD17-1377AC (genomic DNA)
TPP12	$thiC_{10cd}$-U186A	AD17-1378AC (genomic DNA)
Cloning in plasmid pfr-delta		
TPP9-10	pfr-delta-$thiC_{11cd}$	PCR1: 1429AC-473AC (genomic DNA) PCR2: 474AC-473AC (PCR1)

Supplementary Table 5. Summary of oligonucleotides used in this study.

Oligonucleotides	Sequences 5'-3'
5AL	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGCCGGTCCTGTGAGTTAATAG
275AL	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGTTCTCAACGGGGTGCCAC
473AC	TCGAATTCCCGGGTTGTTCGCGGCGGGTCAGTTTTGTTGCAGA
474AC	TGGACCGAATTCGGGCACCCCAGGCTTTACACTTTAT
510AC	GCCACGACGGATGAAGCAAGAGACG
511AC	CCGTCGTGGCACAAGCCACGTCCTTAACTT
663AC	TGAAGCAAGAGACGATCGCCGCAG
664AC	GTGGCTTGTCAGACGACGGATGAAGC
721AC	AAAAAGTTAAGGTCGTGGCTTGTCAGACG
722AC	CGTCTGACAAGCCACGACCTTAACTTTTT
765LB	CGGGCTGGTATCCTGTG
827AC	TTGTTCGCGGCGGGTCAGTTTTGTTGCAGTCATAGCTCATTC
918AC	CAACAGCAGGGGCAGACATTTTTTTAACAC
989MG	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGGGCTTAAGTATAAGGAGGAAAAAATATG
1041AC	CGCCCGCGTCAAACATCCTGCTTGA
1168AC	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGTACTACCTGCGCTAGCGCA
1178AC	GTTTTGTTGCAGTCATAGCTCATTCCA
1179AC	TGGAATGAGCTATGACTGCAACAAAAC
1194AC	GCACGTTGGCATCAGAAAGCA
1231MG	AAACCCCTCCGTTTAGAGAGGGGTTATGCTAGTTAGGCGCGTAAAA ATGCGCTC
1232MG	AAACCCCTCCGTTTAGAGAGGGGTTATGCTAGTTAGAACGCTTCCC AGGTAAGATCTTC
1372AC	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGCTGCGATTTATCATCGCAACCAAAC
1377AC	AACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACTTCGCGG CGGGTCAGTTTTGTTGCAGTCAT
1378AC	GTGTGATAAAGAAAGTTAAAATGCCGGATCTTCGCGGCGGGTCAGT TTTGTTGCAGTCAT
1429AC	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGTAATTTCTTGTCGGAG
1489FPJ	5'biotin-GCGGGTCAGTTTTGTTGCAGA
1494FPJ	GACATAGCTCATTCCAAAAAG
1495FPJ	GACATAGCTCATTCCAAAAAGTTAAG
1697AC	5'biotin-ATGCGATAGCAGATTAACTCTTGTTCCC
1698AC	5'biotin-CAGGGGCGATGCGATAGCAGATTAA
1704AC	5'biotin-GTTGTTCGCGGCGGGTCAGTTTTGT
1710AC	TTGTTCGCGGCGGGTCAGTTTTGTTGCAGA
1789FPJ	TGCGATAGCAGATTAACTCTTGT
1790FPJ	GGGCGATGCGATAGCAGATTAAC

2019AC	ACCAAAAGAGGAAAGTAGCGTCTGATTCAT
2020AC	TTGCGCTGAACCCAGCAGGTCGACT
2174AC	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATTTAACTCT TGTCGGAGTGCCTTAACTGG
2247AC	5'biotin-GACGACGGATGAAGCAAGAGACG
2248AC	5'biotin-GTTAAGGACGTGGCTTGTCAGACG
2249AC	5'biotin-TTTGTTGCAGACATAGCTCATTCC
AD5	CGGATCCCGAATAAACGGTCTC
AD13	GTGTGATAAAGAAAGTTAAAATGCCGGATCTTCGCGGCGGGTCAGT TTTGTTGCAGACAT
AD17	ACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATTAATTTCT TGTCGGAGTGCC
AD18	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACTTCGCG GCGGGTCAGTTTTGTTGCAGACAT
AD113	CCGGCGTGATGATGCCCTGGCGGGCGTAGTG
Oligo-stop	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACCATAGC TGTTTCCTGTGTGATAAAGAAAGTTAAAATGCCGGATC
P1	GCACTCCGAC
RBS	GCTCATTCCA
RibBL1	TTCCCTGTACTGATAGGTGTTG
SR-G31C(2)	ATCCCGAATAAACGGTCTGAGCCAGTT
SR-G31C(3)	AACTGGCTCAGACCGTTTATTCGGGAT

Supplementary Table 6. PCR constructs used for *in vitro* RNA synthesis.

Constructions	Oligonucleotides
***In vitro* transcription assays**	
pLacUV5-btuB-315	5AL-765LB (genomic DNA)
pLacUV5-lysC-406	1168AC-1194AC (genomic DNA)
pLacUV5-thiB-158	275AL-918AC (genomic DNA)
pLacUV5-thiM-180	1372AC-2020AC (genomic DNA)
pNat-ribB-282	ribBL1-2019AC (genomic DNA)
pNAT-thiC-138	1041AC-663AC (genomic DNA)
pNAT-thiC-143	1041AC-816AC (genomic DNA)
pNAT-thiC-158	1041AC-664AC (genomic DNA)
pNat-thiC-215	1041AC-1710AC (genomic DNA)
pNat-thiC-215-U186A	1041AC-827AC (genomic DNA)
pNat-thiC-215-U160A	PCR1: 1041AC-721AC (genomic DNA) PCR2: 722AC-1710AC (genomic DNA) PCR3: 1041AC-1710AC (PCR1-2)
***In vitro* transcription-translation assays**	
pLacUV5-lacZ	98MG-1231MG (genomic DNA)
pLacUV5-thiC$_{315cd}$	1429AC-1232MG (genomic DNA)
pLacUV5-thiC$_{315cd}$-G31C	PCR1: 1429AC-SR-G31C(2) (genomic DNA) PCR2: SR-G31C(3)-1232MG (genomic DNA) PCR3: 1429AC-1232MG (PCR1-2)
pLacUV5-thiC$_{315cd}$-U186A	PCR1: 1429AC-1178AC (genomic DNA) PCR2: 1179AC-1232MG (genomic DNA) PCR3: 1429AC-1232MG (PCR1-2)
Transcription elongation complexes	
pNat-thiC-EC100	1041AC-1697AC (genomic DNA)
pNat-thiC-EC108	1041AC-1698AC (genomic DNA)
pNat-thiC-EC138	1041AC-2247AC (genomic DNA)
pNat-thiC-EC158	1041AC-2248AC (genomic DNA)
pNat-thiC-EC187	1041AC-2249AC (genomic DNA)
pNat-thiC-EC187-Anti-P1 mutant	PCR1: 1041AC-510AC (genomic DNA) PCR2: 511AC-2249AC (genomic DNA) PCR3: 1041AC-2249AC (PCR1-2)
pNat-thiC-EC197	1041AC-1489FPJ (genomic DNA)
pNat-thiC-EC207	1041AC-1704AC (genomic DNA)

REFERENCES

1. Weinberg, Z. & Breaker, R. R. R2R--software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12,** 3 (2011).
2. Caron, M. P. *et al*. Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A* **109,** E3444-53 (2012).
3. Nahvi, A. *et al*. Genetic control by a metabolite binding mRNA. *Chem Biol* **9,** 1043 (2002).
4. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419,** 952–956 (2002).
5. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17,** 2688–2697 (2003).
6. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A* **99,** 15908–15913 (2002).
7. Masse, E. & Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A* **99,** 4620–4625 (2002).
8. Masse, E., Escorcia, F. E. & Gottesman, S. Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev* **17,** 2374–2383 (2003).
9. Mandin, P. & Gottesman, S. A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Mol Microbiol* **72,** 551–565 (2009).
10. Yu, D. *et al*. An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97,** 5978–5983 (2000).

CHAPITRE 2

ÉTUDE EN TEMPS RÉEL DU REPLIEMENT CO-TRANSCRIPTIONNEL D'UN ARN

2.1 Présentation de l'article et contribution

Pour étudier la structure de l'ARN, de nombreuses techniques sont actuellement utilisées et vont permettre d'obtenir une idée globale de la structure adoptée par l'ARN. La technique FRET à l'échelle de la molécule unique (smFRET) permet non seulement d'observer des changements structuraux à l'échelle de l'ARN mais également d'observer sa dynamique (Roy et al., 2008). Cet outil puissant a été plusieurs fois employé pour caractériser différentes étapes de la transcription comme l'initiation de la transcription par exemple (Chakraborty et al., 2012; Harden et al., 2016; Sreenivasan et al., 2016), cependant la structure de l'ARN en complexe transcriptionnel n'a jamais été analysée par cette méthode. Cette absence d'étude structurale est en partie expliquée par l'impossibilité d'introduire un nucléotide fluorescent par l'ARNp d'*E. coli* (observation personnelle du laboratoire).

Dans cette deuxième publication, nous avons développé une nouvelle méthode pour obtenir un complexe transcriptionnel fonctionnel fluorescent et analyser la dynamique structurale de l'ARN en smFRET. Cette méthode repose sur la transcription par étape pour incorporer spécifiquement un nucléotide modifié qui sera par la suite marqué avec un fluorophore. Le marquage de l'ARN peut s'effectuer en présence de l'ARNp grâce à la chimie click sans cuivre (Agard et al., 2004). Comme pour le premier article, nous avons formé notre complexe transcriptionnel grâce au barrage biotine-streptavidine et celui-ci permet également de lier le complexe transcriptionnel sur une lame de microscope biotinylée.

Cette méthode nous a permis d'analyser la dynamique de repliement d'un autre riborégulateur sensible au TPP (*tbpA*) en complexe transcriptionnel. En arrêtant l'ARNp au niveau de ses pauses naturelles, nous avons identifié une structure de type Anti-P1 qui empêche la liaison subséquente du ligand au riborégulateur. Enfin, nous avons été capables d'analyser l'effet de la cinétique de transcription sur le repliement du riborégulateur. Dans ce contexte, nous avons observé que l'efficacité de liaison du ligand était inversement proportionnelle à la vitesse de transcription, définissant ainsi une fenêtre de liaison du TPP. Cette fenêtre est également délimitée par une pause située au début de la région codante du gène contrôlé par le riborégulateur.
Cet article soulève l'hypothèse qu'il existe une compétition entre la formation de l'Anti-P1 et la liaison du ligand, coordonnée par la vitesse de transcription. En effet, des expériences de transcription en temps réel sur lame de microscope viennent conforter cette idée en montrant une dispersion dans l'événement de liaison du TPP.

Patrick St-Pierre, Cibran Perez-Gonzalez et Carlos Penedo ont analysé les résultats de microscopie. Jonathan Grondin a effectué toute l'analyse sur le transcrit synthétique. Jean-François Nadon a participé au développement de la transcription par étape ainsi qu'à la méthode de marquage utilisant la chimie click. Maxime Simoneau-Roy a réalisé les expériences *in vivo* utilisant un gène rapporteur. Pierre Turcotte et Mélanie Geffroy ont développé et effectué les expériences de traduction *in vitro*. J'ai mis au point la méthode de transcription par étape et réalisé toutes les autres expériences en plus de rédiger le manuscrit et de faire les figures avec les Pr. Daniel Lafontaine et Carlos Penedo. L'ensemble du laboratoire Lafontaine a participé à la rédaction de ce manuscrit.

2.2 Manuscrit

A. Chauvier, P. St-Pierre, J.F. Nadon, J.P. Grondin, P. Turcotte, M. Geffroy, C. Perez-Gonzalez, S.H. Eschbach, M. Simoneau-Roy, A.M. Lamontagne, E. Massé, J.C. Penedo and D.A. Lafontaine (2017). Real-time single-molecule imaging of riboswitch regulatory dynamics during transcription elongation. (Science, in submission process)

Real-time single-molecule imaging of riboswitch regulatory dynamics during transcription elongation

Authors: A. Chauvier[1], P. St-Pierre[1]†, J.F. Nadon[1]†, J.P. Grondin[1], P. Turcotte[1], M. Geffroy[1], C. Perez-Gonzalez[2], S.H. Eschbach[1], M. Simoneau-Roy[1], A.M. Lamontagne[1], E. Massé[4], J.C. Penedo[2,3]* and D.A. Lafontaine[1]*

Affiliations:

[1]Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Quebec, Canada, J1K 2R1.

[2]Laboratory for Biophysics and Biomolecular Dynamics, SUPA School of Physics and Astronomy, University of St. Andrews, St Andrews, UK.

[3]Laboratory for Biophysics and Biomolecular Dynamics, Biomedical Sciences Research Complex, School of Biology, University of St. Andrews, St. Andrews, UK.

[4]Department of Biochemistry, Faculty of Medicine and Health Sciences, RNA Group, Université de Sherbrooke, Quebec, Canada, J1E 4K8.

*Corresponding authors. E-mail: jcp10@st-andrews.ac.uk and daniel.lafontaine@usherbrooke.ca

†These authors contributed equally to this work.

Abstract:

Cotranscriptional regulation ensures the timely and selective control of many biological processes. Riboswitches are cotranscriptional regulators that modulate gene expression by sensing cellular metabolites. We describe a single-molecule fluorescent approach to monitor the structural dynamics of the *Escherichia coli tbpA* nascent riboswitch transcripts within bacterial elongation complexes. By imaging nascent transcripts, we demonstrate that transcriptional pause sites are specifically tailored to favor transient metabolite sensing within a narrow transcriptional window located upstream of the translation start codon. Real-time monitoring of *E. coli* RNA polymerase transcription reactions reveals that the first metabolite-RNA binding event is only efficiently performed within a sensing hotspot, which is associated to cotranscriptional folding. These results provide a molecular mechanism describing how transcription elongation coordinates nascent transcript folding and metabolite sensing for regulation. The versatility of our approach suggests that it could be used to study nascent or elongating transcripts in other prokaryotic or eukaryotic transcription machineries.

One Sentence Summary:

Single-molecule imaging of nascent transcripts in elongation complexes reveals the molecular basis for riboswitch metabolite sensing and gene control.

Main Text :

Cotranscriptional RNA folding is involved in coordinating numerous regulatory steps (*1–3*), ranging from bacterial transcription termination (*4*) to eukaryotic mRNA splicing (*5*). Metabolite-sensing riboswitches are prime examples of cotranscriptional RNA regulators that directly coordinate metabolite sensing and regulation by undergoing cotranscriptional folding (*4*, *6*). Unfortunately, due to the transient nature of nascent transcripts, the impact of cotranscriptional RNA folding on genetic regulation is particularly difficult to study (*1*, *7*).

Single-molecule Fluorescence Resonance Energy Transfer (smFRET) is unique in monitoring specific conformational dynamics using strategically positioned fluorescent reporters (*8*). However, since bacterial and eukaryotic RNA polymerases (RNAP) do not incorporate fluorescent nucleotides during transcription elongation (*9*, *10*), smFRET studies attempting to characterize cotranscriptional RNA folding have been restricted to the use of synthetic constructs or bacteriophage T7 RNAP-made fluorescent transcripts (*8*). Such approaches intrinsically do not allow to study cotranscriptional RNA folding in a native context [supplementary materials section 2.1, (SM 2.1)](*3*), which depends on RNAP pausing, elongation rates and protein-RNA contacts (*11–13*).

Here, we describe a novel approach allowing smFRET analysis of cotranscriptional RNA folding. In contrast to previous smFRET studies (*14–16*), our approach allows to monitor in real-time the folding of nascent transcripts within *Escherichia coli* elongation complexes. We applied this method to study the thiamin pyrophosphate (TPP) *tbpA* riboswitch from *E. coli* (fig. S1A), which controls the synthesis of thiamin and TPP transporters (fig. S1B) (*17*) and for which *Salmonella typhimurium* antibiotic tolerance mechanisms have been predicted (*18*).

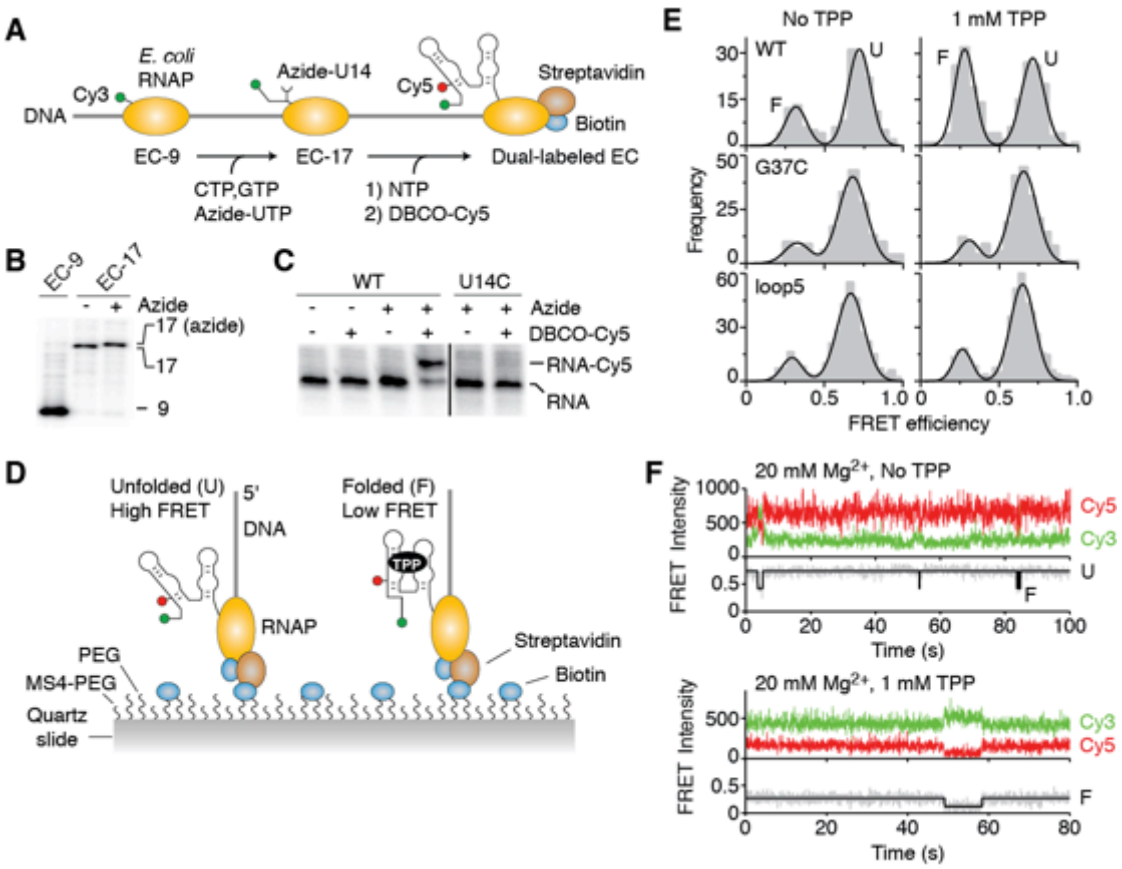


Fig. 1. smFRET analysis of *tbpA* ECs.

(A) Scheme describing the approach to obtain Cy3-Cy5 dual-labeled ECs. While Cy3 is introduced by initiating transcriptions with a Cy3-GUU trinucleotide, the Cy5 is incorporated at position U14 through a copper-free click reaction. **(B)** PAGE separation of transcripts corresponding to EC-9 and EC-17 obtained with UTP (-) or azide-UTP (+). **(C)** PAGE separation of WT and U14C transcripts obtained with UTP (-) or azide-UTP (+) reacted with or without DBCO-Cy5. **(D)** Scheme depicting labeled ECs immobilized to PEG through the transcriptional roadblock. **(E)** smFRET histograms of EC-88 WT, G37C and loop5 mutants. **(F)** smFRET-time traces of EC-88. A Cy5 blinking step is occurring at ~50 s with TPP, resulting in E ~0.15.

The first step of our approach (Fig. 1A) consists in initiating transcription using a fluorescent trinucleotide (Cy3-GUU) and a subset of nucleotides (nt) (fig. S2A), yielding Cy3-labeled elongation complexes stalled at position 9 (EC-9). The second step, resulting in EC-17 formation (fig. S2B), allows the incorporation of a single azide-UTP

analog at U14 (Fig. 1B). The last step involves transcribing complexes to a biotin-streptavidin roadblock (fig. S2B) (*19*) and coupling DBCO-Cy5 specifically to the U14 azide group through a copper-free click reaction (Fig. 1C and SM 2.2) (*20*). Importantly, RNase H assays reporting on TPP binding affinity (K_{switch}) (SM 2.3) showed less than a 2.5-fold effect on TPP sensing from either Cy3 or Cy5 dye (fig. S3, A-D).

Fluorescently-labeled elongation complexes, comprising *tbpA* aptamer transcripts (88 nt), were immobilized via the streptavidin roadblock on a biotin-functionalized polyethylene glycol (PEG) surface (Fig. 1D)(*21*). Specific formation, immobilization and labeling of intact EC-88 was confirmed using total internal reflection fluorescence microscopy (fig. S4, A-D and SM 2.4). In the absence of TPP, smFRET analysis of EC-88 showed a major high-FRET population (E ~0.6) and minor contribution from a low-FRET state (E ~0.3), which were assigned to unfolded (U) and folded (F) states, respectively (Fig. 1E, WT and SM 2.5). Addition of 1 mM TPP shifted the equilibrium toward the F population (Fig. 1E, WT). Hidden Markov modeling revealed very brief excursions to the F state without TPP (Fig. 1F and fig. S5, A and B), suggesting that *tbpA* nascent transcripts can transiently adopt a folded-like conformation in the apo state. Mutants preventing ligand binding (G37C and loop5, fig. S1A) did not fold in the presence of TPP (Fig. 1E and fig. S6, A and B) (*22*, *23*). Very similar smFRET profiles to EC-88 WT were obtained from run-off transcripts (fig. S7) or semi-synthetic RNAs (fig. S8), consistent with the *tbpA* 88-nt RNA efficiently responding to TPP.

Riboswitch TPP binding was investigated within ECs stalled at previously described pause sites 104, 117 and 136 (Fig. 2A) (*19*). RNase H assays revealed large variations in TPP affinity among stalled complexes (>70-fold) (Fig. 2B, fig. S9), with EC-136 showing the lowest affinity value (9.9 ± 0.4 μM) (table S1). We reasoned that inefficient TPP sensing could result from anti-P1 stem formation in EC-136 (Fig. 2C), similarly to the *thiC* riboswitch (*19*). Destabilization of the anti-P1 stem rescued TPP affinity by more than 30-fold (Fig. 2B, anti-P1 mutant), confirming the antagonistic role of the anti-P1 stem in TPP sensing.

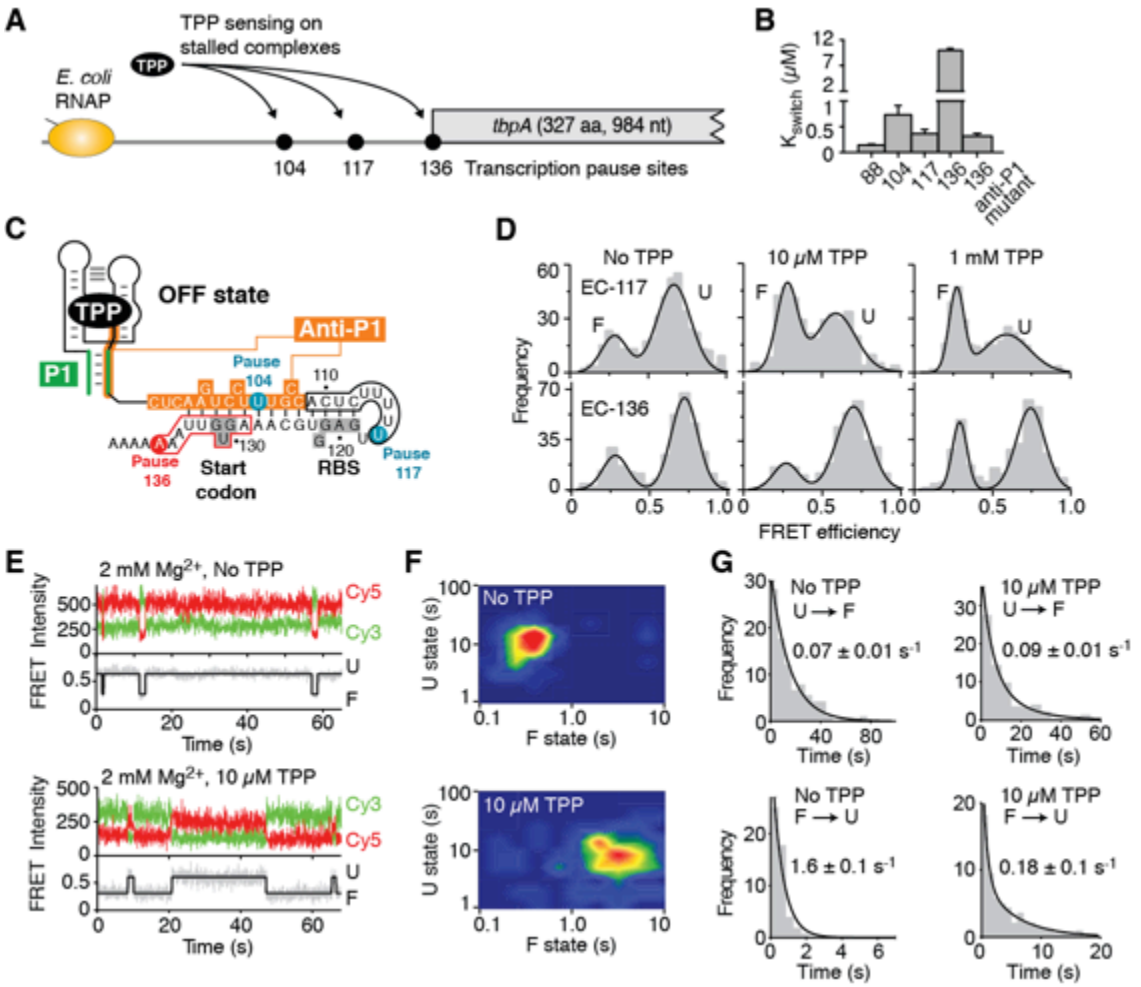


Fig. 2. TPP sensing at transcriptional pause sites.

(A) Genetic context of the *tbpA* riboswitch and transcriptional pause sites. **(B)** TPP binding affinity measured by RNase H assays for various ECs. **(C)** Predicted secondary structure of the TPP-bound riboswitch. Sequences within the RNA-DNA hybrid are shown for pauses 117 (black) and 136 (red). **(D)** smFRET histograms of EC-117 and EC-136. **(E)** smFRET-time traces of EC-117 at 2 mM Mg^{2+}. **(F)** 2D contour plots of U and F states at 2 mM Mg^{2+}. **(G)** Dwell time histograms of the U F and F U transitions at 2 mM Mg^{2+}. Exponential decay values are indicated for each condition.

We next used smFRET assays to explore the structure of nascent transcripts at *tbpA* transcriptional pause sites. Nascent transcript folding was observed in EC-104 and EC-117 with TPP (Fig. 2D, top and figs. S10 and S11). Although no effect was observed for

EC-136 at 10 μM TPP, riboswitch folding occurred at 1 mM TPP (Fig. 2D, bottom and fig. S12), in agreement with K_{switch} values (Fig. 2B). In contrast to EC-136, efficient TPP sensing in EC-117 could be due to RNAP preventing the anti-P1 stem through steric hindrance (Fig. 2C). Strikingly, a 117-nt nascent run-off transcript showed both a very low binding affinity (fig. S13A) and an absence of TPP-induced F state (fig. S13, B-D). These results indicate that RNAP plays a central role in TPP sensing in EC-117 by precluding anti-P1 stem formation.

smFRET kinetic analysis for all stalled complexes showed very rare transitions between U and F states (figs. S5 and S10 to S12), suggesting highly static conformers. Under conditions allowing RNA structural dynamics (2 mM Mg^{2+}) (*23*), EC-117 showed rare transitions from the U to the F conformer without TPP (Fig. 2E, top and fig. S14A). Addition of 10 μM TPP gave rise to frequent and long-lived excursions to the F state (Fig. 2E, bottom and fig. S14B), indicative of TPP binding. smFRET kinetic analysis of the U to F interconversion indicated that TPP plays no role in the U→F transition (Fig. 2, F and G) but stabilizes the F state by decreasing ~10-fold the F→U transition rate (Fig. 2G). These results suggest that nascent riboswitches recognize TPP via a conformational capture mechanism in EC-117 (*23*).

Our data indicate that cotranscriptional TPP binding is performed in a narrow transcriptional window delimited by position 136 (SM 2.7). In support of this model, smFRET data of transcription complexes elongated halfway (0.5 kb) or at the end (~1 kb) of the *tbpA* gene showed inefficient TPP-induced folding (fig. S15, A-D). Furthermore, RNase H assays showed that although fast transcription of EC-136 (1 mM NTP) allowed poor cotranscriptional TPP binding (Fig. 3A, fig. S16, and table S2), lower transcription rate (50 μM NTP) increased binding affinity by >8-fold. No such effect was obtained when transcribing EC-88 lacking the anti-P1 stem (Fig. 3A and fig. S17). Additional evidence for *tbpA* cotranscriptional sensing was obtained from RNase H kinetic assays (fig. S18) and uncoupled transcription-translation *in vitro* assays (fig. S19).

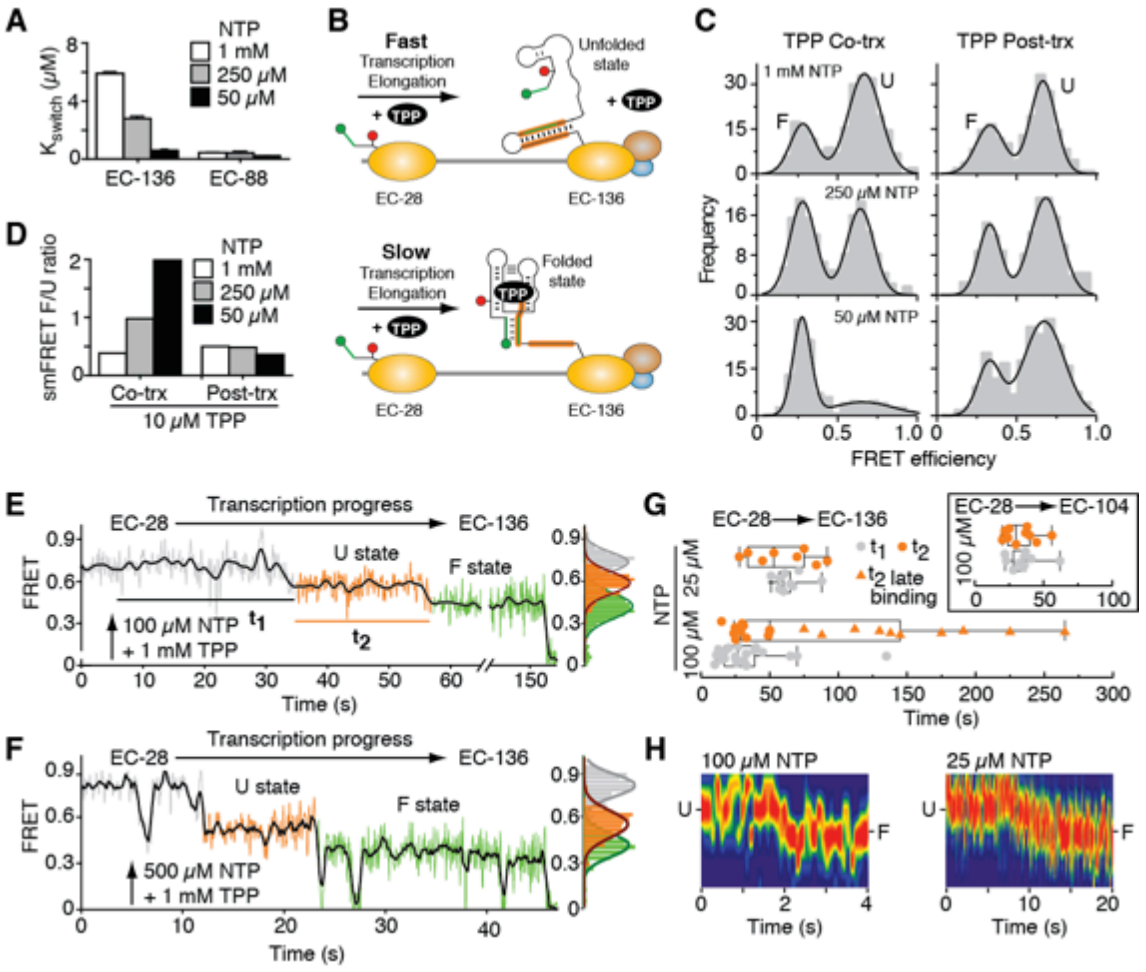


Fig. 3. Cotranscriptional TPP sensing by the *tbpA* riboswitch.

(A) Cotranscriptional TPP binding monitored by RNase H assays for EC-88 and EC-136 at various NTP concentrations. **(B)** Scheme representing the relationship between TPP binding and rate of transcription **(C)** smFRET histograms obtained after transcribing EC-28 EC-136 at various NTP concentrations when 10 μM TPP was added cotranscriptionally or post-transcriptionally. **(D)** Ratios of F to U states obtained from Fig. 3C. **(E, F)** Real-time monitoring of the cotranscriptional folding during transcription elongation using a concentration of 100 μM **(E)** or 500 μM NTP **(F)**. Upward arrows show injections at 5 s. A smoothing line was added using an adjacent-averaging algorithm with N=5 points. **(G)** Dwell-time analysis of t_1 and t_2 states obtained from real-time transcription reactions. **(H)** 2D contour plots of U F transitions observed for transcriptions done with 100 μM or 25 μM NTP.

To investigate the influence of transcription rate on riboswitch TPP-sensing (Fig. 3B), we engineered a Cy3-Cy5 labeled EC-28 (fig. S20 and SM 2.8), which was elongated to position 136 at different NTP concentrations. While stalled EC-28 complexes did not respond to TPP (fig. S21), efficient cotranscriptional TPP sensing was observed for EC-136 when transcribed using low NTP concentrations (Fig. 3C and fig. S22). Real-time imaging of transcription reactions revealed cotranscriptional trajectories sequentially transiting from EC-28 (E ~0.8), TPP-unbound (U) and TPP-bound (F) states (Fig. 3E and fig. S23A). No U F transition was detected without TPP (fig. S23B). Transcriptions performed near physiological conditions (500 μM NTP) (*24*) still yielded a similar three-state FRET profile (Fig. 3F and fig. S23C), indicating that sequential cotranscriptional folding is intrinsic to *tbpA* elongation complexes. Similarly to the adenine riboswitch (*25*), we observed a single metabolite binding event per transcription reaction, in agreement with riboswitch regulation being kinetically controlled. Importantly, in contrast to abrupt post-transcriptional events, real-time E~0.8 U and U F transitions were characterized by a gradual FRET decrease (fig. S24, A-C), suggesting that cotranscriptional folding takes place in the same timescale as transcription elongation (SM 2.10).

Kinetic analysis of cotranscriptional trajectories occurring between NTP injection and U state (Fig. 3E, t_1 and SM 2.9) revealed that the majority of EC-28 EC-136 exhibited an average E ~0.8 dwell-time of ~27 s (Fig. 3G, 100 μM NTP). The approximate RNAP location for U state formation was estimated between positions 83 and 88, as revealed from the absence of the U state in EC-83 (fig. S25A). From this estimation, we obtained an EC-28 EC-88 transcription rate of ~2.3 nt/s, which is only ~2-fold lower than the pause-free transcription rate (*25*). The very low efficiency of EC-83 to perform TPP sensing (fig. S25B) indicates that the upstream boundary for cotranscriptional TPP sensing is contained within the 5 nt window between EC-83 and EC-88, and that transcriptional pausing at position 83 (fig. S25C) might be important to coordinate upstream RNA folding to downstream TPP binding during RNAP elongation.

Dwell-times of the U state (t_2), denoting the first TPP association event to the nascent transcript, could be categorized into two subgroups: an 'early binding' group with t_2 values homogeneously averaging at ~28 s, and a 'late binding' group exhibiting a large spread centered at t_2 ~145 s (Fig. 3G, EC-28 EC-136, 100 μM NTP). Gradually decreasing U F transitions for the early binding group displayed a homogeneous timescale of ~1 s, which increased to ~5 s when using 25 μM NTP (Fig. 3H). In contrast, a variety of U F timescales were observed for the late binding group (fig. S26), suggesting TPP binding by elongating complexes having lower affinities. Both transcribing a shorter expression platform (EC-28 EC-104) or using a lower NTP concentration (EC-28 EC-136, 25 μM) resulted in an absence of t_2 heterogeneity (Fig. 3G), suggesting that heterogeneity is elongation rate-dependent and mostly localized between positions 104 and 136 (SM 2.10). Furthermore, the ~5-fold faster binding time (~11 s) of TPP to a stalled EC-117 complex (Fig. 2G), compared to the lowest t_2 late-binding value (~75 s) (Fig. 3G), suggests that late binding events are restricted to the 117-136 region, which constitutes the downstream boundary of the TPP binding window.

These experiments provide the molecular basis of nascent transcript folding and *tbpA* riboswitch TPP sensing within a narrow transcriptional window. The sequential cotranscriptional folding is intrinsically linked to the displacement rate of elongating complexes, ultimately affecting metabolite recognition. Upon ECs transiting into the sensing hotspot 88-117 nt region (Fig. 4), TPP sensing by nascent *tbpA* transcripts results in early and late binding events leading to riboswitch regulation, ensuring an irreversible decision to inhibit translation initiation. The method presented here may be used to study nascent RNA structures in several processes (SM 2.11), such as RNAP interactions with transcriptional factors (*15*), RNAP II promoter-proximal pausing (*26*) and mRNA splice site selection (*1*). The imaging of cotranscriptional folding using smFRET assays will provide a new and powerful approach to elucidate the role of RNA folding in the cell.

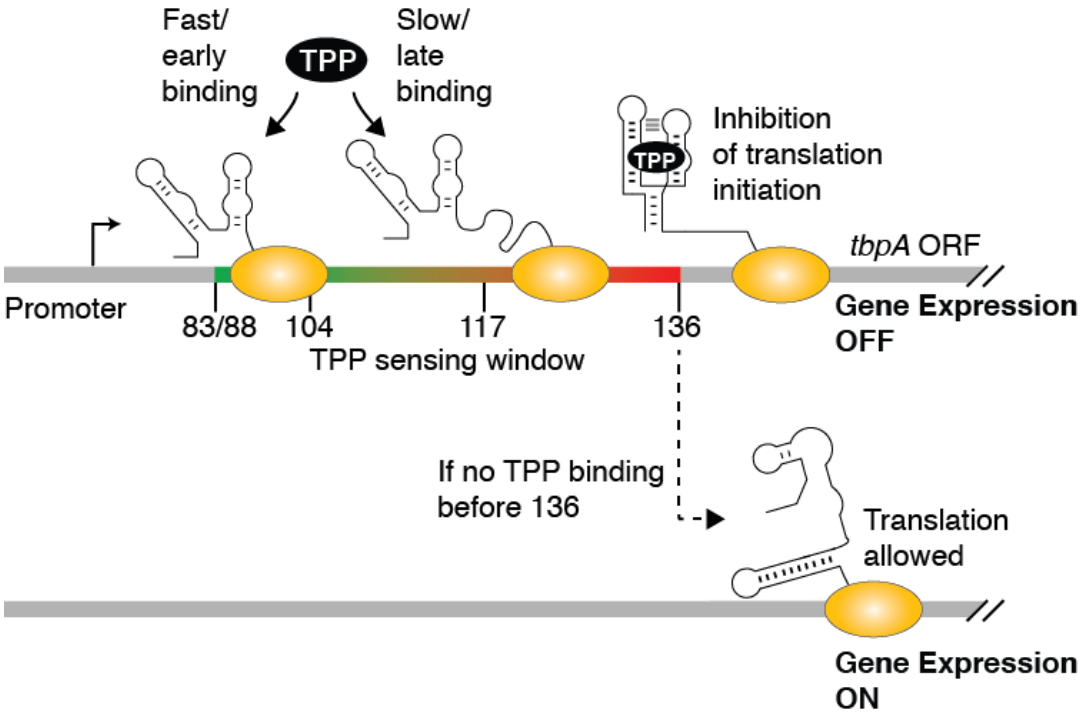


Fig. 4. Cotranscriptional TPP sensing and regulatory mechanism of the *tbpA* riboswitch. Nascent transcripts sense cellular TPP within a transcriptional window. The RNAP location dictates whether early or late TPP binding occurs, both of which ensuring cotranscriptional riboswitch folding and inhibition of translation initiation.

References and Notes :

1. R. Perales, D. Bentley, *Mol. Cell.* **36**, 178–91 (2009).
2. A. Ray-Soni, M. J. Bellecourt, R. Landick, *Annu. Rev. Biochem.* **85**, 319–347 (2016).
3. T. Pan, T. Sosnick, *Annu Rev Biophys Biomol Struct.* **35**, 161–175 (2006).
4. A. Serganov, E. Nudler, *Cell.* **152**, 17–24 (2013).
5. C. J. McManus, B. R. Graveley, *Curr. Opin. Genet. Dev.* **21**, 373–9 (2011).
6. A. Roth, R. R. Breaker, *Annu Rev Biochem.* **78**, 305–334 (2009).
7. K. M. Neugebauer, *J. Cell Sci.* **115**, 3865–71 (2002).
8. P. St-Pierre, K. McCluskey, E. Shaw, J. C. Penedo, D. A. Lafontaine, *Biochim Biophys Acta.* **1839**, 1005–1019 (2014).
9. D. G. Vassylyev *et al.*, *Nature.* **448**, 163–8 (2007).
10. K. D. Westover, D. A. Bushnell, R. D. Kornberg, *Cell.* **119**, 481–9 (2004).
11. D. Lai, J. R. Proctor, I. M. Meyer, *RNA.* **19**, 1461–73 (2013).
12. J. Zhang, R. Landick, *Trends Biochem. Sci.* **41**, 293–310 (2016).
13. R. Schroeder, A. Barta, K. Semrad, *Nat. Rev. Mol. Cell Biol.* **5**, 908–919 (2004).
14. J. Andrecka *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 135–40 (2008).
15. P. P. Hein *et al.*, *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
16. A. Chakraborty *et al.*, *Science.* **337**, 591–5 (2012).
17. D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, *J Biol Chem.* **277**, 48949–48959 (2002).
18. J. Yu, T. Schneiders, *BMC Microbiol.* **12**, 195 (2012).
19. A. Chauvier *et al.*, *Nat. Commun.* **8**, 13892 (2017).
20. M.-L. Winz, A. Samanta, D. Benzinger, A. Jäschke, *Nucleic Acids Res.* **40**, e78 (2012).

21. R. Roy, S. Hohng, T. Ha, *Nat Methods*. **5**, 507–516 (2008).

22. N. Kulshina, T. E. Edwards, A. R. Ferre-D'Amare, *RNA*. **16**, 186–196 (2010).

23. A. Haller, R. B. Altman, M. F. Souliere, S. C. Blanchard, R. Micura, *Proc Natl Acad Sci U S A*. **110**, 4188–4193 (2013).

24. M. H. Buckstein, J. He, H. Rubin, *J Bacteriol*. **190**, 718–726 (2008).

25. K. L. Frieda, S. M. Block, *Science*. **338**, 397–400 (2012).

26. X. Liu, W. L. Kraus, X. Bai, *Trends Biochem. Sci*. **40**, 516–525 (2015).

Acknowledgements:

We thank A. Lavigueur, B. Leblanc, S. Rodrigue and V. Steimle for discussion. This work was supported by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada. JCP wishes to thank the Scottish Universities Physics Alliance (SUPA) and the Engineering and Physical Sciences Research Council equipment grant support.

Supplementary Materials:

Materials and Methods

Supplementary Text

Figures S1 to S26

Tables S1 to S6

Full References

Supplementary Materials for

Real-time single-molecule imaging of riboswitch regulatory dynamics during transcription elongation

A. Chauvier[1], P. St-Pierre[1]†, J.F. Nadon[1]†, J.P. Grondin[1], P. Turcotte[1], M. Geffroy[1], C. Perez-Gonzalez[2], S.H. Eschbach[1], M. Simoneau-Roy[1], A.M. Lamontagne[1], E. Massé[4], J.C. Penedo[2,3]* and D.A. Lafontaine[1]*

[1]Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Quebec, Canada, J1K 2R1.

[2]Laboratory for Biophysics and Biomolecular Dynamics, SUPA School of Physics and Astronomy, University of St. Andrews, St Andrews, UK.

[3]Laboratory for Biophysics and Biomolecular Dynamics, Biomedical Sciences Research Complex, School of Biology, University of St. Andrews, St. Andrews, UK.

[4]Department of Biochemistry, Faculty of Medicine and Health Sciences, RNA Group, Université de Sherbrooke, Quebec, Canada, J1E 4K8.

*Corresponding authors. E-mail: jcp10@st-andrews.ac.uk and daniel.lafontaine@usherbrooke.ca

†These authors contributed equally to this work.

This file includes:

Materials and Methods

1.1 Bacterial strains and DNA oligonucleotides

Strains used in this study are derived from *Escherichia coli* MG1655. Strain phenotypes are listed in table S3. Strain BL21 (DE3) was used for overproduction of RNA polymerase (RNAP) and sigma70 factor. The PM1205 strain (table S3) was used to construct the *tbpA* translational and natural promoter fusion (table S4) The PCR strategy to obtain the *tbpA* riboswitch translational construct is listed in table S5. DNA oligonucleotides used in this study were purchased from Integrated DNA technologies (IDT) and are listed in table S6.

1.2 Beta-galactosidase experiments

Kinetic assays for beta-galactosidase experiments were performed as described previously (*19*, *27*). Briefly, an overnight bacterial culture grown in M63 0.2% glucose minimal medium was diluted to an OD_{600} of 0.02 in 50 mL of fresh medium. The culture was incubated at 37°C until an OD_{600} of 0.1 was obtained. TPP (500 μg/mL) was then added as indicated.

1.3 Preparation of dual-labeled transcription elongation complexes (EC)

The *Escherichia coli* RNAP polymerase (RNAP) and sigma70 factor were purified as previously described (*28*). DNA templates for transcription were produced by PCR using oligonucleotides containing the *lacUV5* promoter sequence (table S5). The formation of stalled ECs was ensured by using DNA templates containing a biotin at the 5'-end of the antisense strand. Streptavidin was mixed to a ratio of 5:1 with DNA (*19*) and added 5 min before transcription initiation. *In vitro* transcription reactions were performed in three steps to allow the specific incorporation of both Cy3 and Cy5 fluorophores at positions 1 and 14, respectively. Transcription reactions were done in a buffer containing 20 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 20 mM NaCl, 14 mM 2-mercaptoethanol and 0.1 mM EDTA. In Step 1 of our procedure, 200 nM DNA template, 600 nM sigma70 factor and 300 nM RNAP were incubated at 37°C for 5 min. Transcription reactions were initiated by adding 25 μM Cy3-GUU trinucleotide and 1.25 μM ATP/CTP/UTP nucleotides at 37°C for 10 min, thus yielding an elongation complex

stalled at position 9 (EC-9). The sample was next passed through G50 columns to remove any free nucleotides. In Step 2, reactions were incubated with 2.5 μM CTP/GTP and 2.5 μM azide-modified UTP analog at 37°C for 5 min to allow the formation of EC-17. Since a single uracil is found at position 14 in the 10-17 nucleotide region, it ensures that nascent transcripts carry a single azide at position 14. The Step 3 of the procedure was performed by using G50 columns to wash EC complexes, which were elongated by adding 1 mM NTPs at 37°C for 1 min, thus yielding stalled complexes at the biotin-streptavidin roadblock. Such stalled complexes were previously shown to be retained during extensive washing procedure (*19*).

The incorporation of the Cy5 fluorophore at the U14 azide group of nascent transcripts was performed by using copper-free strain-promoted click chemistry (*20*). The azide-carrying ECs were incubated with 63 μM DBCO-Cy5 (Jena Bioscience) in the transcription buffer described above for 1 h at 37°C. Transcription complexes were then washed twice using G50 columns to remove unreacted DBCO-Cy5 dyes. Dual-labeled ECs were collected in the transcription buffer and were incubated 5 min with TPP or water prior to be used for smFRET analysis.

1.4 Preparation of dual-labeled EC at position 28 (EC-28)

The preparation of EC-28 complexes was done using a modified DNA template in which a U23A mutation was introduced (SM 2.8). Briefly, following the preparation of EC-17 complexes in Step 2, elongation complexes were walked to position 28 using 2.5 μM ATP/CTP/GTP at 37°C for 3 min, thus resulting in EC-28 complex formation. The coupling of Cy5 with the U14 azide group was performed as described in the previous section. Dual-labeled nascent transcripts were washed and collected using the transcription buffer.

1.5 Preparation of dual-labeled run-off nascent transcripts

A DNA template was assembled corresponding to the first 117 residues of the *tbpA* riboswitch fused to an anchor sequence. Because the DNA template did not contain a biotin at the 5' end of the antisense strand, the additions of NTPs in Step 3 resulted in

run-off transcription and release of transcripts. The coupling of Cy5 with the U14 azide reacting group was performed as indicated above. Transcripts were then purified using G50 columns and hybridized to an anchor 5'-biotin DNA template (table S6), which allowed microscope slide immobilization and sm-FRET imaging.

1.6 Assembly of semi-synthetic fluorescent *tbpA* riboswitches

A semi-synthetic *tbpA* riboswitch was assembled using a modified version of a previously described strategy (*29*). Briefly, the 5' synthetic RNA molecule corresponding to the first 38 nt of the riboswitch was purchased from IDT (2558JG) that contained a Cy3 fluorophore at the 5' end and an azide reacting group at U14 (table S6). The azide group was reacted with DBCO-Cy5 as indicated above to obtain a dual-labeled RNA synthetic strand. The 3' transcript strand encompassing positions 39 to 88 fused to an anchor sequence was transcribed from a double-stranded DNA template (table S5, 3'tbpA-89) by using T7 RNA polymerase (*30*). The homogeneity of the transcript 5' end was ensured by inserting a hammerhead ribozyme sequence upstream of the riboswitch RNA molecule (*31*). The 5'-OH extremity of the 3' transcript was phosphorylated using the T4 polynucleotide kinase (NEB) in the presence of 10 mM ATP and incubated at 37°C for 30 min. In ligation control assays, the 3' transcript was phosphorylated using 100 nM γ-^{32}P [ATP]. The purified synthetic and T7 RNAP-made RNA strands were annealed by heating a mixture (molar ratio 1:1) to 80°C in 10 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$ and 100 mM KCl, which was slowly cooled to room temperature. T4 RNA ligase was added to the reaction and samples were incubated at 37°C for 4 h. Full-length ligated RNA molecules were purified by denaturing polyacrylamide gel electrophoresis and electroelution. The RNA was precipitated with ethanol and redissolved in water. The dual-labeled semi-synthetic *tbpA* riboswitch was hybridized to an anchor 5'-biotin DNA template (table S6, 2559JG), which allowed microscope slide immobilization and sm-FRET imaging.

1.7 Single-molecule imaging

Single-molecule FRET experiments were performed as previously described (*32*, *33*). smFRET traces were recorded from immobilized single elongation complexes using a prism-type total-internal reflection setup including an inverted microscope (Olympus IX71) coupled to a 532-nm laser (Crystalaser) and a back illuminated Ixon EMCCD camera (Andor). Microscope quartz slides were passivated with a 30:1 mixture of PEG (Laysan Biosciences, USA) (*34*) and biotinylated PEG (*35*). A second passivation round was performed for 30 min using 25 mM MS4-PEG (Thermo Scientific) diluted in bicarbonate buffer (pH 8.0) to limit protein-surface interactions (*34*). A concentration of ~250 pM dual-labeled ECs was added to the slide. For the imaging of run-off transcripts, 0.2 mg/mL streptavidin was added to the PEG slide to allow immobilization. Data were acquired using a 50 ms integration time on surface-immobilized molecules in the transcription buffer to which was added 2 mM Trolox, 5 mM 3,4-protocatechuic acid (PCA, Sigma) and 100 nM protocatechuate dioxygenase (PCD, Sigma), pH 7.5, which was used to enhance the photostability of the dye (*36*, *37*). Donor and acceptor fluorescence emissions were separated using dichroic mirrors (DCRLP, Chroma Technology) and imaged onto the left (donor) and right (acceptor) half-chip of the EMCCD camera. The setup allowed to record Cy3 and Cy5 signals simultaneously. Left and right images were corrected for optical aberrations and inhomogeneous evanescence-wave illumination using custom-build routines using IDL software (Exelis, USA). Subsequently, an IDL mapping algorithm was used to correlate the position of Cy3 and Cy5 signals from the same immobilized molecule and a time-dependent trajectory was built for each dye and for each molecule.

1.8 smFRET analysis of stalled ECs and run-off transcripts

Background corrected single-molecule Cy3/Cy5 intensity trajectories were transformed into FRET efficiency trajectories using ($I_A/[I_A+I_D]$), where I_A and I_D are the fluorescence intensities of donor and acceptor, respectively. Single-molecule FRET histograms were generated using the first 15 frames of each FRET trajectory and fitted by one or two Gaussian distributions depending on the specific conditions using Origin 8.0 (OriginLab

Corp, USA). The center of each Gaussian was taken as the FRET value for each particular state. Relative contributions of each FRET state in steady-state conditions were calculated from the ratio of areas under the corresponding Gaussian curve. For clarity, idealized Hidden Markov modelling (*38*) of each experimental FRET trajectory was carried out using the Hammy software package developed at the University of Illinois Urbana-Champaign and the idealized trace was superimposed to the experimental trajectory. Single-molecule kinetic analysis was performed at two levels: average rates for a given molecule and dwell-time histograms of cumulative transitions. Kinetic analysis was carried out using custom-built software in MATLAB (MatWorks, USA). Average rates were plotted as contour maps whereas the dwell-times of individual transitions were plotted as frequency histograms and fitted to an exponential decay function to extract the corresponding rates. To avoid bias towards fast transitions, for each molecule, each transition was weighted a factor considering the number of total transitions in the molecule. Unless state otherwise, all data was collected using an integration time of 100 ms per frame and a total number of frames ranging from 500 to 2000 depending on the dynamic range of the sample.

1.9 Real-time smFRET assays of cotranscriptional folding during transcription elongation

Real-time analysis was performed using a setup similar to the one described in section 1.7. However, minor modifications were made to the microscope slide to allow real-time injection during imaging. For instance, 40 mm cover slides were used to allow the fixation at one end of the slide channel of a yellow 200 μL pipette tip, which was cut to be used as a buffer reservoir. In addition, a 18G1 needle (BD) was fixed on the other side of the slide channel. A syringe was attached to a flexible tube connected to the needle to draw off buffer from the reservoir. For each experiment, 100 μL of 1:1 solution of transcription and imaging buffer was injected to fill the imaging channel. Injections in real-time assays were carried out at 5 s, unless stated otherwise, after movie data collection was started to have a pre-injection FRET level reference. Real-time transcription movies were taken for a sufficient long time window (>1000-3000

frames/movie depending on conditions) to ensure that Cy5, Cy3 or both signals were significantly photobleached by the end of the movie. The observation of single-step Cy5 or Cy3 photobleaching, together with no significant changes in the average total emission, were taken as evidence for fluctuations in FRET signal arising from single elongation complexes. Single-molecule FRET trajectories were calculated as indicated before and only those fulfilling the previous criteria of singleness were considered for data analysis. To help visualization, a smoothing line was added to the experimental FRET trajectory using an adjacent-averaging algorithm with N=5 points in the moving window.

1.10 RNase H probing assays of stalled ECs and run-off transcripts

Transcription reactions were done in 20 mM Tris-HCl pH 8.0, 20 mM $MgCl_2$, 20 mM NaCl, 14 mM 2-mercaptoethanol and 0.1 mM EDTA. The DNA template, sigma70 factor and RNAP were first incubated at 37°C for 5 min. Reactions were initiated by incubating 10 μM GUU initiator, 1.25 μM ATP/CTP nucleotides and [a-^{32}P] UTP at 37°C for 10 min, thus yielding EC-9. After passing samples through G50 columns to remove free nucleotides, transcription reactions were performed by adding 1 mM NTPs at 37°C for 5 min. Reaction aliquots were mixed with 20 μM of a DNA oligonucleotide (table S6, 2477AC) at 37°C for 5 min. RNase H cleavage was started by adding a solution of RNase H (0.12 U/μL) in 5 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$, 100 mM KCl, 50 μM EDTA and 10 mM 2-mercaptoethanol at 37°C for 5 min. Reactions were stopped by adding an equal volume of a stop solution (95% formamide, 20 mM EDTA and 0.4% SDS).

K_{switch} values of stalled ECs and native transcripts were obtained using a procedure similar to that used for T_{50} (*19*, *39*, *40*). Briefly, after transcription reactions are completed, TPP is added at various concentrations (1 nM to 1 mM) and incubated at 37°C for 5 min. RNase H assays are then performed as described above. The quantification assumes a simple 1:1 stoichiometry between the aptamer and TPP as expected from crystal structures of the corresponding thiM aptamer (*41*, *42*).

K_{switch} values obtained at various transcription elongation rates were performed by transcribing the corresponding EC with 1 mM, 250 μM or 50 μM NTP in the presence

of various TPP concentrations (1 nM to 1 mM) at 37°C for 5 min. RNase H assays were performed directly on the newly synthesized ECs (see above).

1.11 RNase H probing analysis of TPP binding kinetics

The kinetics of TPP binding (co- or post-transcriptional binding) was essentially performed as previously described (*19*). Completed transcription reactions were mixed with 200 μM DNA (table S6, 923AC) and RNase H for 15 s at various time points (15 s, 45 s, 90 s, 2 min and 3 min). TPP (10 μM) was either added co- or post-transcriptionally. The reported errors for the TPP-binding rates are the standard error in the fitting, which are assumed to be approximated by the standard deviation (*29*).

1.12 *In vitro* transcription-translation assays

The description of the PCR strategy to obtain DNA templates used for transcription-translation assays can be found in table S5. *In vitro* translation assays were performed as previously reported (*19*). Reactions were done using the PURExpress Kit from New England Biolabs. To verify that TPP does not generally affect transcription-translation efficiency, a DNA template containing the lacUV5 promoter fused to the first 934 codons of *lacZ* followed by a T7 terminator was used as an internal control. TPP was added at the final concentration of 25 μM and transcription was carried out with 0.5 U of the RNA polymerase holoenzyme from Epicentre. For *tbpA* uncoupled transcription-translation assays, reactions were initiated by a transcription step of the placUV5-$tbpA_{306cd}$ construct at 37°C for 15 min using a transcription solution without amino acids/transfer RNA. Rifampicin (250 μg/mL) was next added to the reaction and incubated for 1 min to prevent re-initiation of transcription. The translation step was initiated by incubating with a solution comprising amino acids and transfer RNA at 37°C for 4 h. TPP (100 μM) was either added during the transcription step (cotranscriptional) or the transcription step (post-transcriptional). Reactions were stopped by placing samples on ice for 10 min and next incubated with 4 volumes of acetone 100% at 4°C for 15 min. Samples were precipitated, resuspended in denaturing buffer and resolved on SDS-PAGE.

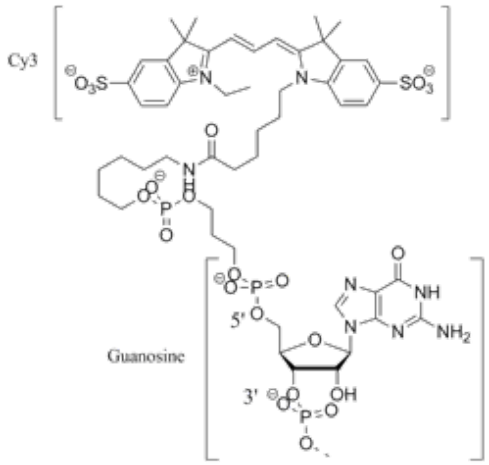
Supplementary Text

2.1 Importance of using the *E. coli* RNAP to study *E. coli tbpA* nascent transcripts

Cotranscriptional RNA folding inherently relies on the sequential and directional polymerization of transcripts emerging from RNAP, which may occur on the same timescale of RNA synthesis (*11*). As a result, the overall speed of transcription can affect transiently adopted structures as well as final transcript conformations (*11*, *43–45*), and ultimately the outcome of associated biological processes. The phage T7 RNAP proceeds five times faster than the *E. coli* RNAP (*46*) and has been shown in several studies to alter the folding of RNA. For example, it was previously demonstrated that T7 RNAP-driven *in vivo* transcription of *rrnB* operon did not lead to the formation of active 70S ribosomes (*47*), suggesting that the ribosome assembly process requires a fine adaptation of the transcription rate of rRNAs and the assembly process. The use of T7 RNAP was also shown to yield inactive RNAII as a replication primer (*48*), to introduce new kinetic traps in the *Bacillus subtilis* RNase P RNA (*49*) and to uncouple metabolite binding from AdoCbl riboswitch folding (*50*). Cotranscriptional folding was also demonstrated to rely on RNAP pausing whereby the discontinuous transcription elongation process dictates the structural outcome of emerging nascent RNAs (*51*). Similarly to the speed of transcription, the identity of the polymerase was shown to be important for RNAP pausing (*52*). For example, while the *E. coli* RNAP efficiently responds to the *his* class-I pause site, the *Bacillus subtilis* RNAP shows complete inability to pause at the *his* site, suggesting the *E. coli* and *B. subtilis* RNAP profoundly differ in their recognition of regulatory sequences (*52*). Thus, these results clearly indicate that the study of cotranscriptional RNA folding needs to be performed using the native RNAP (*3*), i.e., from the same organism as the investigated RNA molecule. Consequently, it is expected that the cotranscriptional folding of the *E. coli tbpA* riboswitch should yield biologically relevant information when transcription reactions are performed using the *E. coli* RNAP.

2.2 Incorporation of fluorescent dyes within nascent transcripts.

During the course of our studies, several approaches were considered to introduce a pair of Cy3 and Cy5 dyes within nascent transcripts. Since *in vitro* transcription reactions rely on the use of polynucleotide initiators (*28*, *53*), we reasoned that the Cy3 dye could be introduced through the use of a Cy3-labeled GUU trinucleotide:



Chemical structure of the Cy3 and guanosine of the Cy3-labeled GUU trinucleotide initiator.

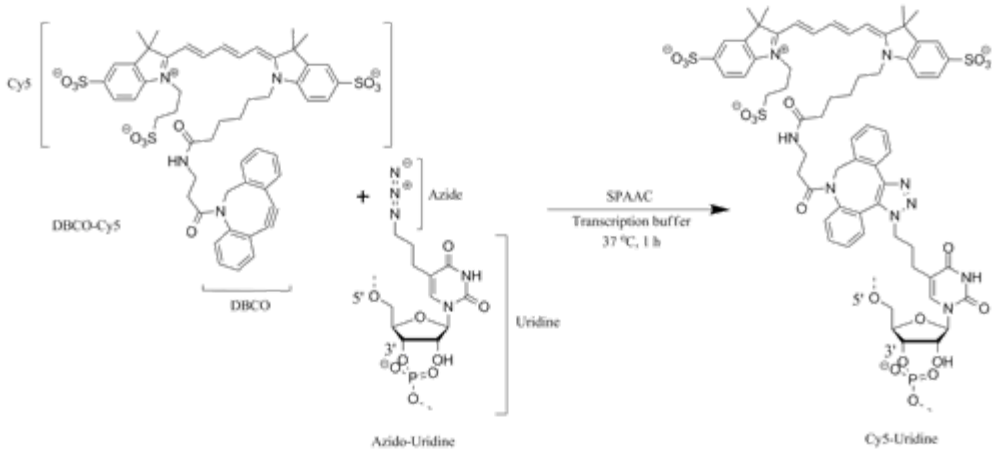
In our conditions, transcription reactions were found to be efficient when using the Cy3-labeled GUU initiator (fig. S2A). The introduction of the Cy5 dye was attempted by performing transcription reactions using the *E. coli* RNAP in the presence of Cy5-UTP. However, no successful Cy5-labeled transcripts were detected (data not shown). We thus reasoned that specific internal labeling of the nascent RNA was required without the direct incorporation of fluorescent nucleotides.

Since click type reactions require simple conditions and a minimum of purification, they constitute an excellent option for the labeling of biological systems (*54*). Most notorious of all click reactions is the copper catalyzed alkyne-azide cycloaddition (CuAAC) (*55*, *56*). CuAAC has already extensively been used for the labeling of biological components. However, some problems concerning effects of the reaction on nucleic acids and/or proteins (components of transcriptional complexes) have previously been observed (*57–60*). Notably, copper oxidizes the imidazole found in histidine side chains

(*57*) and reacts with oxygen to generate radicals which can lead to the degradation of nucleic acids (*58*). Furthermore, the catalytically active form used in CuAAC (copper(I)) is insoluble in water and must be generated *in situ* by reduction using sodium ascorbate, which was shown to disturb non-covalent interactions with proteins (*60*). In addition, dehydroascorbate generated by the click reaction has been shown to react with arginine residues in proteins (*59*).

Copper-free strain-promoted azide-alkyne cycloaddition (SPAAC) (*61*) does not require any catalyst, can be carried out in biological conditions, and therefore represents a more attractive and gentle approach to labeling (*61*). SPAAC has already been shown to be efficient in the labeling of proteins (*62*), DNA (*63*) and RNA, both *in vitro* (*3, 41*) and *in vivo* (*41, 42*). For these reasons, SPAAC presented a more suitable option for the labeling of nascent RNA within an elongation complex.

Since a DBCO group is bulkier than an azide reacting group, it is likely that the *E. coli* RNAP would incorporate the latter with more ease. In addition, since the incorporation of azido-UTP into RNA by T7 and human RNAP have already been described (*41, 43*, *44)*, we hypothesized that it would be possible using the *E. coli* RNAP. In this reaction, a DBCO-Cy5 is coupled to an azido-UTP through SPPAC reaction thus yielding a Cy5-labeled uridine:



Chemical structure of azido-uridine, DBCO-Cy5, and the resulting Cy5-labeled uridine.

Our results showed that the SPAAC reaction was efficiently performed (Fig. 1C), thus enabling to obtain Cy5-labeled transcripts. Importantly, when SPAAC was carried using a transcript not containing a azide reacting group, no gel retardation of the transcript was observed in the presence of DBCO-Cy5 (Fig. 1C). No labeling was obtained when attempting SPAAC using a mutant template (U14C) that does not allow azido-UTP incorporation (Fig. 1C), suggesting that no significant misincorporation of the azide group occurs at other transcript positions.

2.3 Identification and validation of labeled positions for smFRET assays

Identification of fluorophore positions: FRET is a powerful technique that provides inter-fluorophores distance information from 2 to 10 nm (*21*, *68*), therefore allowing to characterize conformational changes of biomolecules. However, important considerations have to be taken into account when introducing fluorophores as they could affect the folding or activity of studied biomolecules. Although it is theoretically possible to incorporate fluorophores at any uracil position in the riboswitch sequence using stepwise transcription and click chemistry reaction (SM 2.2), practical limitations such as the RNAP reinitiation efficiency during the stepwise procedure constraints the introduction of fluorophores in the 5' region of the characterized RNA. We explored the ability of RNAP to use a Cy3-GUU initiator in transcription reactions (fig. S2A), thus placing the Cy3 dye at G1. The position to introduce the Cy5 dye was selected based on crystal structures of the related *thiM* aptamer (*41*, *42*, *69*). The structure shows the presence of an interaction (P3-L5) occurring between the stem P3 and the loop 5, which is stabilized upon TPP binding (*22*, *23*, *70*). As a result, we reasoned that formation of the P3-L5 interaction in presence of TPP should be detected by introducing a dye in close proximity to the P3 stem. Based on these considerations and on the fact that an azide-UTP analog was used for click reactions, the residue U14 (fig. S1A) was identified as a suitable position to introduce Cy5 within the riboswitch sequence. Importantly, the choice for the exact position of internal labeling was directed by the need to reduce the number of transcription steps during the stepwise approach as a way to obtain sufficient amount of materials for subsequent analysis. In addition, the identified position for Cy5

introduction needed to support efficient click reaction with DBCO-Cy5 to obtain dual-labeled Cy3-Cy5 RNA molecules. Indeed, in the course of our study, we observed for some positions that the efficiency of click reaction can dramatically differ depending on the RNA structure. However, efficient Cy5 labeling was observed for position U14, thus allowing smFRET analysis.

Validation of labeled residues for smFRET assays: For the Cy3-Cy5 FRET vector to be meaningful and report on biologically relevant conformational changes, the presence of both dyes should not perturb the folding or activity of the riboswitch. In contrast to *Bacillus subtilis* riboswitches that modulate transcription terminators, *E. coli* riboswitches regulate the initiation of translation (*71*). As a result, while the activity of *B. subtilis* riboswitches may be studied using single-round *in vitro* transcription assays, the characterization of the activity of *E. coli* riboswitches represents a more technically challenging approach as it requires to reconstitute coupled transcription-translation assays allowing to monitor the metabolite-dependent regulation of translation initiation. As an effort to develop a tractable approach to study the ligand-dependent activity of translationally controlling riboswitches, an RNase H-based assay was employed to monitor the ability of nascent riboswitches to undergo metabolite-dependent structural changes (*19*, *50*, *72*, *73*). In this assay, specific riboswitch regions are targeted using DNA oligonucleotides in conditions of increasing ligand concentrations. RNase H cleavage efficiencies are modeled to reveal the ligand-dependent riboswitch conformational change, thus yielding an estimation of ligand binding affinity (K_{switch} values) (*19*, *73*). Using this approach, we determined that singly-labeled *tbpA* riboswitches carrying a Cy3 at the G1 exhibited a very similar K_{switch} value (table S2, 218 nM ± 25 nM) compared to unlabeled riboswitches (227 nM ± 13 nM), indicating a negligible effect of Cy3 on riboswitch activity. Similarly, a small ~2.5-fold effect of the Cy5 dye at U14 (K_{switch} = 521 nM ± 72 nM) was observed from RNase H assays, indicating that the riboswitch TPP-sensing activity is not strongly altered by the presence of Cy5. Therefore, we conclude that the G1-Cy3/U14-Cy5 FRET vector used in this study is well suited to monitor riboswitch conformational changes upon ligand binding without compromising riboswitch function.

2.4 Validation of Dual-Labeled Cy3-Cy5 Elongation Complexes for smFRET Analysis.

Requirements for smFRET detection: Single-molecule fluorescence detection requires the spatial separation of the immobilized complexes so they can be assigned as individual structures. In addition, it requires a sufficiently high signal-to-noise ratio (SNR) to detect the low number of photons available from single-dyes. To achieve such conditions, immobilization specificity and low background—both from impurities and non-specifically adsorbed dyes—are of paramount importance. Moreover, it is known that fluorescently labeled proteins might exhibit an even higher level of non-specific interactions with the glass surface compared to unlabeled complexes (*74*). As a result, our stepwise approach to obtain Cy3-Cy5 labeled nascent transcripts has been developed and optimized to allow Cy3 and Cy5 specific labeling prior to immobilization on the dual-PEGylated quartz slide. Notably, a reverse approach consisting in immobilized ECs on the slide followed by dye labeling would compromise SNR due to the high concentrations of dye required for efficient labeling. Such an approach would require the use of laborious and expensive waveguide-based techniques to allow efficient imaging. In contrast, our two-step approach is efficient and easy-to-implement in most labs with conventional molecular biology kits and with standard single-molecule capabilities.

Validation of the method for visualization: Control experiments were carried out at the single-molecule level to test labeling and immobilization specificity. Single-molecule images of Cy3-labeled intact elongation complexes showed ~ 190 molecules per field of view (70 μm x 70 μm) (fig. S4B). Immobilized complexes displayed a single-step photobleaching indicative of the presence of a single Cy3 dye per fluorescence spot (fig. S4A). No immobilization was observed when using EC complexes lacking the biotin group on the DNA template or the anchoring streptavidin protein (<1 spot per field of view) (fig. S4B). For dual-labeled complexes, Cy3 and Cy5 fluorescent signals were detected in their corresponding channels only in conditions allowing successful bioconjugation (fig. S4C, top panel). Complexes subjected to click reaction but lacking the azide reacting group only showed Cy3 fluorescence (fig. S4C, middle panel),

consistent with specific Cy5 labeling. Similarly, complexes where the reactive azide-UTP was prevented to be incorporated at position 14 (U14C mutant) did not show significant levels of Cy5 fluorescence (figs. S4C, bottom panel and S4D, bottom panel). Together, our control experiments indicate that the Cy5 fluorophore is only incorporated at the U14 residue, thus ensuring the presence of a single Cy3-Cy5 pair per elongation complex.

2.5 Assignment of Folded (F) and Unfolded (U) smFRET populations.

Folded (F) state: Although smFRET is a powerful assay to monitor time-dependent structural changes, the use of a single FRET Cy3-Cy5 vector does not always allow the straightforward assignment of FRET populations to known RNA structures. In our study, smFRET assays of EC-88 showed that the *tbpA* riboswitch exhibits two FRET states corresponding to values of E ~0.3 and E ~0.6. The fact that the addition of TPP yielded an increase of the E ~0.3 population suggests that the latter corresponds to the folded TPP-bound structure (F state). Since a minor proportion of the F state is observed in the absence of TPP, it indicates that nascent transcripts can adopt an F state-like structure in the apo form, as reported for other riboswitches (*23*, *75*). Although no *tbpA* aptamer crystal structure is yet available, the *thiM* riboswitch variant exhibits a similar secondary structure and has been crystallized (*41*, *42*, *69*). From the structure, it was deduced that the bound TPP interacts with the residue G37 and is involved in the stabilization of the P3-L5 tertiary interaction. In agreement with these studies, our smFRET data showed that the TPP-induced F state is not observed in the context of a G37C mutant (perturbing ligand binding) or loop5 mutant (preventing the P3-L5 tertiary interaction). Interestingly, since the last ~8-9 transcribed residues are sequestered by the RNA-DNA hybrid of the elongation complex (*76*), it suggests that the P1 stem is not formed in the context of EC-88. It was previously shown for the *thiM* aptamer that TPP sensing is strongly reduced when the 5' and 3' strands of the P1 stem were removed (*70*), suggesting that the presence of the P1 stem is important for metabolite binding. However, since *thiM* crystal structures showed that TPP-RNA interactions are primarily performed through the P2-P3 and P4-P5 peripheral structural elements (*41*, *42*, *69*), it

suggests that the role of the P1 stem consists primarily in restraining the conformational search of P2-P3 and P4-P5 stems for TPP sensing (*23*, *70*), and not to perform direct metabolite contact. This is in agreement with TPP recognition by the two helical domains (P2-P3 and P4-P5 stems) being significantly faster than P1 stem folding for *thiM* (*70*), suggesting that P1 formation is not involved in initial TPP sensing. As a result, TPP binding is expected to occur, albeit with a reduced affinity, in the absence of P1 stem. The strong deficiency observed for the TPP binding in the context of a shortened *thiM* P1 stem could be explained by misfolding of the P2 stem caused by the shortening of the 5' strand P1 stem (*70*).

The existence of similar F states across ECs (88, 104, 117 and 136) suggests that associated TPP-bound conformers adopt similar structures. Given that the P1 stem is most probably not folded in EC-88 (see above), it suggests that the spatial distance between Cy3 (G1) and Cy5 (U14) fluorophores is not sensitive to P1 folding. This apparent lack of structural sensitivity could be explained by the Cy3-Cy5 inter-distance not being largely modulated upon P1 formation. Notably, FRET populations similar to E ~0.3 were observed in a wide range of constructs, including EC-28 and EC-83 (fig. S21 and S25). In the case of EC-28, which most probably exhibits ~14 nt outside of the RNAP exit channel, it suggests that the adoption of a conformation similar to the F state is an intrinsic property of the 1-14 nt region. Given the inherent dynamics of the structurally similar *thiM* P1 stem (*23*), it suggests that the P1 5' strand may attain a conformational state closely related to that observed in the TPP-bound state, a feature that could be an integral part of TPP sensing.

Unfolded (U) state: Very little structural information is currently available about the apo form of the aptamer as no crystal structure has been solved in the absence of TPP. In the absence of TPP, it is expected that the P3-L5 interaction is not stably formed in the context of the related *thiM* aptamer (*22*, *23*, *77*). In the apo form, our smFRET data showed that the TPP-binding competent EC-88 complex exhibits a FRET value of ~0.6, suggesting a global conformation different from the TPP-bound complex (Fig. 1E). We have also analyzed the folding of *tbpA* nascent transcript in the context of EC-28 and EC-83 complexes, which revealed a high FRET value of ~0.8 for both complexes (fig. S21A and S25A). The high FRET ~0.8 state of EC-28 is most probably explained by a

relatively unstructured conformation, as recently showed for unstructured 15-nt single-stranded nucleic acids (*78*). Given that both EC-28 and EC-83 complexes do not show efficient TPP binding (fig. S21A and S25B), it suggests that the adoption of the U state in EC-88 is associated with the ability to perform efficient ligand binding. The FRET change observed between EC-83 and EC-88 likely results from the RNA adopting different conformers in EC-83 and EC-88 (SM 2.10). For example, while base pairing interactions could occur between residues G1-C4 and G50-C53 in EC-83, EC-88 could allow the formation of the P4 stem, which would lead to the formation of the pyrophosphate sensor helix involved in TPP binding (*41*, *42*, *69*). As a result, the U state conformation could exhibit a core domain important for TPP sensing, such as a P2/P4 scaffold suggested in the case of the *thiM* aptamer (*23*).

2.6 Optimization of RNA ligation to obtain semi-synthetic *tbpA* RNAs

By using an enzymatic ligation to assemble chemically synthesized RNA fragments, it is possible to incorporate modified nucleotide internally at specific locations on the riboswitch. For our experiments, the enzymatic ligation was completed using the T4 RNA ligase (NEB) (*79*), which is an ATP-dependent enzyme that catalyzes the formation of phosphodiester bonds between an acceptor oligonucleotide with a free 3'-hydroxyl and a donor oligonucleotide with a 5'-monophosphate. When using the T4 RNA ligase, it has been shown that some sites are more efficient for enzymatic ligation (*80*). The ligation sites efficiency is dependent of the 3' and 5'-terminal nucleosides. Usually, pyrimidines perform with a higher efficiency than the purine as a donor (providing the 5'-monophosphate) (pC > pU = pA > pG). Among acceptors, 3'-terminal adenosines are preferred over cytidine and guanosine, and the uridine residue is the poorest acceptor. While the efficiency of the ligation will depend on the 3' and 5'-terminal residue of the oligonucleotides, the most important factor is the RNA secondary structure. The donor and acceptor RNA should be able to assemble into a ligation competent complex that reconstitute the secondary structure of the ligated product (*79*).

In our experiments, we chose the ligation site to occur between the two adenosines at positions 39 and 40. The location of the ligation site allowed to use a chemically

synthesized 1-39 strand carrying a Cy3 dye at U1 and an azide reacting group at U14. Thus, by performing a click reaction with DBCO-Cy5, this approach allowed to obtain a dual-labeled Cy3-Cy5 strand. The 3' strand was obtained through T7 RNAP transcription of a hammerhead-3' strand fusion construct. A ligation competent complex was obtained by hybridizing the 5' synthetic and 3' transcript strands. Our results showed that this site allowed the efficient ligation between the acceptor and donor RNA (fig. S8B).

2.7 Implications of riboswitch transcriptional binding window for genetic regulation

Our study of the *tbpA* riboswitch indicated large variations in TPP sensing (>70-fold) where the lowest affinity was observed for the EC-136 complex (Fig. 2B). The inefficiency of binding in EC-136 was attributed to the presence of the anti-P1 stem since the TPP binding affinity was restored in a mutant destabilizing the anti-P1 stem (Fig. 2B). This situation is highly similar to that of the *thiC* riboswitch, which was showed to perform TPP sensing only within a narrow transcriptional window (*19*). The inefficiency of the *tbpA* riboswitch to recognize TPP at EC-136 (Fig. 2B) suggests that metabolite sensing is only performed within a transcriptional window in which the downstream boundary is delimited by position 136. The presence of a transcriptional window is consistent with the riboswitch not attaining thermodynamic equilibrium with TPP before RNAP reaching EC-136, as found for the *thiC* and *ribD* riboswitches (*19*, *81*). Therefore, the rate of transcription elongation and the kinetics of TPP binding are important for riboswitch regulation. As expected from such a kinetically driven mechanism, our K_{switch} data showed that the rate of transcription was important for TPP sensing where a slower transcription rate (50 μM NTP) allowed efficient TPP sensing (Fig. 3A). In agreement with this model, a TPP binding kinetic analysis showed that TPP binding was ~5-fold faster in a cotranscriptional context (fig. S18). The importance of TPP recognition by the riboswitch within the transcriptional window was also suggested from uncoupled transcription-translation *in vitro* assays, which showed that the TPP-dependent inhibition of translation initiation was only observed upon cotranscriptional

TPP addition (fig. S19). Thus, as found for other transcriptionally regulating riboswitches (*81*), these data support a model where the translationally regulating *tbpA* riboswitch modulates genetic regulation by binding TPP within a defined transcriptional window allowing efficient metabolite sensing.

2.8 Requirements for the use of EC-28 in real-time smFRET cotranscriptional assays

Although smFRET assays of stalled ECs give crucial information about the structure of nascent transcripts at defined pause sites (Fig. 2D and fig. 10A), no real-time data can be obtained from these experiments since the analysis is strictly performed at the post-transcriptional level. To monitor real-time cotranscriptional RNA folding, it requires a dual-labeled Cy3-Cy5 elongation complex positioned upstream of the region to be transcribed, which will be subsequently elongated upon the addition of NTP. After exploring different options within the RNA sequence, we obtained dual-labeled EC-28, which is ideally located to monitor the folding of the riboswitch downstream sequence. To obtain this fluorescent elongation complex, several criteria needed to be fulfilled. First, Cy5 labeling at U14 required RNAP to be positioned at a downstream location to allow the azide-U14 modified residue to be accessible for click reaction. Furthermore, to obtain sufficient amount of dual-labeled EC-28, we reduced the number of stepwise reactions by introducing an U23A mutation, thus allowing to perform EC-17 EC-28 (see sequence in fig. S2B). The U23A mutation was identified based on its non-conserved nature (*17*) and on crystal structures of the related *thiM* aptamer (*41*, *42*, *69*) showing that U23A should not perturb TPP sensing. Furthermore, although the making of EC-28 inherently requires stepwise reactions at position 17 and 28, we reasoned that RNAP pausing in this region should not strongly affect subsequent RNA folding since only a small portion of the riboswitch has been transcribed at that point. As expected, K_{switch} assays showed that transcribing EC-28 EC-136 at 1 mM or 50 μM NTP (figure S20, A and B) did not alter metabolite recognition compared to the EC-9 EC-136 wild-type sequence at corresponding NTP concentrations (fig. S16, A and C), thus justifying the use of EC-28 to study *tbpA* cotranscriptional folding. Importantly, EC-28 was found

to be highly competent for transcription elongation (fig. S20C) and for DBCO-Cy5 labeling at position 14 (fig. S20D).

2.9 Analysis and interpretation of real-time FRET traces.

Single-molecule FRET trajectories obtained from real-time transcription elongation complexes were analyzed to extract t_1 and t_2 values. t_1 represents the time in seconds from NTP injection (EC-28, E ~ 0.8) to observing the formation of a long-lived U state (E ~0.6). t_2 represents the time interval from U to the formation of the F state. t_1 and t_2 values were determined by manual inspection of the traces.

Analysis of the t_1 dwell-time: At 100 μM NTPs and 1 mM TPP concentrations (Fig. 3G), a box plot of t_1 values yielded a range of 55 s from a minimum value of 10 s to a maximum of 65 s with an average value of 27 s. Our data do not provide direct information about the exact RNAP position at a given time and therefore the elongation rate of the nascent transcript. However, lower and upper limits for the length of the transcribed RNA during the t_1 phase can be inferred indirectly from the smFRET values obtained for ECs in steady-state conditions. For instance, smFRET histograms of EC-83 complexes showed predominantly a E ~0.8 value, whereas EC-88 displayed the U state (fig. S25A). Therefore, if we assume the conformational transition from EC-28 (E ~0.8) towards a binding competent conformation represented by the U state occurs at the latest at residue 88, we can estimate an elongation length of ~60 nt. Using this value and the average t_1 value of 27 s, we obtained a transcription rate of ~2.3 nt/s at 100 μM NTP concentration. This transcription rate is similar to the pause-free transcription rate (~6-7 nt/s) determined using *E. coli* RNAP at 100 μM NTP using sm-force measurements (*82*). The observed minor ~2.5-fold difference in transcription rates could arise from the existence of pause sites occurring within the *tbpA* riboswitch aptamer or from the uncertainty related to exact RNAP position within the riboswitch sequence.

Analysis of the t_2 dwell-time: The t_2 dwell time, which corresponds to the dwell time of the U state, was assumed to be defined by the time interval from the formation of a binding competent state (EC-88) until a binding event resulted in the TPP-induced formation of the F state. This assignation is supported by the smFRET data of ECs and

also by the fact that real-time FRET traces done in the absence of TPP did not show the F state (fig. S23B). For EC28→EC-136, an average t_2 value of ~98 s was obtained at 100 μM NTP. However, because the spread of values was much higher than for t_1, ranging from 15 s to >300 s, we obtained a median value of ~50 s. Interestingly, a box plot of t_2 values revealed that while ~50% of traces exhibited an average t_2 value of ~28 s ('early TPP binding'), the other half of traces showed an average t_2 value of ~145 s ('late TPP binding'). The 'early TPP binding' group showed a progressive decrease in FRET signal from U to F lasting for ~1 s at 100 μM NTP concentration and ~5 s at 25 μM NTP (Fig. 3H). However, in the case of the 'late TPP binding' group, t_2 spread was assigned to the heterogeneity of the U→F transition appearance across the time window. We interpret the t_2 heterogeneity by the fact that TPP binding may occur on ECs located at various positions along the transcriptional window or in a post-transcriptional manner (SM 2.10).

2.10 Cotranscriptional smFRET transitions observed in real-time transcription reactions

EC-28 (E~0.8)→U transition: Single-molecule analysis of real-time transcription reactions showed gradual FRET decreases for E~0.8→U and U→F transitions (Fig. 3, E and F and fig. S24B). The E~0.8→U (E ~0.6) transition was assigned within a 5 nt window located between EC-83 and EC-88 (SM 2.9). The FRET change observed between EC-83 and EC-88 likely results from the RNA adopting different structures in EC-83 and EC-88 (SM 2.5), where the pyrophosphate sensor helix important for TPP binding (*41*, *42*, *69*) could be adopted in EC-88. The gradual decrease of the transition suggests that the formation of this RNA structure can happen cotranscriptionally, i.e., on the timescale of transcription elongation (*25*, *44*, *45*, *83–86*), where the successive addition of NTP provides supplementary residues participating in the gradual formation of the U state. The gradual FRET transition decrease was observed in all traces with or without TPP, suggesting efficient cotranscriptional folding of this region in all observed conditions.

U→F transition: Based on our smFRET data, the U→F transition consists in the formation of the P3-L5 interaction that is stabilized through TPP binding (Fig. 1E). In the context of EC-88, given that the RNA-DNA hybrid of the elongation complex is ~8-9 base pairs (*76*), it indicates that the P1 stem is not formed and thus not essential for TPP sensing (Fig. 2B). These data suggest a model where TPP could be initially recognized by nascent transcripts through the P3-L5 interaction, which could lead to the stabilization of the P1 stem upon RNAP elongation to a downstream position (e.g., the 104 transcriptional pause site). Alternatively, TPP sensing could be performed when RNAP is already located in a downstream location allowing P1 formation, thus enabling remote coupling between P3-L5 and P1 stem during ligand binding, as suggested by P1 stem stabilization leading to the increased adoption of the folded state (*23*). Although our data do not distinguish between the two modes of TPP recognition, they suggest that metabolite sensing is performed homogeneously by ~50% of ECs exhibiting an 'early TPP binding' profile (t_2 values averaging at ~28 s) (Fig. 3G). In agreement with the cotranscriptional nature of the U→F conformational change, the time of the transition was found to be inversely proportional to the NTP concentration where ~1 s and ~5 s transitions were observed at 100 μM and 25 μM NTP, respectively (Fig. 3H). The presence of synchronized transitions for both transcription rates suggest that similar conformational changes are occurring upon TPP recognition by ECs in the 'early binding' group, which could correspond to the gradual formation of the P3-L5 interaction coupled to the formation of the P1 stem.

In contrast, ECs exhibiting t_2 'late binding' heterogeneous values are indicative of TPP sensing across the transcriptional landscape by nascent riboswitches exhibiting various conformations and ligand binding competency. For example, nascent transcripts within ECs located in the 117-136 region are likely to exhibit a partially formed anti-P1 stem that could delay to some degree TPP binding. However, in case of a productive TPP binding event, the formation of the P3-L5 interaction and P1 stem possibly lead to a progressive change in FRET efficiency due to a probable competition between P1/anti-P1 stem formation occurring prior to TPP-induced folding. The presence of various smFRET profiles in the 'late binding' group suggests that the TPP-dependent folding

could occur via various mechanisms, which will require more work to be fully explained.

Cotranscriptional vs post-transcriptional binding: In contrast to cotranscriptional monitoring of transcription reactions providing information about the first TPP binding event to the riboswitch, it is expected that stalled transcription complexes have undergone successive binding events prior to smFRET detection. Therefore, the presence of abrupt conformational changes in post-transcriptional binding (fig. S24A) suggests that RNA folding is fast compared to the timescale of image recording (~50 ms) and could reflect the conformational sampling of the P3-L5 interaction without extensively altering the secondary structure of the riboswitch. Thus, in contrast to cotranscriptional folding, the absence of a gradual FRET decrease in stalled complexes could be explained by the fact that the P1 stem is not interchanging structures upon the binding and release of TPP. More work will be required to elucidate the implication of P1 formation in the U F transition.

2.11 Applicability of the technique to study other transcription processes

Several lines of evidence suggest that the approach described here may be applied to study the dynamics of nascent transcripts in the context of other transcription machineries. For example, the *Saccharomyces cerevisiae* RNAP II has been shown to initiate promoter-dependent transcription reactions using trinucleotide initiators (*87*), suggesting that fluorescently-labeled trinucleotides could be used to introduce a Cy3 dye at the 5' extremity of the nascent transcript. Furthermore, since *S. cerevisiae* RNAP II transcription elongation complexes support stepwise transcription conditions (*87*), it indicates that the RNAP II could be positioned to site-specifically introduce the required analog. There is a very high probability that an azide-UTP analog would be recognized by the *S. cerevisiae* RNAP II since it was shown that mammalian RNA polymerases I, II and III efficiently incorporate click-compatible UTP analogs to allow subsequent fluorescent labeling (*64*, *88*). This is in agreement with the RNAP core structure being conserved among the three domains of life (*89*). The latter observation also suggests that the biotin-streptavidin roadblock would be suitable to generate stalled RNAP II

elongation complexes, thus enabling the immobilization of such complexes for smFRET analysis. Although there is less available evidence about the feasibility of our approach for human RNAP, current *in vitro* systems using dinucleotide transcription initiators (*90*) provides a working system that could be amenable to smFRET analysis.

Fluorescent dyes have also been incorporated within RNA molecules through the synthetic reconstitution of bacterial or eukaryotic nucleic scaffolds (*14*, *15*, *91*). The fact that both bacterial and eukaryotic RNAP support similar architectures to initiate transcription indicates that the presence of a labeled trinucleotide should be efficiently elongated during transcription.

Furthermore, this approach allows to site-specifically introduce any click-compatible analogs, such as crosslinking reagents or spin labels, for RNA and RNA-protein structural analyses. In contrast to synthetic nucleic scaffolds (*14–16*), our method relies on PCR-made DNA templates, thus providing a fast and inexpensive way to produce dual-labeled nascent transcripts in the kilobase range.

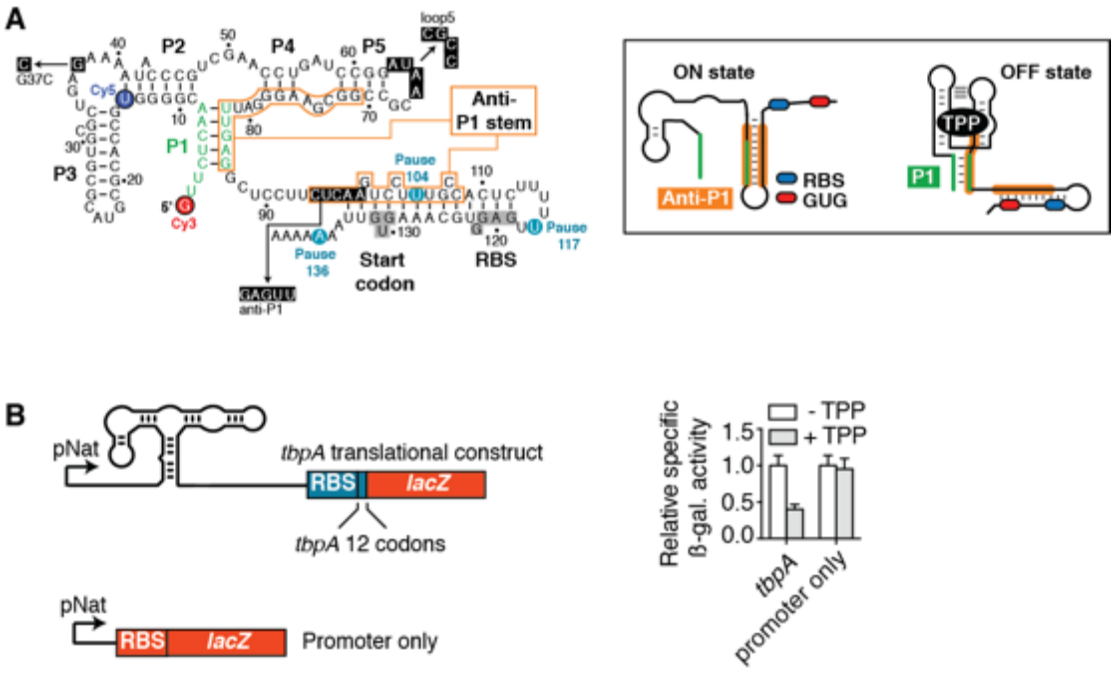


Figure S1. Regulation mechanism of the TPP-sensing *tbpA* riboswitch.

(A) Sequence and secondary structure of the *tbpA* riboswitch shown in the predicted ligand-bound conformation (OFF state). Based on the riboswitch regulation model, TPP binding induces the formation of the P1 stem and sequesters the RBS and start codon. The anti-P1 stem is predicted to only form in the absence of TPP (ON state). The P3-L5 interaction (*41*, *42*) between the stem P3 (C31) and the loop 5 (A65-A66) is not shown for clarity purpose. Mutations inhibiting ligand binding (G37C), P3-L5 interaction (loop5) and the anti-P1 stem are shown (*22*). The transcriptional pause sites 104, 117 and 136 were determined from a previous study (*19*). The ribosome binding site (RBS, positions 119-122) and GUG start codon (positions 129-131) are shown. The insert shows the *tbpA* riboswitch in the proposed ON and OFF states, which are characterized by the presence of the anti-P1 (orange) and P1 stem (green), respectively. **(B)** Left, schematic description of beta-galactosidase constructs monitoring *tbpA* expression. A construct is also shown which monitors the expression only in the context of the natural promoter (pNat). Right, beta-galactosidase assays of translational and promoter-only fusions when cells were grown in the absence or presence of 500 μg/mL TPP. Values were normalized to data obtained in the absence of TPP. The average values of three independent experiments with standard deviations are shown.

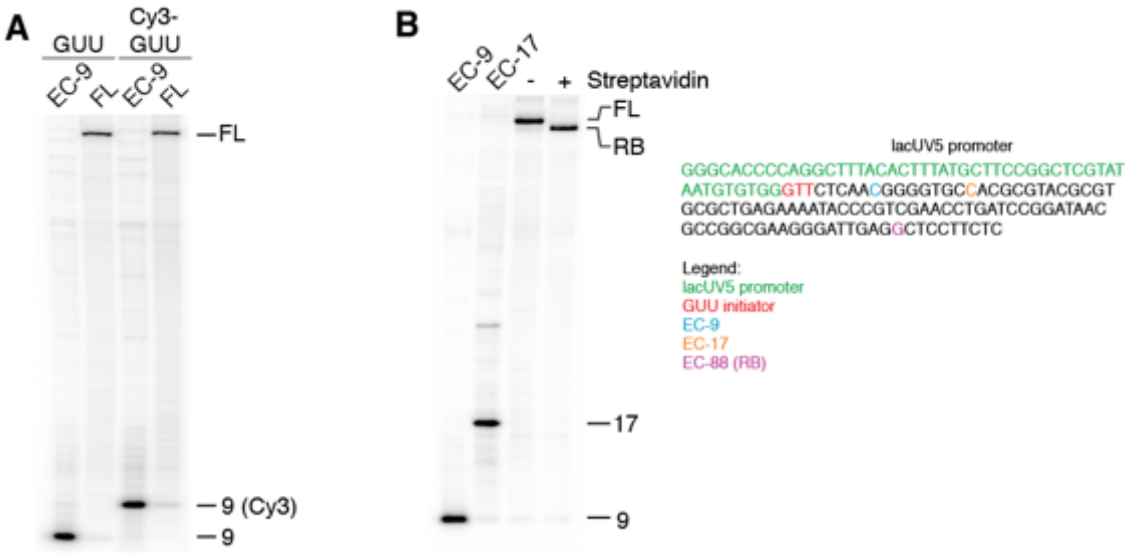


Figure S2. *In vitro* transcription reactions of the *tbpA* riboswitch.

(A) *In vitro* transcription reactions performed in the presence of GUU or Cy3-GUU trinucleotides. The EC-9 and full-length (FL) transcription elongation products are shown. As expected, a gel retardation for the 9-nt transcript is observed when Cy3-GUU is used to initiate the transcription. The structure of Cy3-GUU is showed in SM 2.2. **(B)** Left, *in vitro* stepwise transcription reactions showing the formation of EC-9, EC-17 and of elongated reactions performed in the absence (full-length, FL) or presence (roadblock, RB) of streptavidin. Right, sequence of the DNA template used for the stepwise transcription showing the lacUV5 promoter and the different elongation steps. Coupling of the 5' extremity of the antisense strand to a biotin ensures that the presence of streptavidin stalls RNAP 9 nt before the template end (*19*).

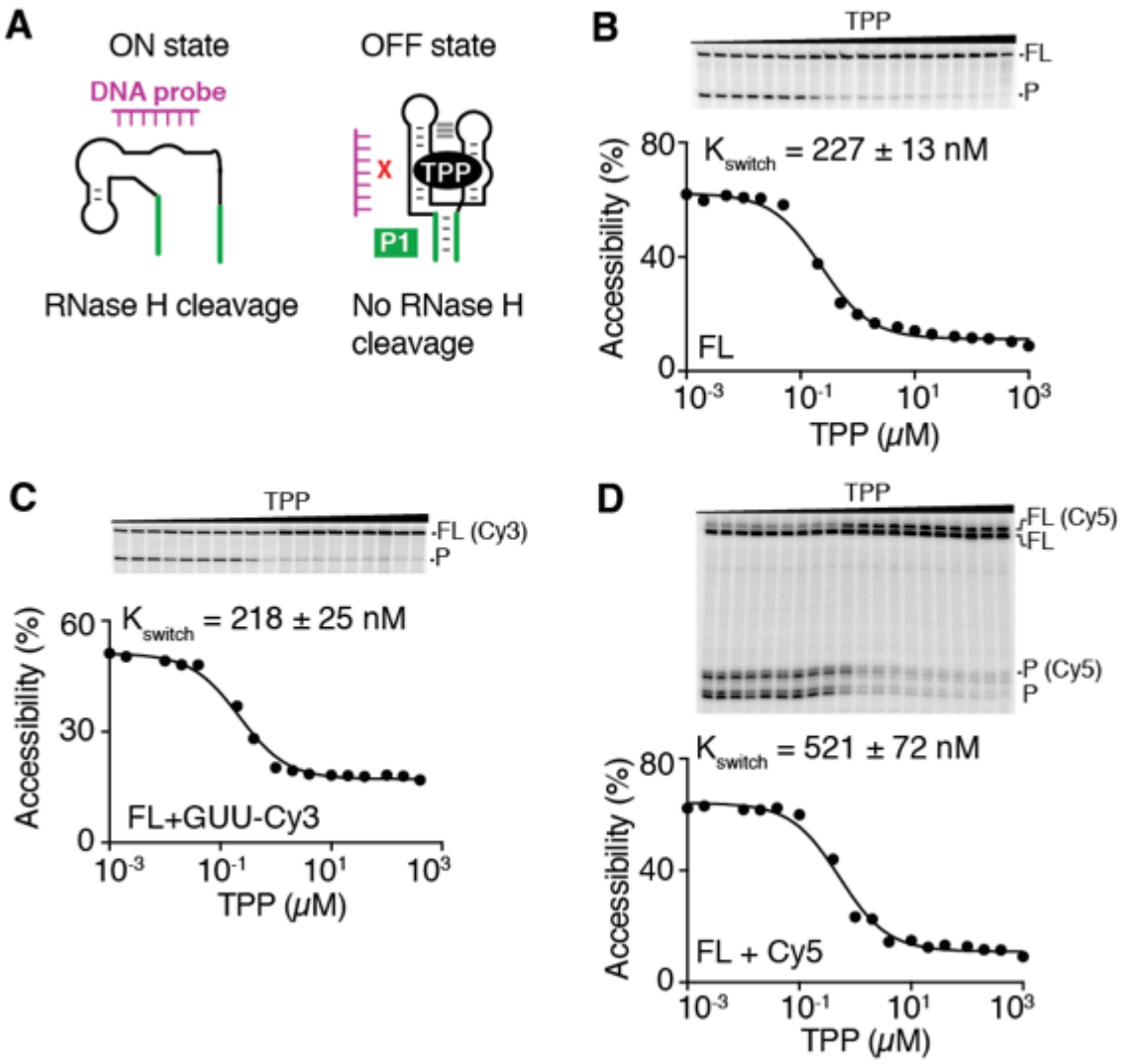
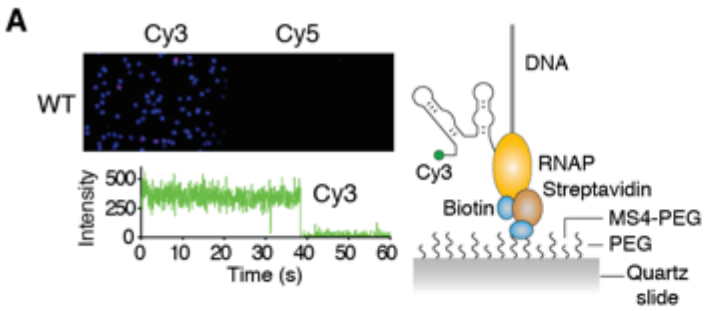


Figure S3. K_{switch} determination of fluorescently labeled *tbpA* riboswitches.

(A) Schematic description of RNase H assays in which a DNA probe is used to target the J2-4 region (residues 40-49). In the absence of TPP, the DNA oligonucleotide hybridizes to the RNA which is cleaved by RNase H. However, TPP binding stabilizes the aptamer structure and does not allow DNA hybridization and RNase H cleavage. **(B-D)** K_{switch} determination of the unmodified **(B)**, Cy3-labeled **(C)** or Cy5-labeled **(D)** *tbpA* transcript done in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. RNase H assays performed on *tbpA* riboswitches submitted to Cy5 click reaction showed the presence of two full-length (FL) and product (P) species, suggesting that a fraction of riboswitch transcripts were not labeled with Cy5. The full-length (FL) and

cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown below the gel.



B

Elongation Complexes	Average number of molecules per field of view
Intact complexes	190
No streptavidin	1
No biotin on DNA	1
No DNA	6
No RNAP	6
No Cy3	2

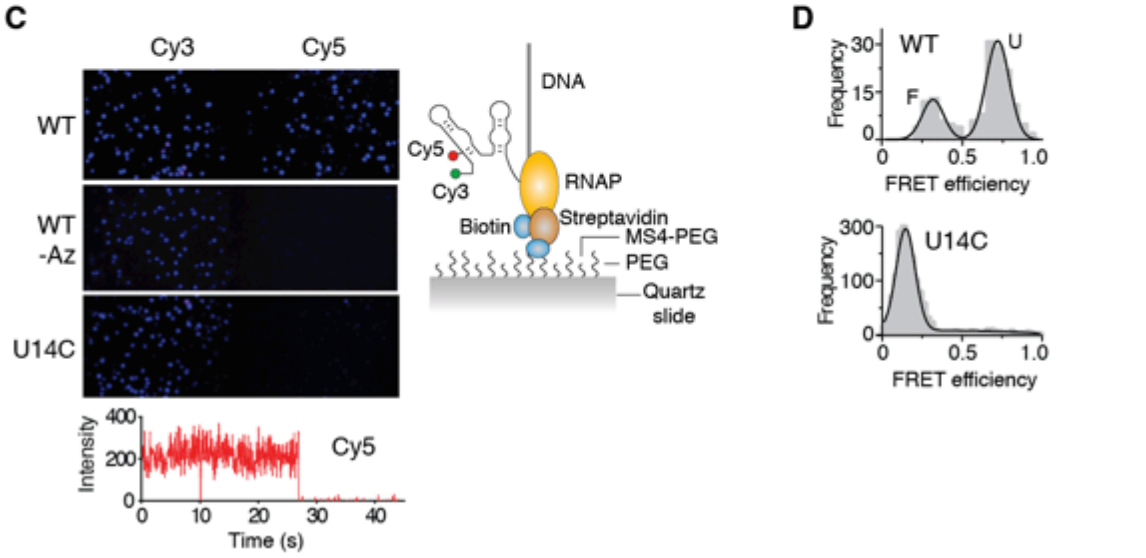


Figure S4. Single-molecule imaging of ECs.

(A) Representative single-molecule image of Cy3-labeled ECs immobilized through the biotin-streptavidin roadblock. Spatially separated emission spots showed a single photobleaching step (below the image), thus confirming the presence of a single Cy3 dye per EC. The scheme represents the visualized complex immobilized on the microscope slide. **(B)** Summary of control experiments carried out to demonstrate specific immobilization through the roadblock. Significant signal is only obtained in the presence of all constituents of the Cy3-labeled EC. **(C)** Representative single-molecule images of Cy3-Cy5 labeled ECs. While Cy5 fluorescent spots are observed when using the wild-type EC (top), no Cy5 signal is observed without the azide-UTP analog (middle) or when using a U14C mutant not allowing incorporation of the UTP analog (bottom). A single-photobleaching Cy5 step is observed confirming the incorporation of a single Cy5 dye in the WT complex (below the image). The scheme shows the dual-labeled wild-type complex attached to the microscope slide. **(D)** Comparison of single-

molecule dwell-time histograms obtained for WT (top) and U14C complexes (bottom). For U14C complexes that are unable to incorporate the Cy5 dye, the single-molecule histogram showed a predominant Gaussian (~89%) centered (zero peak) at E ~0.10-0.15 corresponding to Cy3-only complexes. A 10-15% leakage of the Cy3 signal through the DCLP645 dichroic mirror leads to an apparent FRET efficiency of E ~0.10-0.15.

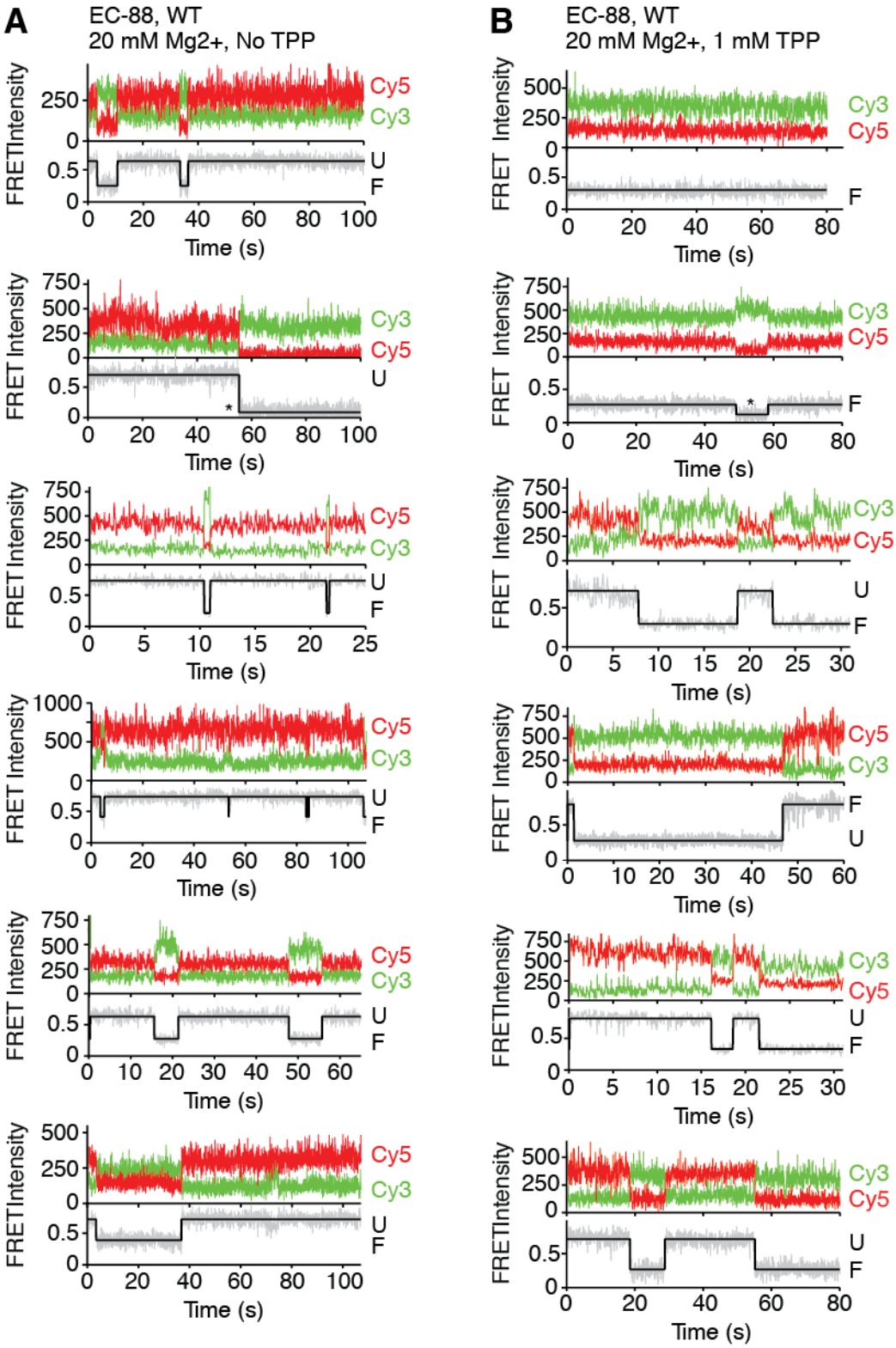


Figure S5. smFRET trajectories of EC-88.

(A,B) Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained

at 20 mM $MgCl_2$ in the absence **(A)** or presence of 1 mM TPP **(B)**. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching and blinking events are indicated by asterisks.

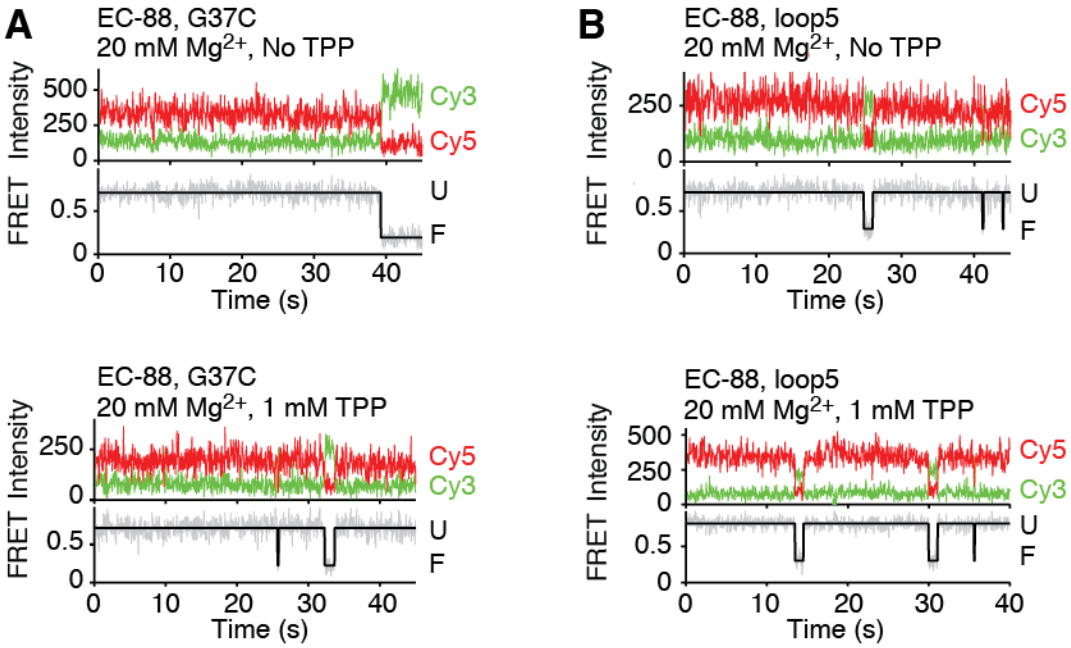


Figure S6. smFRET trajectories of G37C and loop5 EC-88 mutants.

(A-B) Representative smFRET trajectories of the G37C **(A)** and loop5 **(B)** mutants showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence (top) or presence of 1 mM TPP (bottom). Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. In both mutants, only very transient formation of the F state is observed in the absence and presence of 1 TPP.

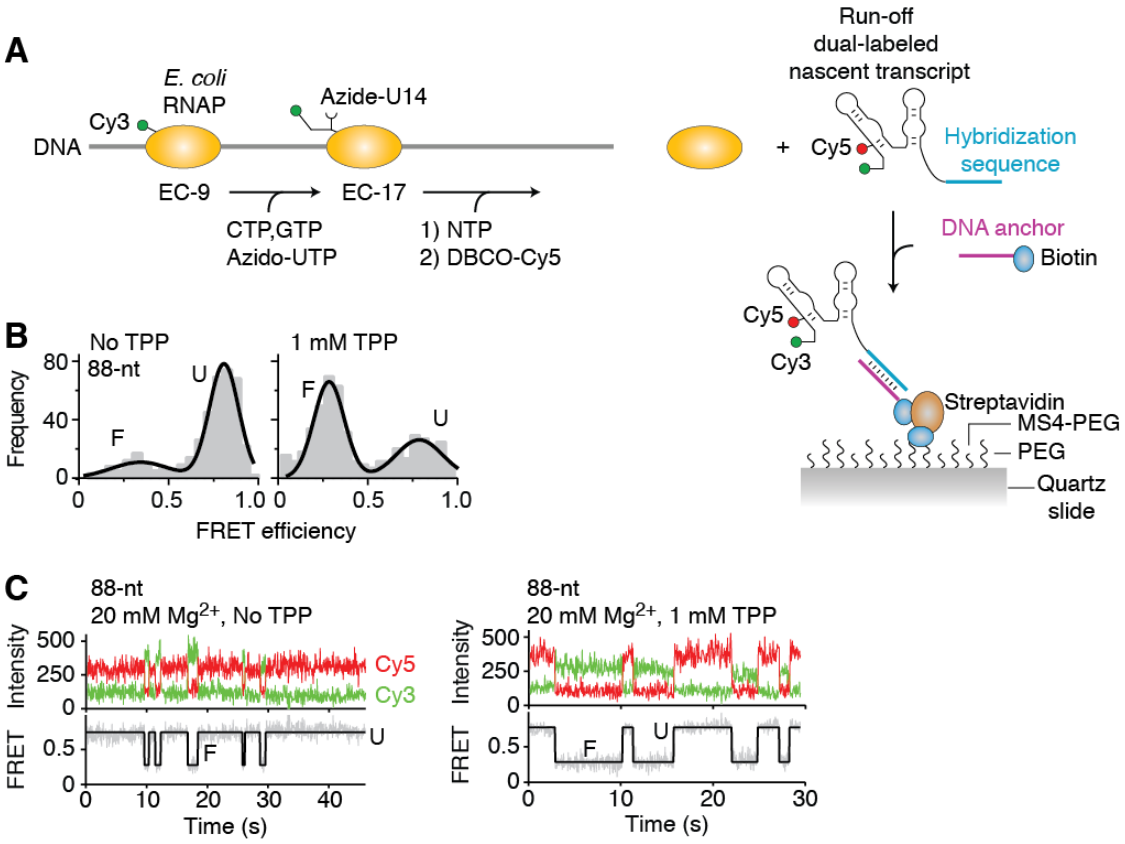


Figure S7. smFRET analysis of the 88-nt *tbpA* nascent transcript.

(A) Schematic representation of the approach used to obtain the Cy3-Cy5 dual-labeled nascent transcript. While the formation of EC-9 and EC-17 is described in Fig. 1A, the absence of streptavidine in the third step allows run-off transcription and release of the Cy3-labeled RNA from EC. DBCO-Cy5 is then allowed to react with the azide-containing transcript to generate the Cy3-Cy5 dual-labeled RNA. A biotinylated DNA oligonucleotide (anchor) is used to immobilize the nascent transcript on the microscope slide. **(B)** smFRET histograms of the 88-nt *tbpA* nascent transcript in the absence or presence of 1 mM TPP. **(C)** Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right.

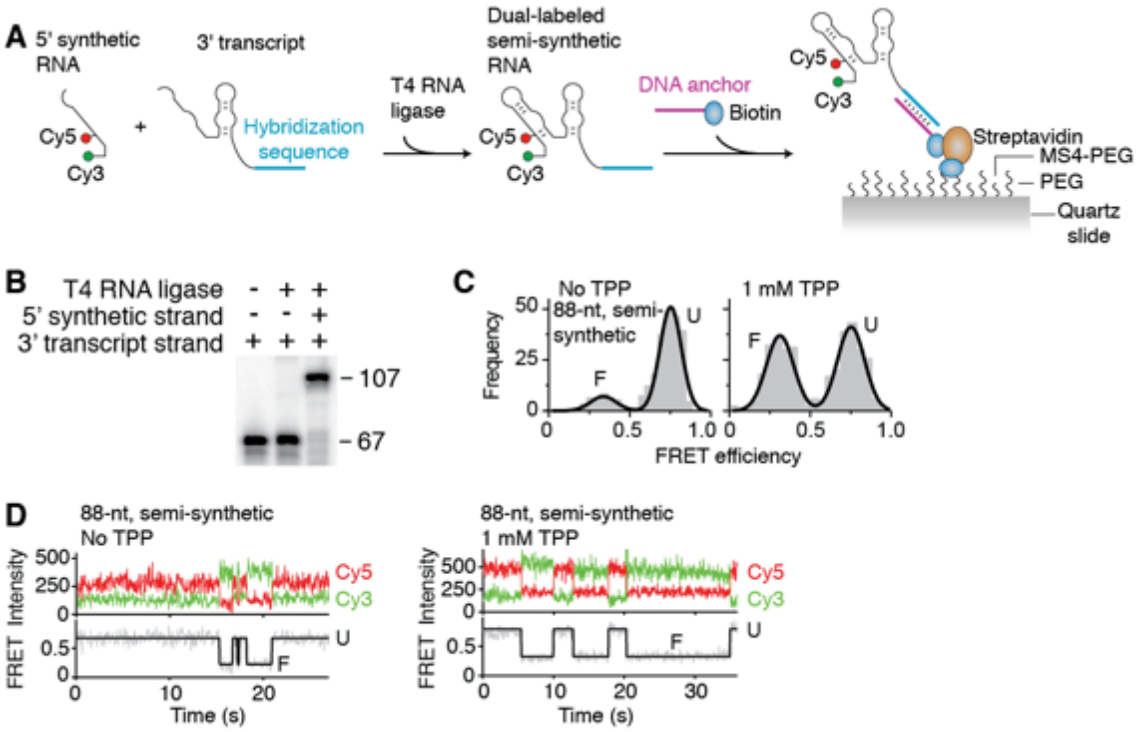


Figure S8. smFRET analysis of a semi-synthetic 88-nt *tbpA* RNA.

(A) Schematic representation of the methodology used to produce the semi-synthetic *tbpA* RNA. A synthetic RNA strand carrying Cy3 and Cy5 dyes at positions 1 and 14, respectively, is ligated to a T7 RNAP-purified transcript to allow the reconstitution of the *tbpA* aptamer. A biotinylated DNA oligonucleotide (anchor) is used to immobilize the semi-synthetic transcript on the microscope slide. **(B)** T4 RNA ligation of the 5' synthetic and 3' transcript strands. Formation of the full-length RNA was confirmed by using a 5'-radiolabeled 3' transcript strand. The expected length of the construct is 107 nt. **(C)** smFRET histograms in the absence or presence of 1 mM TPP. **(D)** Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right.

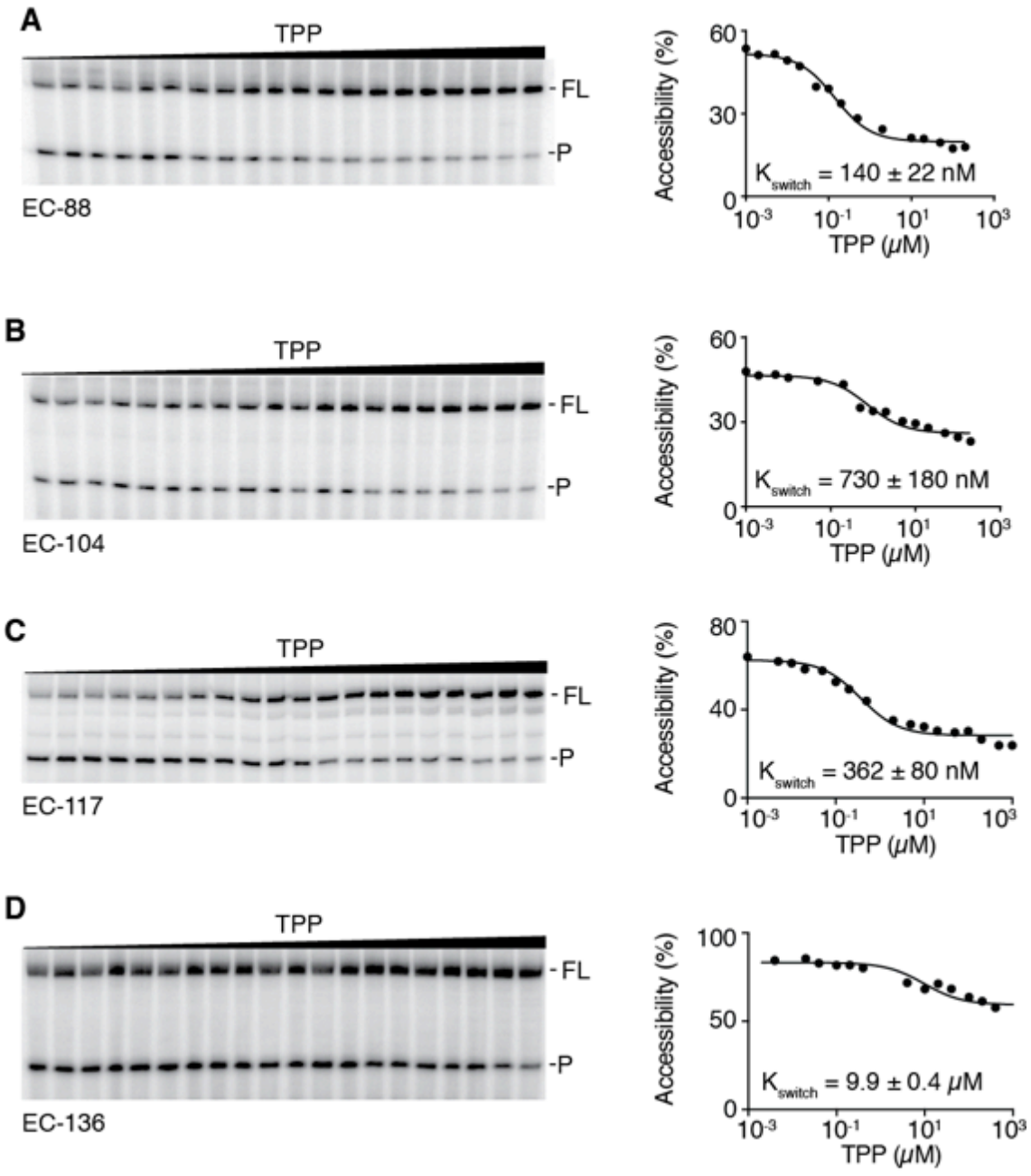


Figure S9. K_{switch} determination of nascent transcripts at EC-88 and transcriptional pause sites.

(A-D) K_{switch} determination of EC-88 **(A)**, EC-104 **(B)**, EC-117 **(C)** and EC-136 **(D)**. RNase H assays were performed in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. The full-length (FL) and cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right.

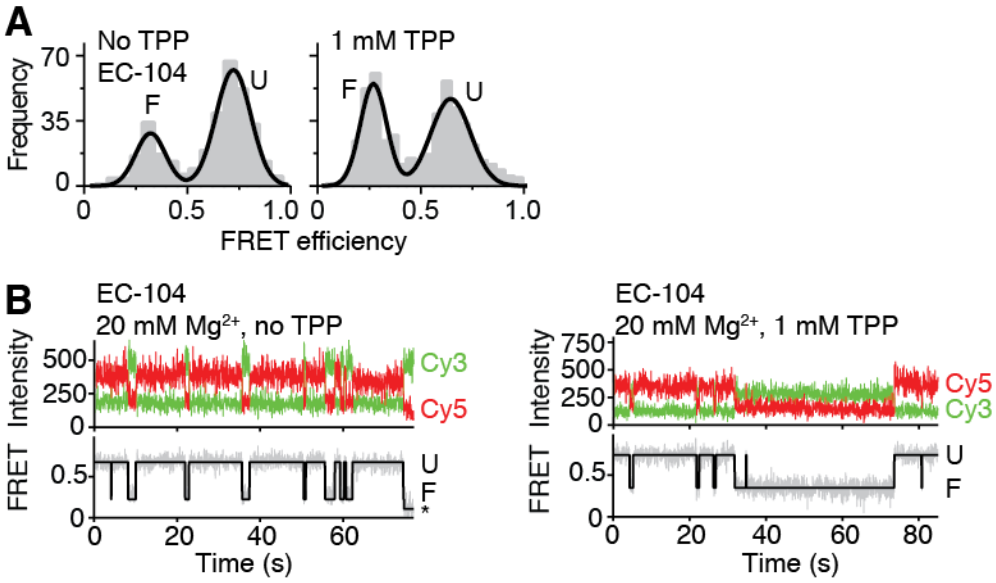


Figure S10. smFRET analysis of EC-104.

(A) smFRET histograms in the absence or presence of 1 mM TPP. **(B)** Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching events are indicated by asterisks.

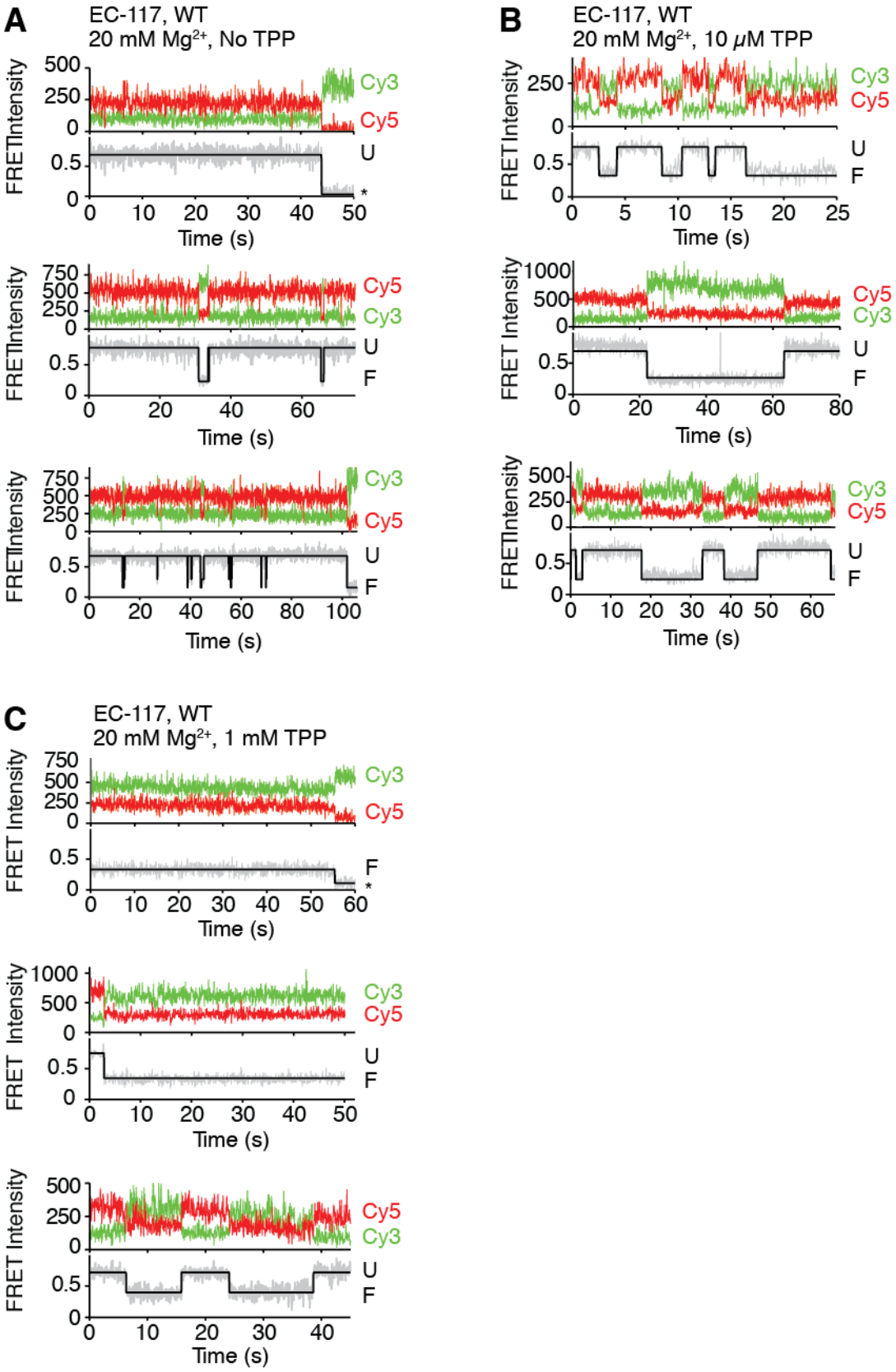

A
EC-117, WT
20 mM Mg2+, No TPP
FRETIntensity
Cy3
Cy5
U
F
Time (s)
B
EC-117, WT
20 mM Mg2+, 10 μM TPP
C
EC-117, WT
20 mM Mg2+, 1 mM TPP

Figure S11. smFRET trajectories of EC-117.

(A-C) Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence **(A)** or presence of 10 μM **(B)** or 1 mM TPP **(C)**. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching events are indicated by asterisks.

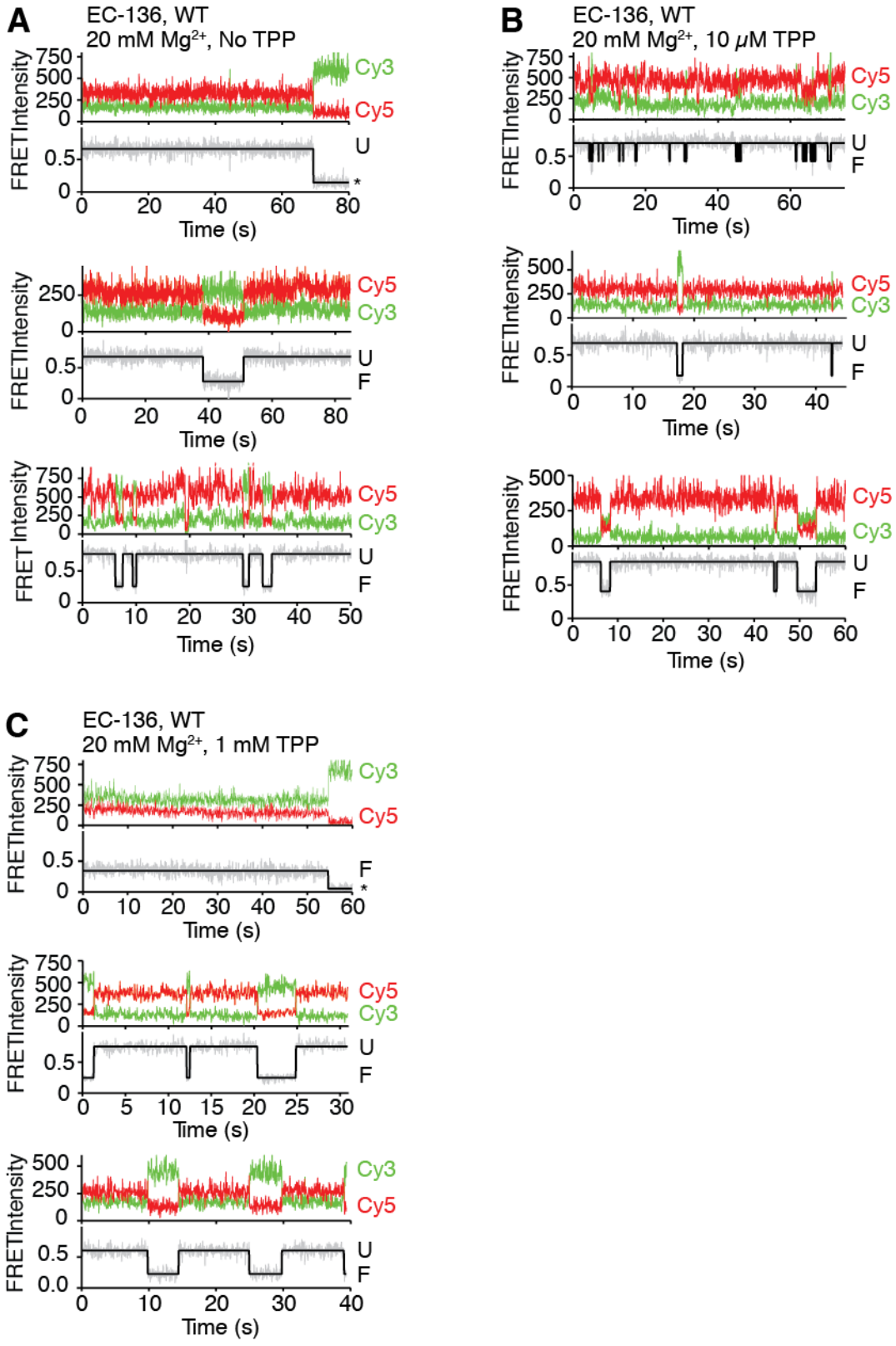

A
EC-136, WT
20 mM Mg2+, No TPP
FRETIntensity
Cy3
Cy5
U
F
*
Time (s)
B
EC-136, WT
20 mM Mg2+, 10 μM TPP
C
EC-136, WT
20 mM Mg2+, 1 mM TPP

Figure S12. smFRET trajectories of EC-136.

(A-C) Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence **(A)** or presence of 10 μM **(B)** or 1 mM TPP **(C)**. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching events are indicated by asterisks.

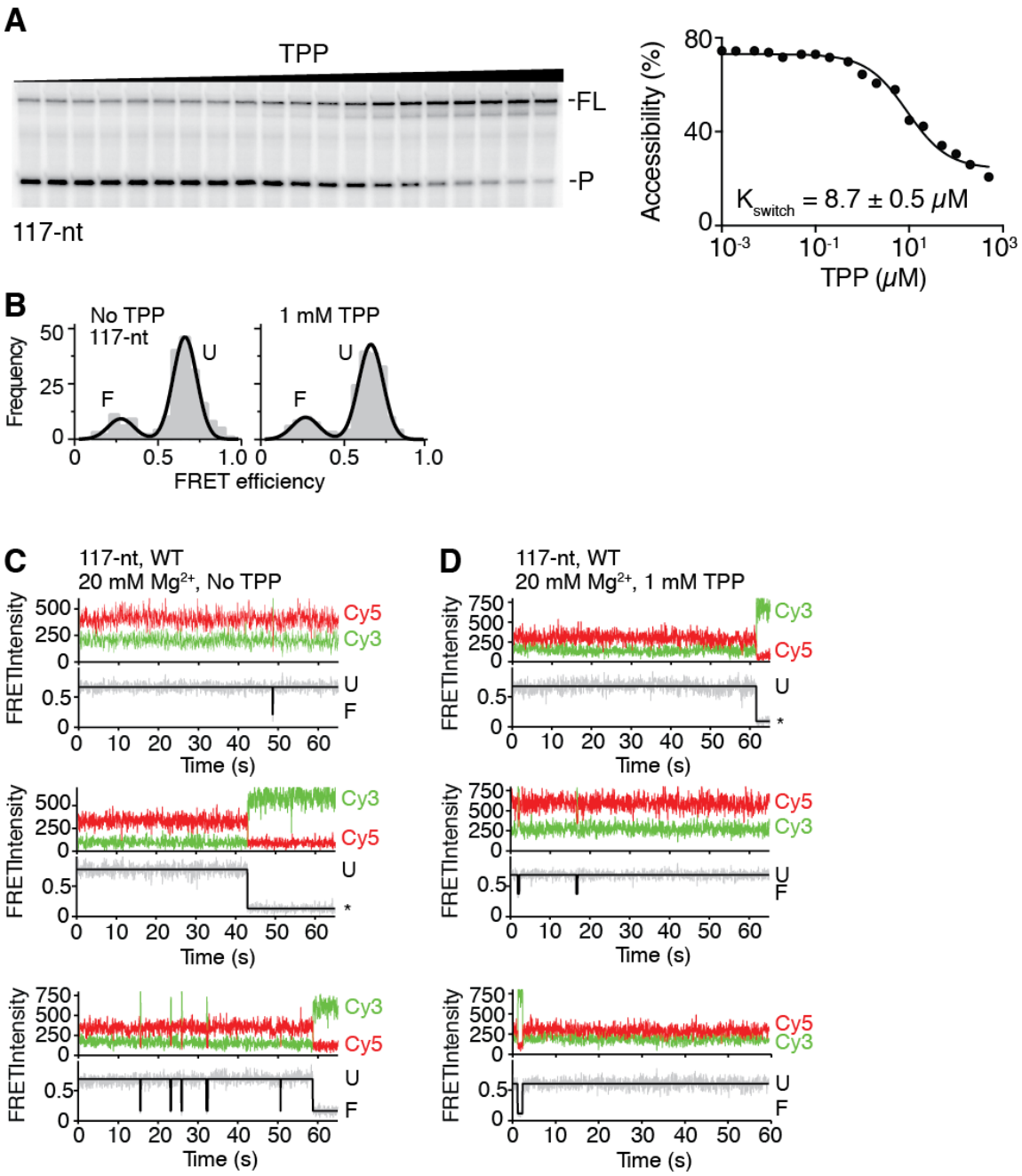


Figure S13. Biochemical and smFRET analysis of the 117-nt *tbpA* nascent transcript.

(A) K_{switch} determination of the 117-nt transcript using RNase H degradation assays performed in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. The full-length (FL) and cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right. **(B)** smFRET histograms obtained in the absence or presence of 1 mM TPP. **(C-D)** Representative

smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence **(C)** or presence **(D)** of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching events are indicated by asterisks.

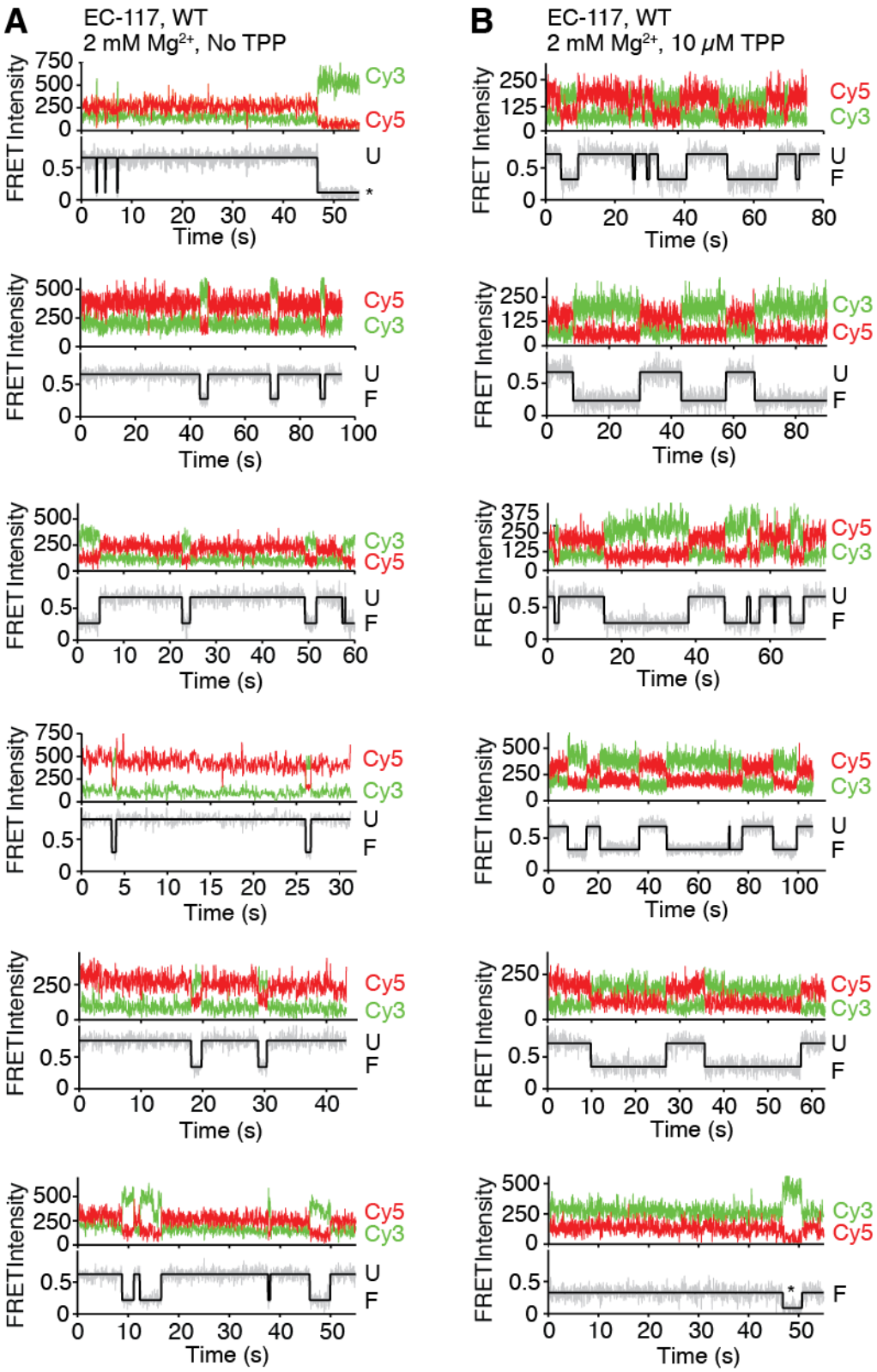

A
EC-117, WT
2 mM Mg2+, No TPP
B
EC-117, WT
2 mM Mg2+, 10 μM TPP
FRET Intensity
Cy3
Cy5
U
F
*
Time (s)

Figure S14. smFRET trajectories of EC-117.

(A-B) Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 2 mM $MgCl_2$ in the absence **(A)** or presence of 10 μM TPP **(B)**. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right. Photobleaching and blinking events are indicated by asterisks.

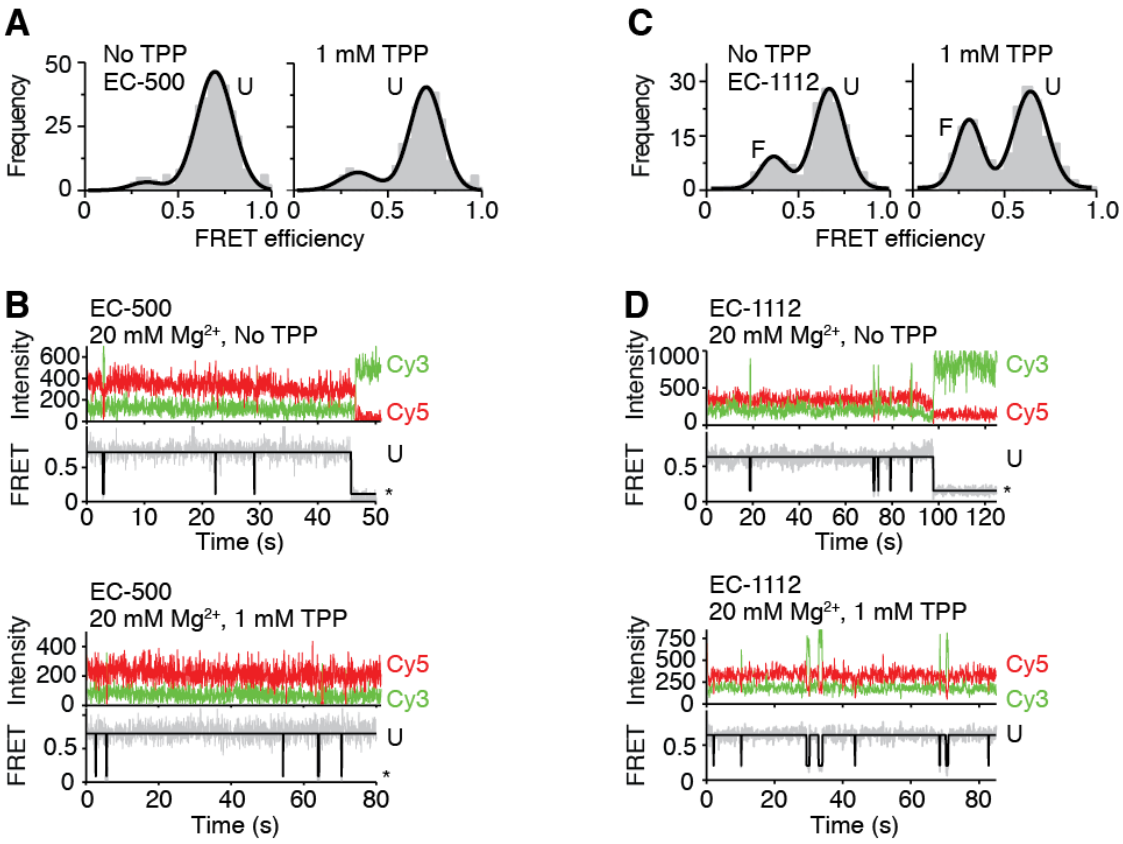


Figure S15. smFRET analysis of EC-500 and EC-1112.

(A) smFRET histograms of EC-500 obtained in the absence or presence of 1 mM TPP. **(B)** Representative smFRET trajectories of EC-500 showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. **(C)** smFRET histograms of EC-1112 obtained in the absence or presence of 1 mM TPP. **(D)** Representative smFRET trajectories of EC-1112 showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Unfolded (U) and folded (F) conformations are identified on the right.

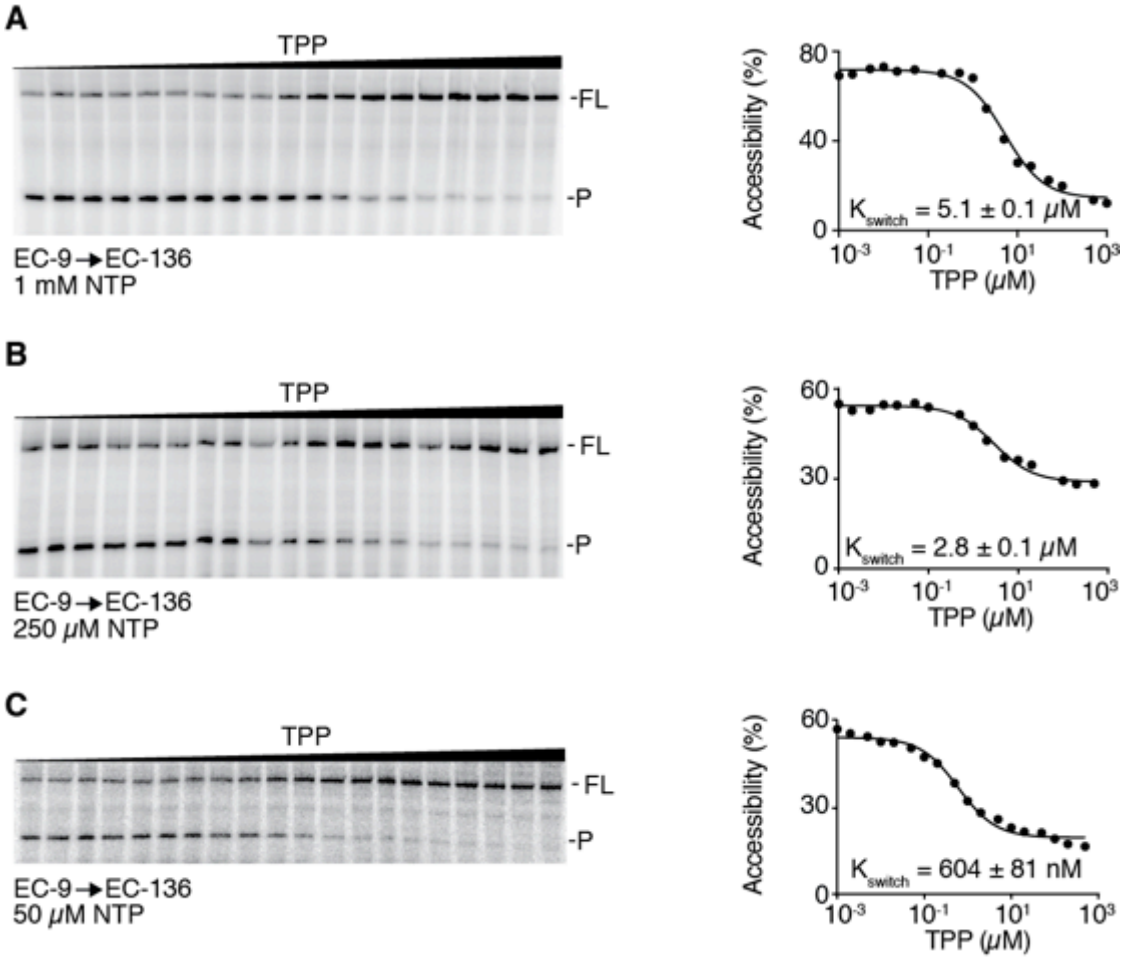


Figure S16. K_{switch} determination of EC-136 transcribed at different NTP concentrations.

K_{switch} determination of EC-136 transcribed with 1 mM, 250 μM and 50 μM NTP. Transcriptions were performed in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. The full-length (FL) and RNase H cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right.

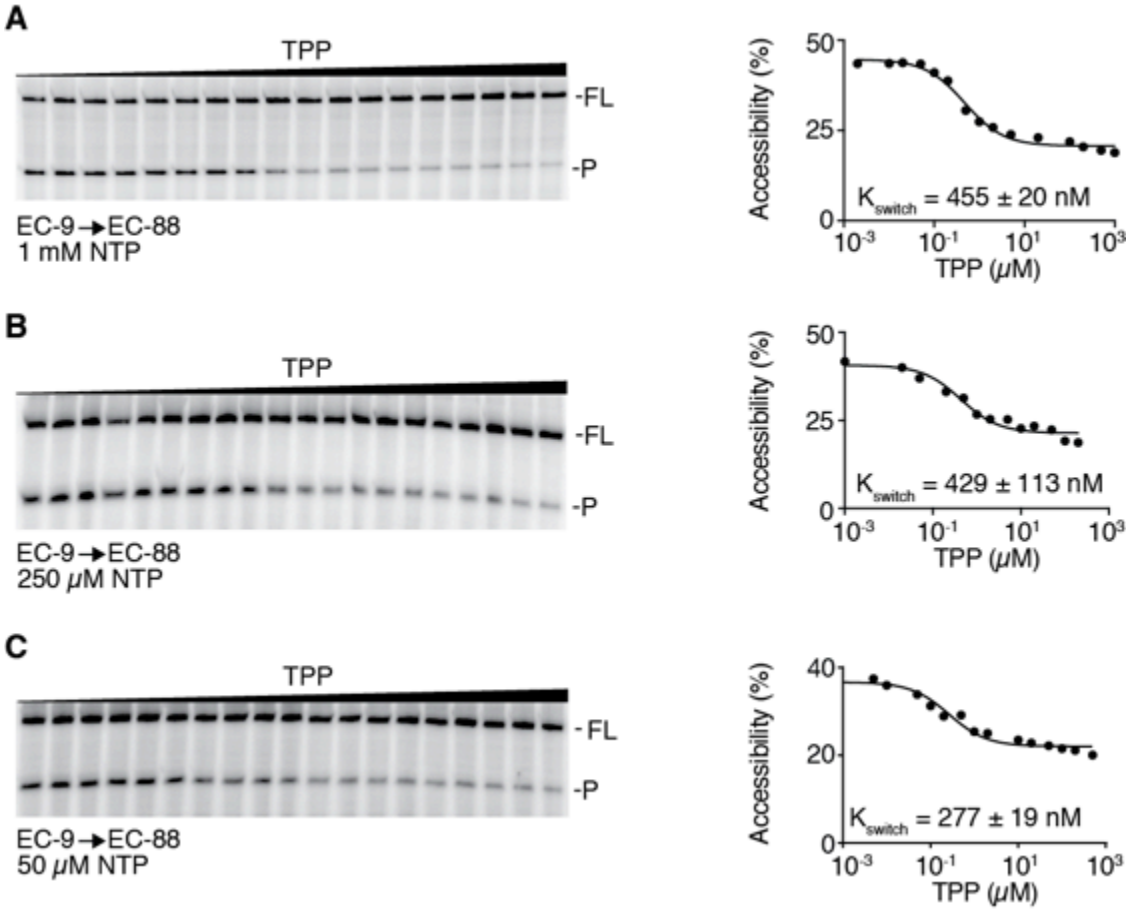


Figure S17. K_{switch} determination of EC-88 transcribed at different NTP concentrations.

K_{switch} determination of EC-88 transcribed with 1 mM, 250 μM and 50 μM NTP. Transcriptions were performed in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. The full-length (FL) and RNase H cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right.

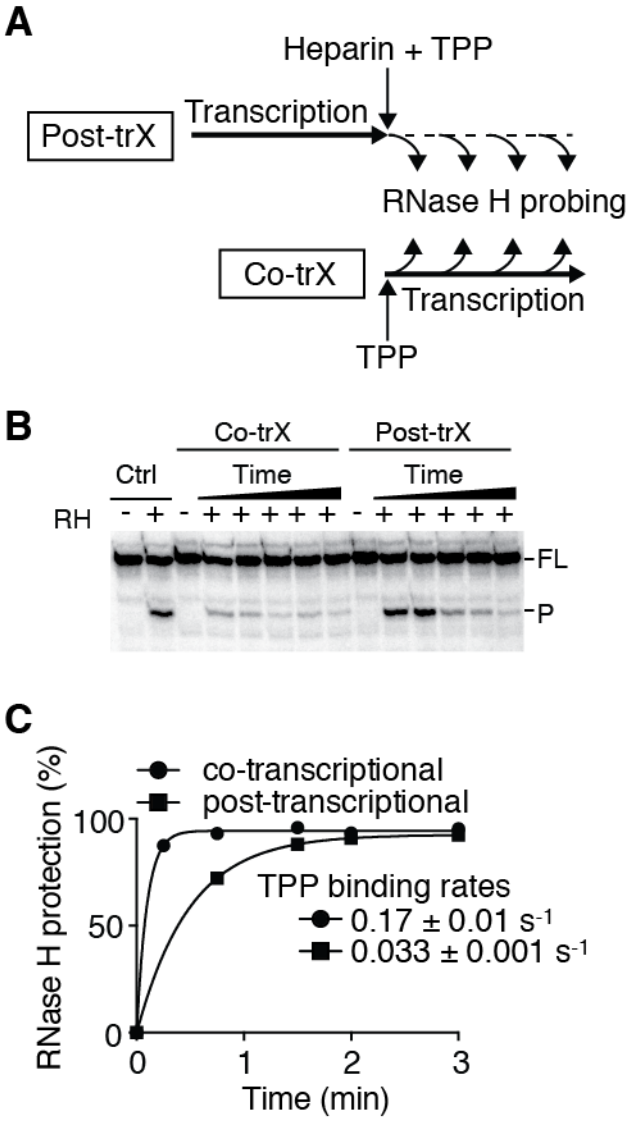


Figure S18. RNase H degradation assays monitoring cotranscriptional or post-transcriptional TPP binding to the *tbpA* riboswitch.

(A) Experimental approach monitoring TPP binding. Post-transcriptional TPP binding (Post-trx) was assessed by performing transcription in absence of TPP. After stopping transcription, TPP is added and RNase H degradation assays are performed at (15 s, 45 s, 90 s, 2 min and 3 min). Cotranscriptional TPP binding (Co-trx) was assessed by adding TPP during transcription elongation and by performing RNase H degradation assays during transcription. **(B)** RNase H probing assays monitoring TPP binding to the riboswitch. Control reactions (Ctrl) done in presence of RNase H (RH) show a cleaved product (P) without TPP. Post-transcriptional and cotranscriptional reactions were

performed as indicated in panel A. The FL and P are indicated on the right. **(C)** Quantification analysis of RNase H experiments. Data were fitted to a single-exponential model and yielded apparent TPP binding rates of 0.17 ± 0.01 s^{-1} and 0.033 ± 0.001 s^{-1} for cotranscriptional and post-transcriptional reactions, respectively. The ~5-fold faster cotranscriptional binding rate of TPP suggests that it is performed preferentially over post-transcriptional binding.

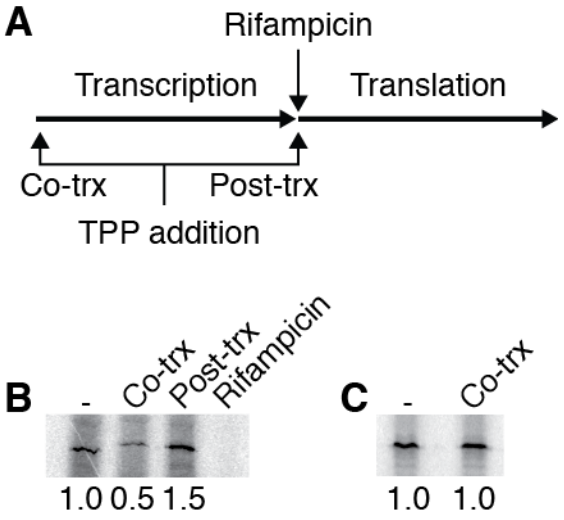


Figure S19. Transcription-translation *in vitro* assays of the *tbpA* riboswitch.

(A) Experimental setup monitoring TPP binding using transcription-translation in vitro assays. The effect of post- transcriptional (Post-trx) TPP binding was assessed by performing transcription reactions in absence of TPP. After the addition of rifampicin, amino acids and TPP were added to allow TPP binding to the riboswitch during translation. The cotranscriptional (Co-trx) TPP binding was assessed by adding TPP at the beginning of transcription. **(B)** Transcription-translation *in vitro* assays monitoring TPP binding to the riboswitch. Reactions were performed in absence (-) or presence of 100 μM TPP added either cotranscriptionally (Co-trx) or post-transcriptionally (Post-trx) as indicated in panel A. The absence of post-transcriptional regulation suggests that TPP sensing is more efficient when done cotranscriptionally. No product was obtained when rifampicin was added at the beginning of the transcription step (Rifampicin), consistent with rifampicin being efficient to block transcription initiation (*92*). **(C)** Transcription-translation *in vitro* assays using a *lacZ* construct. As expected, control reactions performed without (-) or with cotranscriptional addition of 25 μM TPP (Post-trx) did not affect the protein level.

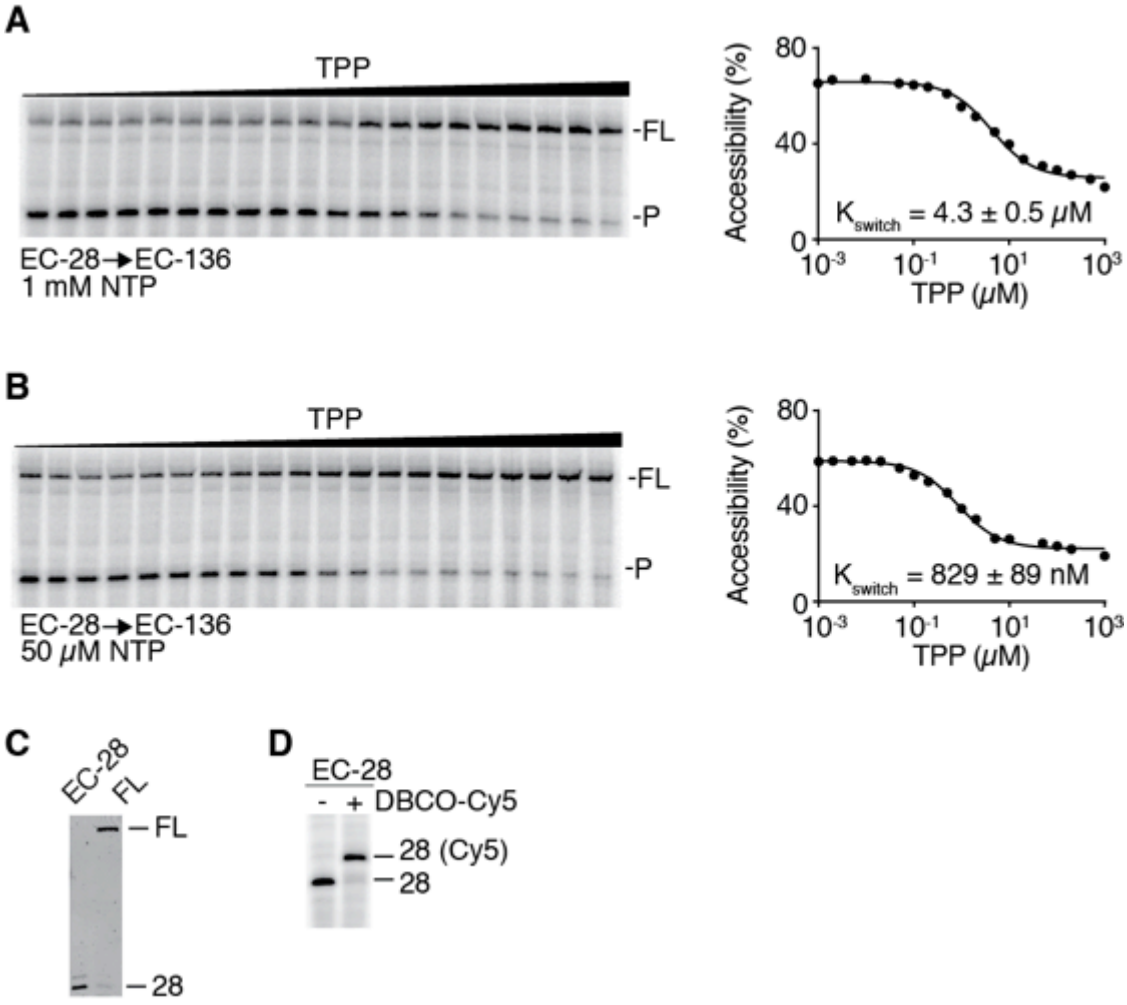


Figure S20. Validation of EC-28 for the study of riboswitch folding.

(A-B) K_{switch} determination of EC-28 EC-136 when transcribing with 1 mM NTP **(A)** or 50 μM NTP **(B)**. TPP concentrations ranging from 1 nM to 1 mM were added cotranscriptionally and RNase H assays were performed on transcription products. The full-length (FL) and cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right. The similar K_{switch} values obtained for EC-28 EC-136 or EC-136 (table S2) indicate that the introduction of the U23A mutation or the pausing at EC-28 does not alter riboswitch folding. **(C)** Transcription reactions of EC-28 and full-length transcript using a Cy3-GUU initiator. Detection was performed using Cy3 fluorescence. **(D)** PAGE separation of 28-nt transcripts obtained with or without reaction with DBCO-Cy5.

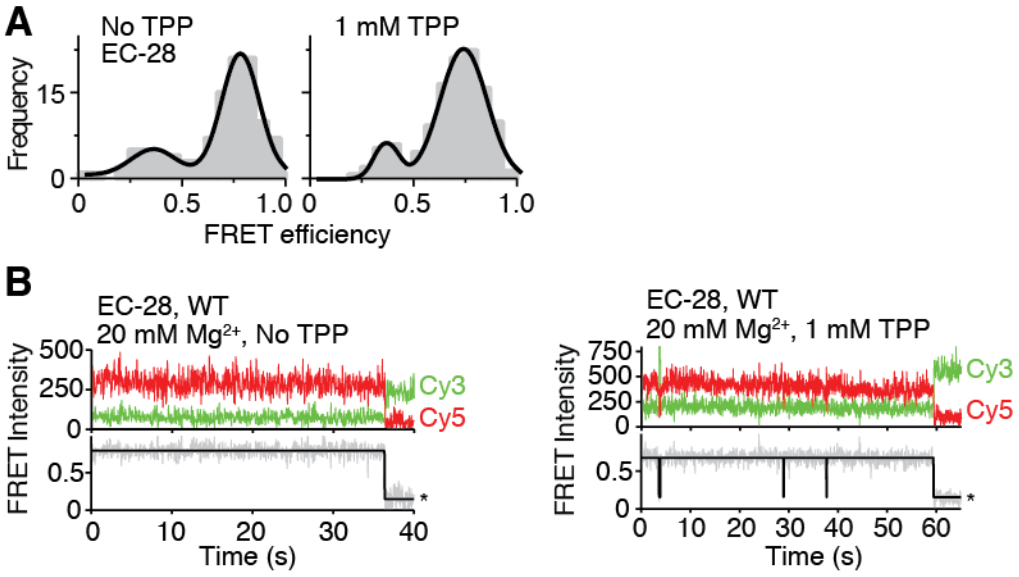


Figure S21. smFRET analysis of EC-28.

(A) smFRET histograms obtained in the absence or presence of 1 mM TPP. **(B)** Representative smFRET trajectories showing the anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces obtained at 20 mM $MgCl_2$ in the absence or presence of 1 mM TPP. Black lines represent the hidden Markov modelling of the FRET traces. Photobleaching events are identified by an asterisk.

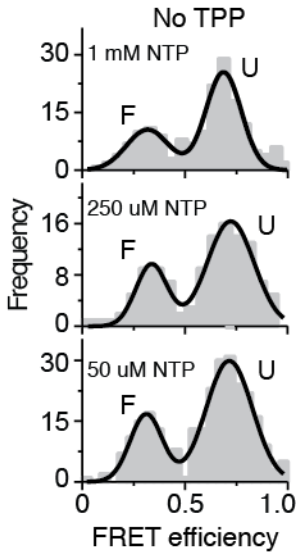


Figure S22. smFRET histograms obtained after transcribing EC-28 EC-136 in the absence of TPP.

Transcription elongation reactions were performed in the presence of 1 mM, 250 μM or 50 μM NTP.

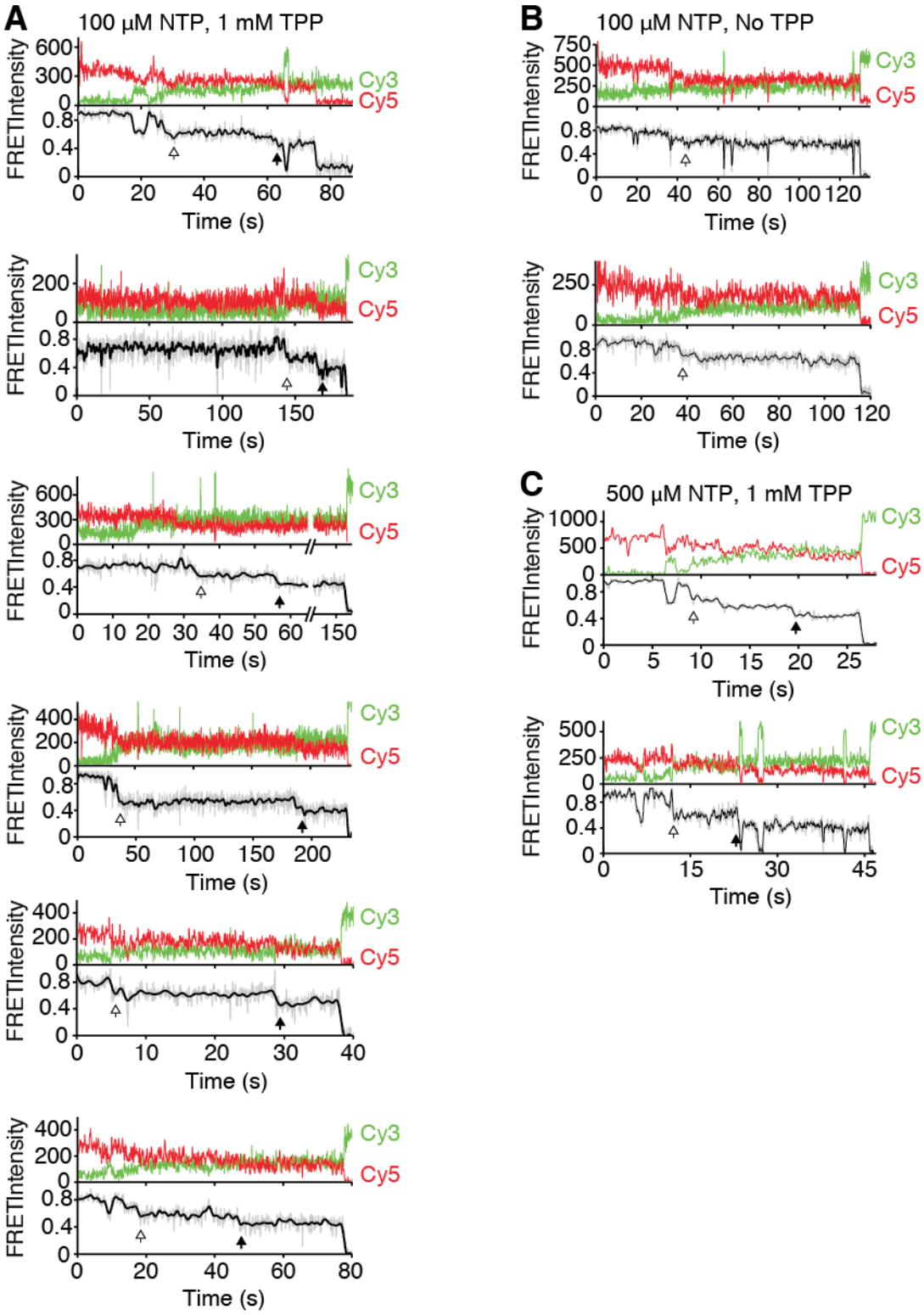

A
100 µM NTP, 1 mM TPP
FRETIntensity
Cy3
Cy5
Time (s)
B
100 µM NTP, No TPP
C
500 µM NTP, 1 mM TPP

Figure S23. Real-time cotranscriptional folding of *tbpA* nascent transcripts.

(A-C) Real-time monitoring of the cotranscriptional folding during transcription elongation using a concentration of 100 μM NTP in the absence **(A)** or presence **(B)** of 1 mM TPP, or using 500 μM NTP and 1 mM TPP **(C)**. Transcriptions were performed using the EC-28->EC-136. The anti-correlated donor (green) and acceptor (red) emission intensities, together with the resulting FRET traces are shown. White and black arrows represent ~0.8 U and U F transitions, respectively. A smoothing line was added using an adjacent-averaging algorithm with N=5 points. Injections were done at 5 s.

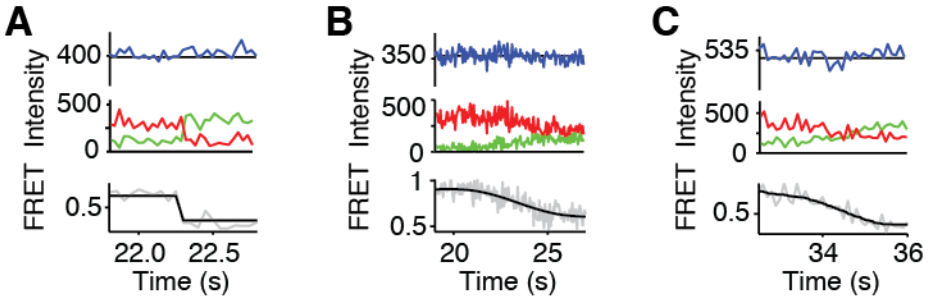


Figure S24. smFRET transitions obtained on stalled and elongating RNAP complexes.

(A) U F transition observed within a stalled EC-117 complex in the presence of 2 mM $MgCl_2$ and 10 μM TPP. **(B-C)** smFRET transitions EC-28 (E ~0.8) U **(B)** and U F **(C)** observed during real-time transcription of EC-28 EC-136 elongating complexes. Total intensity (top panel), Cy3 (green) and Cy5 (red) intensities and FRET (bottom panel) trajectories in the presence of 100 μM NTP and 1 mM TPP.

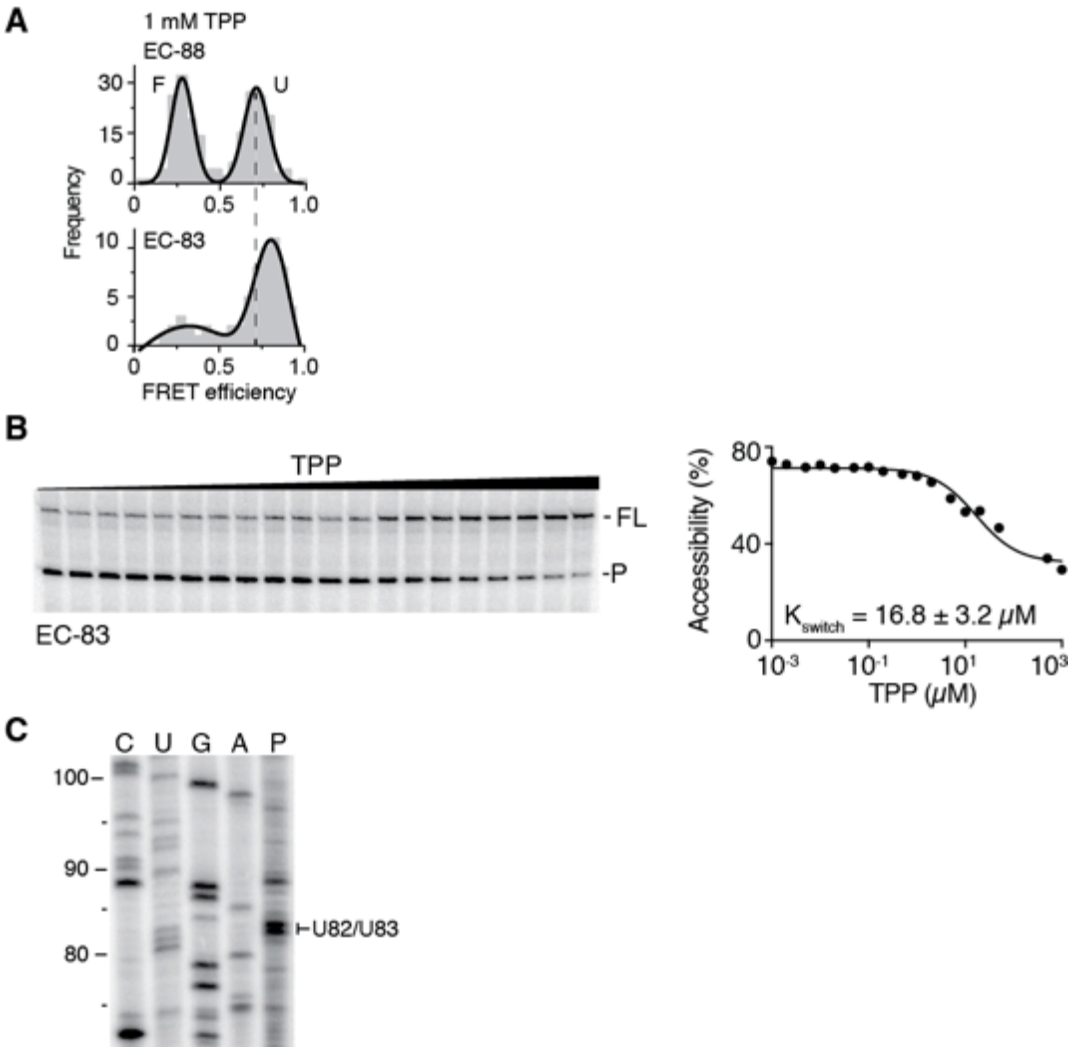


Figure S25. smFRET and biochemical analyses of the EC-83.

(A) smFRET histograms of EC-88 and EC-83 obtained in the presence of 20 mM $MgCl_2$ and 1 mM TPP. The dotted line shows that the U state is not present in EC-83. **(B)** K_{switch} determination of EC-83 using RNase H degradation assays performed in the presence of increasing TPP concentrations ranging from 1 nM to 1 mM. The full-length (FL) and cleaved products (P) are indicated on the right of the gel. Quantification and fitting analysis of the data are shown to the right. **(C)** Mapping of the 83-nt transcriptional intermediate done in the absence of TPP. Transcription reactions were done in the presence of 25 μM NTP and stopped after 45 s (P). Pause sites were mapped by using 3'-O-Methyl NTP in the transcription elongation mix.

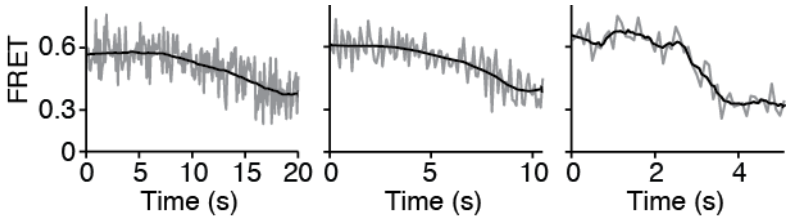


Figure S26. Representative single-molecule U F transitions associated to the 'late binding' t_2 group.

A gradual decrease in FRET efficiency was observed over a heterogeneous time scale ranging from several seconds (left and middle panels) to ~1 s (right panel).

Table S1. K_{switch} values determined for different constructs.

Template	K_{switch}
Elongation complexes	
EC-83	16770 nM ± 3200 nM
EC-88	140 nM ± 22 nM
EC-104	730 nM ± 180 nM
EC-117	362 nM ± 80 nM
EC-136	9970 nM ± 380 nM
EC-136 (anti-P1 stem)	315 nM ± 54 nM
Run-off transcripts	
88-nt	227 nM ± 13 nM
88-nt (1-Cy3)	218 nM ± 25 nM
88-nt (1-Cy5)	521 nM ± 72 nM
117-nt	8690 nM ± 495 nM

Table S2. K_{switch} values determined at various NTP concentrations.

Template	[NTP]	K_{switch}
EC-88	1 mM	455 nM ± 20 nM
	250 μM	429 nM ± 113 nM
	50 μM	277 nM ± 19 nM
EC-136	1 mM	5090 nM ± 113 nM
	250 μM	2800 nM ± 141 nM
	50 μM	604 nM ± 81 nM
EC-136 (U23A)	1 mM	4300 nM ± 500 nM
	50 μM	829 nM ± 89 nM

Table S3. Summary of strains or plasmids used in this study.

Strains	Relevant marker	References
EM1055	MG1655 *ΔlacZ* X174	(*93*)
PM1205	*lacI*'::P_{BAD}-*cat*-*sacB*-*lacZ*, mini I tet^R	(*94*)
TPP1	PM1205 lacI′::P_{NAT}-$TbpA_{12cd}$-LacZ	This study
TPP2	PM1205 lacI′::P_{NAT}- LacZ	This study

Table S4. Summary of *lacZ* fusions used in this study.

Strains	Constructs	Oligonucleotides
TPP1	$TbpA_{12cd}$	MSR8-MSR6 (genomic DNA)
TPP2	P_{NAT}-*tbpA*	MSR8-MSR9 (genomic DNA)

Table S5. PCR constructs used for *in vitro* RNA synthesis.

Constructions	Oligonucleotides
***In vitro* transcription assays**	
pLacUV5-tbpA-158	275AL-918AC (genomic DNA)
pLacUV5-tbpA-88	275AL-2440AC (genomic DNA)
pLacUV5-tbpA-117	275AL-2434AC (genomic DNA)
pLacUV5-tbpA-88-extBio	275AL-2439AC (genomic DNA)
pLacUV5-tbpA-117-extBio	275AL-2437AC (genomic DNA)
***In vitro* transcription-translation assays**	
pLacUV5-lacZ	PCR1: 988MG-986MG (genomic DNA) PCR2: 989MG-986MG (PCR1)
pLacUV5-tbpA$_{306cd}$	275AL-2666AC (genomic DNA)
Transcription elongation complexes	
pLacUV5-tbpA-EC83	275AL-1579JFN (genomic DNA)
pLacUV5-tbpA-EC88	275AL-1697AC (genomic DNA)
pLacUV5-tbpA-EC88 (U14C)	275AL-2390AC (genomic DNA)
pLacUV5-tbpA-EC88 (G37C)	PCR1: 275AL-2384AC (genomic DNA) PCR2: 2385AC-1578JFN (genomic DNA) PCR3: 275AL-1578JFN (PCR1-2)
pLacUV5-tbpA-EC88 (loop5)	PCR1: 275AL-2560AC (genomic DNA) PCR2: 2561AC-1578JFN (genomic DNA) PCR3: 275AL-1578JFN (PCR1-2)
pLacUV5-tbpA-EC104	275AL -2394AC (genomic DNA)
pLacUV5-tbpA-EC117	275AL -2108AC (genomic DNA)
pLacUV5-tbpA-EC136	275AL -1575JFN (genomic DNA)
pLacUV5-tbpA-EC500	275AL-2722AC (genomic DNA)
pLacUV5-tbpA-EC1112	275AL-2667AC (genomic DNA)
pLacUV5-tbpA-EC136 (Anti-P1 stem)	PCR1: 275AL-2633AC (genomic DNA) PCR2: 2634AC-1575JFN (genomic DNA) PCR3: 275AL-1575JFN (PCR1-2)
pLacUV5-tbpA-EC104 (U23A)	PCR1: 275AL-2547AC (genomic DNA) PCR2: 2548AC-2394AC (genomic DNA) PCR3: 275AL-2394AC (PCR1-2)
pLacUV5-tbpA-EC136 (U23A)	PCR1: 275AL-2547AC (genomic DNA) PCR2: 2548AC-1575JFN (genomic DNA) PCR3: 275AL-1575JFN (PCR1-2)
T7 RNAP transcription	
3' tbpA-89	1987PTSP-1988PTSP (genomic DNA)

Table S6. Summary of oligonucleotides used in this study.

Oligonucleotides	Sequences 5'-3'
MSR6	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGCACAG CAACAGCAGGGG
MSR8	CGGGTAGCAAAACAGATCGAAGAAGGGGTTGAATCGCAGGCCAGG GGCAACGCAGTATGC
MSR9	TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACCATAGC TGTTTCCTGTGTGAGTGGCACCCCGTTGAGAAC
275AL	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGTTCTCAACGGGGTGCCAC
918AC	CAACAGCAGGGGCAGACATTTTTTTAACAC
923AC	TTTGCACCTC
986MG	AAACCCCTCCGTTTAGAGAGGGGTTATGCTAGTTACATCTGAACTTC AGCCTCCAGTAC
988MG	GGGCTTAAGTATAAGGAGGAAAAAATATGGTCGTTTTACAACGTCC TGACTG
989MG	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGGGCTTAAGTATAAGGAGGAAAAAATATG
1575JFN	5'biotin-AGACATTTTTTTAACACTTTGCACCTC
1578JFN	5'biotin-TGAGAAGGAGCCTCAAATCCC
1579JFN	5'biotin-AGGAGCCTCAAATCCCTTCG
1987PTSP	GGTAATACGACTCACTATAGGGTATTTCTGATGAGGCCTTCGGGCC GAAACGGTGAAAGCCGTAAAATACCC
1988PTSP	GGCCGGGGGGCCGGGGAGCCTCAAATCCCTTCGCCGGCGTTATCCC GATCAGGTTCGACGGGTATTTTACGGCTTTCACCGTTTCGGCC
2108AC	5'biotin-TGCACCTCAAAAAAGAGTGGCAAAGG
2344AC	5'biotin-AGACCACGTTGAAAGATTGGGTTTG
2384AC	GACGGGTATTTTGTCAGCGCACGCG
2385AC	CGCGTGCGCTGACAAAATACCCGTC
2390AC	GGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATAATGTGT GGGTTCTCAACGGGGCGCCACGCGTACG
2394AC	5'biotin-AGAGTGGCAAAGGACTTGAGAAGGA
2434AC	AAAAAGAGTGGCAAAGGACTTGAGAAG
2437AC	AGACCACGTTGAAAGATTGGGTTTGCAAAAAAGAGTGGCAAAGGA CTTGAGA
2439AC	AGACCACGTTGAAAGATTGGGTTTGGAGCCTCAAATCCCTTCGCCG
2440AC	GCCTCAAATCCCTTCGCCGG
2477AC	GACGGGTATT
2547AC	CTCAGCGCACGCGTTCGCGTGGC
2548AC	GCCACGCGAACGCGTGCGCTGAG
2558JG	5'Cy3-GUUCUCAACGGGG-iAzideN-GCCACGCGUACGCGUGCGCUGA GA
2559JG	5'biotin-GGCCGGGGGGCCGGGG

2560AC	CCTTCGCCGGCGGGCGCCGGATCAGGTT
2561AC	AACCTGATCCGGCGCCCGCCGGCGAAGG
2633AC	GTGGCAAAGGACAACTCAAGGAGCCTCAA
2634AC	TTGAGGCTCCTTGAGTTGTCCTTTGCCAC
2666AC	AAACCCCTCCGTTTAGAGAGGGGTTATGCTAGTTCGTGAACTCCAACGTGGTTGC
2667AC	5’Biotin-AGCCGGGAATTAACGGCTGACGGCG
2722AC	5’biotin-GAAGGCGAAGTAGCCATAATCAAACG

References and Notes :

1. R. Perales, D. Bentley, "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell.* **36**, 178–91 (2009).

2. A. Ray-Soni, M. J. Bellecourt, R. Landick, Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85**, 319–347 (2016).

3. T. Pan, T. Sosnick, RNA folding during transcription. *Annu Rev Biophys Biomol Struct.* **35**, 161–175 (2006).

4. A. Serganov, E. Nudler, A decade of riboswitches. *Cell.* **152**, 17–24 (2013).

5. C. J. McManus, B. R. Graveley, RNA structure and the mechanisms of alternative splicing. *Curr. Opin. Genet. Dev.* **21**, 373–9 (2011).

6. A. Roth, R. R. Breaker, The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem.* **78**, 305–334 (2009).

7. K. M. Neugebauer, On the importance of being co-transcriptional. *J. Cell Sci.* **115**, 3865–71 (2002).

8. P. St-Pierre, K. McCluskey, E. Shaw, J. C. Penedo, D. A. Lafontaine, Fluorescence tools to investigate riboswitch structural dynamics. *Biochim Biophys Acta.* **1839**, 1005–1019 (2014).

9. D. G. Vassylyev *et al.*, Structural basis for substrate loading in bacterial RNA polymerase. *Nature.* **448**, 163–8 (2007).

10. K. D. Westover, D. A. Bushnell, R. D. Kornberg, Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell.* **119**, 481–9 (2004).

11. D. Lai, J. R. Proctor, I. M. Meyer, On the importance of cotranscriptional RNA structure formation. *RNA.* **19**, 1461–73 (2013).

12. J. Zhang, R. Landick, A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem. Sci.* **41**, 293–310

(2016).

13. R. Schroeder, A. Barta, K. Semrad, Strategies for RNA folding and assembly. *Nat. Rev. Mol. Cell Biol.* **5**, 908–919 (2004).

14. J. Andrecka *et al.*, Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 135–40 (2008).

15. P. P. Hein *et al.*, RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).

16. A. Chakraborty *et al.*, Opening and closing of the bacterial RNA polymerase clamp. *Science*. **337**, 591–5 (2012).

17. D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, M. S. Gelfand, Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. *J Biol Chem*. **277**, 48949–48959 (2002).

18. J. Yu, T. Schneiders, Tigecycline challenge triggers sRNA production in Salmonella enterica serovar Typhimurium. *BMC Microbiol*. **12**, 195 (2012).

19. A. Chauvier *et al.*, Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat. Commun*. **8**, 13892 (2017).

20. M.-L. Winz, A. Samanta, D. Benzinger, A. Jäschke, Site-specific terminal and internal labeling of RNA by poly(A) polymerase tailing and copper-catalyzed or copper-free strain-promoted click chemistry. *Nucleic Acids Res*. **40**, e78 (2012).

21. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat Methods*. **5**, 507–516 (2008).

22. N. Kulshina, T. E. Edwards, A. R. Ferre-D'Amare, Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA*. **16**, 186–196 (2010).

23. A. Haller, R. B. Altman, M. F. Souliere, S. C. Blanchard, R. Micura, Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A*. **110**, 4188–4193 (2013).

24. M. H. Buckstein, J. He, H. Rubin, Characterization of nucleotide pools as a function of physiological state in Escherichia coli. *J Bacteriol*. **190**, 718–726 (2008).

25. K. L. Frieda, S. M. Block, Direct observation of cotranscriptional folding in an adenine riboswitch. *Science*. **338**, 397–400 (2012).

26. X. Liu, W. L. Kraus, X. Bai, Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci*. **40**, 516–525 (2015).

27. M. P. Caron *et al*., Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A*. **109**, E3444-53 (2012).

28. E. Nudler, I. Gusarov, G. Bar-Nahum, Methods of walking with the RNA polymerase. *Methods Enzym*. **371**, 160–169 (2003).

29. J. F. Lemay, J. C. Penedo, R. Tremblay, D. M. Lilley, D. A. Lafontaine, Folding of the adenine riboswitch. *Chem Biol*. **13**, 857–868 (2006).

30. J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res*. **15**, 8783–8798 (1987).

31. G. Di Tomasso, A. Salvail-Lacoste, J. Bouvette, J. G. Omichinski, P. Legault, in *Methods in enzymology* (2014; http://www.ncbi.nlm.nih.gov/pubmed/25432744), vol. 549, pp. 49–84.

32. P. A. Dalgarno *et al*., Single-molecule chemical denaturation of riboswitches. *Nucleic Acids Res*. **41**, 4253–4265 (2013).

33. B. Heppell *et al*., Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem Biol*. **7**, 384–392 (2011).

34. S. D. Chandradoss *et al*., Surface passivation for single-molecule protein studies. *J. Vis. Exp*. (2014), doi:10.3791/50549.

35. T. D. Craggs, R. D. Hutton, A. Brenlla, M. F. White, J. C. Penedo, Single-molecule characterization of Fen1 and Fen1/PCNA complexes acting on flap

substrates. *Nucleic Acids Res*. **42**, 1857–72 (2014).

36. C. E. Aitken, R. A. Marshall, J. D. Puglisi, An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J*. **94**, 1826–35 (2008).

37. T. Cordes, J. Vogelsang, P. Tinnefeld, On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc*. **131**, 5018–9 (2009).

38. S. A. McKinney, C. Joo, T. Ha, Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys J*. **91**, 1941–1951 (2006).

39. S. Blouin, D. A. Lafontaine, A loop-loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA*. **13**, 1256–1267 (2007).

40. J. F. Lemay *et al*., Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms. *PLoS Genet*. **7**, e1001278 (2011).

41. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature*. **441**, 1167–1171 (2006).

42. T. E. Edwards, A. R. Ferre-D'Amare, Crystal Structures of the Thi-Box Riboswitch Bound to Thiamine Pyrophosphate Analogs Reveal Adaptive RNA-Small Molecule Recognition. *Structure*. **14**, 1459–1468 (2006).

43. T. Pan, X. Fang, T. Sosnick, Pathway modulation, circular permutation and rapid RNA folding under kinetic control. *J. Mol. Biol*. **286**, 721–31 (1999).

44. S. L. Heilman-Miller, S. A. Woodson, Perturbed folding kinetics of circularly permuted RNAs with altered topology. *J. Mol. Biol*. **328**, 385–94 (2003).

45. S. L. Heilman-Miller, S. A. Woodson, Effect of transcription on folding of the Tetrahymena ribozyme. *RNA*. **9**, 722–33 (2003).

46. M. Chamberlin, J. Ring, Characterization of T7-specific ribonucleic acid polymerase. 1. General properties of the enzymatic reaction and the template

specificity of the enzyme. *J. Biol. Chem.* **248**, 2235–44 (1973).

47. B. T. Lewicki, T. Margus, J. Remme, K. H. Nierhaus, Coupling of rRNA transcription and ribosomal assembly in vivo. Formation of active ribosomal subunits in Escherichia coli requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* **231**, 581–93 (1993).

48. M. Y. Chao, M. C. Kan, S. Lin-Chao, RNAII transcribed by IPTG-induced T7 RNA polymerase is non-functional as a replication primer for ColE1-type plasmids in Escherichia coli. *Nucleic Acids Res.* **23**, 1691–5 (1995).

49. T. N. Wong, T. R. Sosnick, T. Pan, Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci U S A.* **104**, 17995–18000 (2007).

50. G. A. Perdrizet 2nd, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci U S A.* **109**, 3323–3328 (2012).

51. S. M. Uptain, C. M. Kane, M. J. Chamberlin, Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem.* **66**, 117–172 (1997).

52. I. Artsimovitch, V. Svetlov, L. Anthony, R. R. Burgess, R. Landick, RNA polymerases from Bacillus subtilis and Escherichia coli differ in recognition of regulatory signals in vitro. *J Bacteriol.* **182**, 6027–6035 (2000).

53. M. Kashlev *et al.*, Histidine-tagged RNA polymerase: dissection of the transcription cycle using immobilized enzyme. *Gene.* **130**, 9–14 (1993).

54. H. C. Kolb, M. G. Finn, K. B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed. Engl.* **40**, 2004–2021 (2001).

55. R. Huisgen, R. Knorr, L. Moebius, G. Szeimies, Einige Beobachtungen zur Addition organischer Azide an CC-Dreifachbindungen. *Chem. Ber.* **98**, 4014–4021 (1965).

56. R. Huisgen, L. Mobius, G. Szeimies, Kinetik der Additionen organischer Azide an CC-Mehrfachbindungen. *Chem. Ber.* **100**, 2494–2507 (1967).

57. K. Uchida, S. Kawakishi, Selective Oxidation of Imidazole Ring in Histidine Residues by the Ascorbic Acid - Copper Ion System. *Biochem. Biophys. Res. Commun.* **138**, 659–665 (1986).

58. E. Paredes, S. R. Das, Click chemistry for rapid labeling and ligation of RNA. *ChemBioChem.* **12**, 125–131 (2011).

59. S. I. Presolski, V. P. Hong, M. G. Finn, Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **3**, 153–162 (2011).

60. Q. Wang *et al.*, Bioconjugation by Copper (I) -Catalyzed Azide-Alkyne [3+2] Cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3193 (2003).

61. N. J. Agard, J. A. Prescher, C. R. Bertozzi, A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).

62. I. Nikić, J. H. Kang, G. E. Girona, I. V. Aramburu, E. a Lemke, Labeling proteins on live mammalian cells using click chemistry. *Nat. Protoc.* **10**, 780–791 (2015).

63. X. Ren, A. H. El-Sagheer, T. Brown, *Nucleic Acids Res.*, in press, doi:10.1093/nar/gkw028.

64. A. A. Sawant *et al.*, A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **44**, e16 (2016).

65. F. Li *et al.*, A Covalent Approach for Site-Specific RNA Labeling in Mammalian Cells. *Angew. Chemie Int. Ed.* **54**, 4597–4602 (2015).

66. H. Rao, A. a Tanpure, A. a Sawant, S. G. Srivatsan, Enzymatic incorporation of an azide-modified UTP analog into oligoribonucleotides for post-transcriptional chemical functionalization. *Nat. Protoc.* **7**, 1097–112 (2012).

67. H. Rao, A. a. Sawant, A. a. Tanpure, S. G. Srivatsan, Posttranscriptional chemical functionalization of azide-modified oligoribonucleotides by bioorthogonal click and Staudinger reactions. *Chem. Commun.* **48**, 498 (2012).

68. D. M. Lilley, T. J. Wilson, Fluorescence resonance energy transfer as a structural tool for nucleic acids. *Curr Opin Chem Biol*. **4**, 507–517 (2000).

69. S. Thore, M. Leibundgut, N. Ban, Structure of the Eukaryotic Thiamine Pyrophosphate Riboswitch with Its Regulatory Ligand. *Science (80-.).*, 1208–1211 (2006).

70. K. Lang, R. Rieder, R. Micura, Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res*. **35**, 5370–5378 (2007).

71. M. Mandal, R. R. Breaker, Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*. **5**, 451–463 (2004).

72. F. J. Grundy, S. C. Lehman, T. M. Henkin, The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci U S A*. **100**, 12057–12062 (2003).

73. A. Lussier, L. Bastet, A. Chauvier, D. A. Lafontaine, A Kissing Loop Is Important for btuB Riboswitch Ligand Sensing and Regulatory Control. *J Biol Chem*. **290**, 26739–26751 (2015).

74. Y. Lu, W. Wang, M. W. Kirschner, Specificity of the anaphase-promoting complex: A single-molecule study. *Science (80-.)*. **348**, 1248737–1248737 (2015).

75. M. F. Soulière *et al*., Tuning a riboswitch response through structural extension of a pseudoknot. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3256-64 (2013).

76. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*. **89**, 33–41 (1997).

77. V. K. Duesterberg, I. T. Fischer-Hwang, C. F. Perez, D. W. Hogan, S. M. Block, Observation of long-range tertiary interactions during ligand binding by the TPP riboswitch aptamer. *Elife*. **4** (2015), doi:10.7554/eLife.12362.

78. M. C. Murphy, I. Rasnik, W. Cheng, T. M. Lohman, T. Ha, Probing single-

stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **86**, 2530–7 (2004).

79. L. D. Sherlin *et al.*, Chemical and enzymatic synthesis of tRNAs for high-throughput crystallization. *RNA.* **7**, 1671–1678 (2001).

80. K. Lang, R. Micura, The preparation of site-specifically modified riboswitch domains as an example for enzymatic ligation of chemically synthesized RNA fragments. *Nat. Protoc.* **3**, 1457–66 (2008).

81. J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell.* **18**, 49–60 (2005).

82. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, S. M. Block, Direct observation of base-pair stepping by RNA polymerase. *Nature.* **438**, 460–5 (2005).

83. J. Boyle, G. T. Robillard, S. H. Kim, Sequential folding of transfer RNA. A nuclear magnetic resonance study of successively longer tRNA fragments with a common 5' end. *J. Mol. Biol.* **139**, 601–25 (1980).

84. F. R. Kramer, D. R. Mills, Secondary structure formation during RNA synthesis. *Nucleic Acids Res.* **9**, 5109–24 (1981).

85. S. L. Brehm, T. R. Cech, Fate of an intervening sequence ribonucleic acid: excision and cyclization of the Tetrahymena ribosomal ribonucleic acid intervening sequence in vivo. *Biochemistry.* **22**, 2390–7 (1983).

86. T. Pan, I. Artsimovitch, X. W. Fang, R. Landick, T. R. Sosnick, Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci U S A.* **96**, 9545–9550 (1999).

87. N. Komissarova, M. L. Kireeva, J. Becker, I. Sidorenkov, M. Kashlev, Engineering of elongation complexes of bacterial and yeast RNA polymerases. *Methods Enzymol.* **371**, 233–51 (2003).

88. C. Y. Jao, A. Salic, Exploring RNA transcription and turnover in vivo by using

click chemistry. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15779–84 (2008).

89. F. Werner, D. Grohmann, Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9**, 85–98 (2011).

90. J. F. Kugel, J. A. Goodrich, In vitro studies of the early steps of RNA synthesis by human RNA polymerase II. *Methods Enzymol.* **370**, 687–701 (2003).

91. H. Kettenberger, K.-J. Armache, P. Cramer, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell.* **16**, 955–65 (2004).

92. E. di Mauro *et al.*, Rifampicin sensitivity of the components of DNA-dependent RNA polymerase. *Nature.* **222**, 533–7 (1969).

93. E. Masse, S. Gottesman, A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc Natl Acad Sci U S A.* **99**, 4620–4625 (2002).

94. P. Mandin, S. Gottesman, A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Mol Microbiol.* **72**, 551–565 (2009).

CONCLUSION

Les travaux présentés dans cette thèse avaient pour but de mieux comprendre l'importance du processus de transcription sur le repliement des riborégulateurs et son impact sur le résultat de l'expression génique. Plus spécifiquement, ces études ont été réalisées sur deux riborégulateurs liant le TPP, pour lesquels il existait peu d'évidences expérimentales à leur sujet.

Dans un premier temps, mes travaux ont permis de caractériser le mécanisme employé par le riborégulateur *thiC* pour réguler l'expression génique. Ce riborégulateur inhibe l'initiation de la traduction et est également la cible de la protéine Rho en présence du ligand. Ce résultat vient renforcer l'idée que les riborégulateurs sont parfaitement adaptés pour réguler finement l'expression génique. Même si le riborégulateur utilise deux mécanismes de régulation distincts, la cinétique de transcription est déterminante pour permettre une liaison efficace du ligand. Cette observation est confortée par les expériences de traduction *in vitro* puisque que le TPP n'est pas capable d'induire une répression traductionnelle si il n'est pas présent pendant le processus de transcription. De plus, les expériences de sondage de l'ARN en complexe transcriptionnel ont permis de mettre en évidence une structure transitoire (structure Anti-P1) qui définie directement la capacité du riborégulateur à lier efficacement le ligand. Dans le cas précis d'un complexe transcriptionnel arrêté au niveau de la pause dans la région d'initiation de la traduction (position 187), les essais de dégradation à la RNase H et de SHAPE montrent clairement l'existence d'une telle structure et son impact négatif sur la liaison subséquente du ligand. Enfin, l'existence d'une telle pause pour les autres riborégulateurs présents chez *E. coli* suggère que ce mécanisme serait plus largement retrouvé au sein des riborégulateurs employant le mécanisme de régulation traductionnel.

La suite de mes travaux a principalement permis de développer une nouvelle méthode afin d'étudier la dynamique structurale de l'ARN en complexe transcriptionnel. En effet, grâce à l'utilisation du barrage biotine-streptavidine, nous avons été en mesure

d'observer le changement de conformation du riborégulateur *tbpA* au cours de la transcription et ce, à l'échelle de la molécule unique par smFRET. L'analyse de la dynamique de repliement de l'aptamère démontre que celui-ci lie le TPP par un mécanisme de capture de conformation comme il est proposé dans la littérature (Haller et al., 2013). Cette technique de marquage de l'ARN en conditions natives nous a permis de démontrer l'importance de l'ARNp pour le repliement du riborégulateur. Celle-ci va directement influer sur l'affinité du riborégulateur pour le TPP en stabilisant une autre structure Anti-P1, comme pour le riborégulateur *thiC*. Enfin les expériences de transcription à différentes vitesses et en temps réel, nous permettent d'établir un modèle cinétique de repliement. Dans ce modèle, la liaison du ligand est directement en compétition avec la formation de l'Anti-P1 et le «vainqueur» sera déterminé par la vitesse de transcription.

Finalement, ces travaux ont permis de décrypter l'influence des pauses transcriptionnelles sur le repliement des riborégulateurs et de faire le lien avec la fonction régulatrice de l'ARN. Les résultats obtenus établissent un modèle général de repliement pour les riborégulateurs employant l'inhibition de la traduction pour réguler l'expression génique. Ce mécanisme de repliement étant strictement dépendant de la cinétique de transcription, il souligne ainsi l'importance du processus de transcription pour étudier la structure des riborégulateur (Batey, 2015; Pan and Sosnick, 2006). Enfin, les méthodes développées et employées constituent de nouveaux outils pour étudier ce type d'ARN et sont actuellement employées au laboratoire pour réaliser des recherches similaires à celles que j'ai réalisées.

DISCUSSION ET PERSPECTIVES

Les études réalisées pendant mon doctorat ont contribué à une meilleure compréhension des réarrangements structuraux qui se produisent pendant la transcription des riborégulateurs. Cependant, ce travail soulève également d'autres questions quant à l'importance du contexte transcriptionnel pour décrypter les mécanismes de régulation employés par les riborégulateurs. Certains aspects du processus de transcription ont particulièrement attiré mon attention et seront discutés dans cette section.

Les autres pauses identifiées chez les riborégulateurs

L'impact des pauses de l'ARNp a été largement étudié dans tous les domaines du vivant et leur implication dans la régulation de l'expression génique n'est plus à démontrer (Jonkers and Lis, 2015; Zhang and Landick, 2016). L'importance de telles pauses a déjà été évaluée dans le contexte des riborégulateurs transcriptionnels (Lemay et al., 2011; Wickiser et al., 2005). Ces pauses permettent de laisser plus de temps au ligand pour lier l'aptamère avant que la plateforme d'expression ne se forme. Dans un exemple plus récent, une pause permet à la protéine Rho de rattraper l'ARNp pour effectuer sa fonction terminatrice (Hollands et al., 2014). Ces travaux ont permis de souligner la fonction de ces pauses mais il serait légitime d'imaginer d'autres fonctions associées à ce processus transcriptionnel. Au cours de notre étude du riborégulateur *thiC*, nous avons également mis en évidence la position des pauses transcriptionnelles au sein de tous les riborégulateurs chez *E. coli*. Cette identification ne se limite pas seulement aux pauses dans la région d'initiation de la traduction, et d'autres pauses identifiées dans l'aptamère ou la plateforme d'expression méritent une attention particulière. Il serait intéressant de déterminer les signaux dans la séquence d'ARN qui conduisent à ces pauses de l'ARNp ainsi que leurs impacts sur le fonctionnement des riborégulateurs.

D'un point de vue structural, si une pause est nécessaire pour obtenir la structure finale de l'aptamère qui sera compétente pour lier le ligand, alors l'étude de ces

réarrangements de conformation pendant la transcription prend tout son sens comme pour le cas de la RNase P (Wong et al., 2007).

D'un point de vue génétique, ces pauses pourraient également permettre le recrutement de facteurs accessoires participant au mécanisme de régulation génique comme la protéine Rho par exemple. En effet, il est à présent clair que Rho effectue son activité terminatrice lorsque l'ARNp est pausée au site de terminaison (Dutta et al., 2008), le signal de chargement de la protéine sur l'ARN est par contre beaucoup moins connu. Il serait ainsi possible de spéculer que des pauses situées plus en amont du site de terminaison favorisent l'accessibilité du site *rut* par réarrangement structural (Kriner and Groisman, 2017) ou laissent plus de temps à Rho pour lier l'ARN naissant et débuter son activité hélicase.

Les facteurs de transcription

L'une des principales conclusions de mes travaux est que la présence de l'ARNp en cours de transcription va moduler la structure du riborégulateur en protégeant physiquement l'ARN et en empêchant la formation de structures clés pour la régulation.

Cet effet de protection de l'enzyme n'était pas surprenant puisque l'empreinte de l'ARNp sur la matrice d'ADN est connue depuis plusieurs années et que l'effet «chaperonne» de l'enzyme a déjà été démontré (Nudler et al., 1996; Pan and Sosnick, 2006). Cependant, comme il a été mentionné dans la section introduction, le complexe transcriptionnel ne se limite pas à l'ARNp transcrivant l'ADN, puisque les facteurs de transcription font partie intégrante de ce complexe. Cette affirmation est renforcée par les travaux de l'équipe du Dr. Landick dans lesquels il a été montré que des facteurs de transcription comme NusA ou Rho pouvaient se lier à l'ARNp dès l'initiation de la transcription et perdurer jusqu'à la terminaison (Mooney et al., 2009).

L'implication de tels facteurs dans le repliement de l'ARN est aujourd'hui encore peu connue même si l'effet de NusA au niveau des pauses de classe I est en partie expliqué par son interaction avec la boucle de la tige (Toulokhonov and Landick, 2003; Toulokhonov et al., 2001). La reconstitution d'un complexe transcriptionnel *in vitro,*

incluant les facteurs de transcription, permettrait d'observer la structure de l'ARN dans des conditions se rapprochant le plus possible de celles présentes dans la cellule.
Enfin nous avons démontré que la vitesse de transcription détermine l'efficacité de liaison du ligand des riborégulateurs traductionnels. Ces riborégulateurs fonctionnent donc sous un régime cinétique et les facteurs modulant la vélocité de l'ARNp auront également un impact (même indirect) sur la régulation de l'expression génique. Afin de mieux mesurer l'importance de ces facteurs, il est donc nécessaire de les considérer de façon indépendante pour chacun des riborégulateurs et dans un système purifié. Ainsi, dans le cas des riborégulateurs traductionnels, l'utilisation des essais de dégradation à la RNase H pour évaluer leur affinité en fonction des conditions de transcription, constitue un outil particulièrement puissant pour répondre à ces questions (Lussier et al., 2015). Cette technique est relativement simple à employer et la méthode est présentée en Annexe (Chauvier and Lafontaine, 2015).

Le ribosome et la traduction

Pendant mon doctorat, j'ai exclusivement travaillé avec des riborégulateurs qui emploient le mécanisme de régulation traductionnel. Nos connaissances sur ce mécanisme restent limitées à la caractérisation d'une tige qui séquestre la région d'initiation de la traduction. Chez les procaryotes, la traduction et la transcription sont couplées conduisant ainsi à la formation d'un complexe que l'on appelle «expressome» qui comporte le ribosome ancré sur la séquence SD suivi de l'ARNp en cours de transcription (Kohler et al., 2017). Il serait ainsi intéressant de caractériser la structure d'un riborégulateur dans le contexte de l'expressome pour évaluer l'impact de la machinerie de traduction sur la structure de la plateforme d'expression. L'ARN protégé par les machineries de traduction et de transcription combinées pourrait correspondre à l'entièreté de petits riborégulateurs comme par exemple celui liant le PréQ1, suggérant que le premier ribosome lié aura un impact irréversible sur une liaison post-transcriptionnelle du ligand.
La dynamique de liaison du ribosome sur sa séquence SD et le lien avec la structure de l'ARN sont également des sujets sur lesquels nous pourrions porter attention. La

méthode de transcription fluorescente que j'ai développée pourrait permettre d'observer la dynamique de repliement de l'ARN en présence du ribosome ou bien encore d'observer en temps réel la liaison d'un ribosome fluorescent (Dorywalska et al., 2005). L'objectif sous-jacent serait d'observer l'effet de la liaison du ligand sur la dynamique de liaison du ribosome afin de déterminer avec précision le mécanisme d'inhibition de la traduction. Une récente étude de l'équipe du Dr. Walter a mimé la présence du ribosome sur l'ARN grâce à des oligo-nucléotides d'ADN qui s'hybrident sur la région SD (Rinaldi et al., 2016). Cette analyse à l'échelle de la molécule unique suggère que la présence du ligand diminue le temps pendant lequel la séquence SD est accessible pour lier le ribosome, posant ainsi les bases de ce domaine de recherche.

Mes travaux ont aussi mis en évidence une régulation à deux niveaux pour le riborégulateur *thiC* en présence du TPP. En effet, la présence du ligand inhibe l'initiation de la traduction et conduit également à une terminaison de la transcription dépendante de la protéine Rho. Ce dernier mécanisme est direct puisque la régulation au niveau de l'ARN n'est pas dépendante de l'inhibition de l'initiation de la traduction (voir section 3.4.3 de l'introduction). Ce résultat suggère que la régulation traductionnelle se produirait seulement dans le cas où la terminaison de la transcription n'a pas pu se produire, agissant comme un contrôle supplémentaire pour réguler l'expression génique à plusieurs niveaux comme dans le cas d'une liaison tardive du TPP par exemple.

Le contexte cellulaire

Depuis la première détection d'une pause transcriptionnelle *in vivo* (Landick et al., 1987), le développement de nouvelles techniques comme le NET-seq (pour *nascent elongating transcript sequencing* en anglais), ont permis de répertorier un nombre important de site de pause (Larson et al., 2014; Vvedenskaya et al., 2014). La comparaison de la banque de donnée de site de pause avec les résultats obtenus *in vitro* mène à plusieurs hypothèses et avenues futures.

Premièrement, certaines pauses comme celles dans la région d'initiation de la traduction des riborégulateurs *thiC, thiM* et *lysC,* sont retrouvées, suggérant leur importance pour le couplage transcription-traduction (Chauvier et al., 2017; Larson et al., 2014). Une étude plus approfondie de ces pauses peut être envisagée directement dans la bactérie pour établir ce lien. En effet, la constitution d'un blocage transcriptionnel *in vivo* (Proshkin et al., 2010) au niveau de ces pauses permettrait de tester cette hypothèse dans différentes conditions comme en inhibant la traduction artificiellement avec un antibiotique ou bien encore en présence du ligand.

Ensuite, certaines pauses comme celles du riborégulateur *btuB* sont seulement observées *in vitro* (Chauvier et al., 2017; Perdrizet et al., 2012) suggérant l'intervention d'autres facteurs pendant la transcription dans la bactérie. L'identification de tels facteurs grâce à des criblages *in vitro* permettrait de comprendre cette différence et d'en apprendre d'avantage sur la régulation de ce riborégulateur.

Enfin, il est important de noter que ces expériences *in vivo* se réalisent majoritairement en milieu riche ce qui ne permet pas de comparer des conditions en présence et en absence de notre ligand. La détection des pauses trancriptionnelles est aussi possible grâce à la technique d'empreinte au permanganate de potassium (Sasse-Dwight and Gralla, 1989, 1991) qui oxyde les thymines de l'ADN dans la bulle de transcription. Ainsi en changeant les conditions de cultures il serait possible de conduire une analyse *in vivo* des pauses transcriptionnelles pour les riborégulateurs.

Etendre le domaine d'analyse

La transcription par l'ARNp bactérienne est connue et maîtrisée depuis de nombreuses années et a considérablement contribué à mieux comprendre les différents processus régulant la synthèse de l'ARN. En laboratoire, les éléments nécessaires pour transcrire un ARN bactérien donné sont faciles d'accès et peu couteux, ce qui permet l'obtention d'un nombre considérable de données rapidement. Même si la transcription eucaryote est également possible *in vitro* (Cheng and Price, 2009), celle-ci nécessite plus de facteurs accessoires pour initier le processus de synthèse d'ARN et l'ARNp II qui est majoritairement utilisée n'est pas disponible commercialement. Néanmoins, des études

ont montré que les ARNp eucaryotes et procaryotes pouvaient répondre aux mêmes signaux au cours de la transcription comme le consensus de pauses, suggérant qu'il serait possible d'étudier un ARN eucaryote de la même manière dont j'ai procédé pendant mes travaux (Larson et al., 2014; Walter et al., 2003; Zamft et al., 2012). Partant de ce principe, il serait particulièrement intéressant de vérifier si le barrage biotine-streptavidine bloque la progression de l'ARNp I, II ou III comme nous avons observé pour l'ARNp d'*E. coli*. Une telle propriété permettrait d'analyser le repliement de l'ARN en population ou à l'échelle de la molécule unique pour des transcrits eucaryotes comme par exemples les riborégulateurs TPP chez *Arabidopsis thaliana*, qui sont un peu plus proches de mon domaine d'expertise. Les récentes études à l'échelle du génome ont mis en évidence que l'épissage alternatif se produit pendant la transcription et préférentiellement lorsque l'ARNp est pausée (Bentley, 2014). Ainsi, le développement de nouveaux outils pour analyser ce processus directement en laboratoire constituerait une avancée majeure pour élucider les paramètres déterminant ce mécanisme.

«Science sans conscience, n’est que ruine de l'âme»
- Rabelais - Pentagruel (1542)

ANNEXE

Chauvier, A., and Lafontaine, D.A. (2015). Probing of Nascent Riboswitch Transcripts. Methods Mol. Biol. *1334*, 109–114.

Probing of native RNA molecules

Adrien Chauvier and Daniel A. Lafontaine*

Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Quebec, Canada, J1K 2R1.

Keywords: Riboswitch, RNA folding, helical junction, metabolite-sensing
Running title: Riboswitch folding strategies are tuned for ligand sensing

*To whom correspondence should be addressed.

Address correspondence to: Daniel Lafontaine, 2500, Blvd Université, Sherbrooke, Quebec, Canada, J1K 2R1.
E-mail: Daniel.Lafontaine@USherbrooke.ca

Summary

The study of biologically significant and native structures is vital to characterize RNAbased regulatory mechanisms. Riboswitches are cis-acting RNA molecules that are involved in the biosynthesis and transport of cellular metabolites. Because riboswitches regulate gene expression by modulating their structure, it is vital to employ native probing assays to determine how native riboswitch structures perform highly efficient and specific ligand recognition. By employing RNase H probing, it is possible to determine the accessibility of specific RNA domains in various structural contexts. Herein, we describe how to employ RNase H probing to characterize nascent mRNA riboswitch molecules as a way to obtain information regarding the riboswitch regulation control mechanism.

Keywords: RNA structure, Nascent mRNA, Native structures, RNase H probing.

1. INTRODUCTION

Riboswitches are RNA molecules that are involved in the control of transport and/or biosynthesis of cellular metabolites (1, 2). They control various genetic levels such as transcription, translation, mRNA decay and splicing (1, 2). Riboswitches are composed of an aptamer domain and an expression platform, which perform ligand recognition and genetic regulation, respectively (1, 2). Upon metabolite binding, riboswitches experience a variety of structural changes important for the modulation of gene expression. Recently, it was shown that ligand binding to the *pbuE* adenine riboswitch exclusively occurred cotranscriptionally (3, 4, 5), suggesting that the intrinsic polarity of the transcription process is of primordial importance for correct riboswitch folding. Transcription was also previously shown to be important for the folding of *Tetrahymena* group I intron (6) and RNase P (7), indicating that co-transcriptional folding can be highly important for the outcome of genetic expression. Given that the relevance of co-transcriptional folding may only be emphasized within systems allowing *in vitro* reconstitution assays, it is highly probable that the general importance of co-transcriptional folding is underestimated by a large degree.

Most probing assays interrogating the structure of RNA molecules (or sub-domains) rely on empirically-established renaturation protocols involving variations in temperature and/or salts. By undergoing these various controlled steps, it is expected that RNA molecules are adopting structures that are favored in defined conditions. However, such conditions-dependent enriched RNA structures may not always represent the native structure required for biological activity. It is therefore important to choose experimental conditions where nascent mRNA molecules can be studied using structural probing.

Ribonuclease H is a well-known non-specific endonuclease that catalyzes RNA via a hydrolytic mechanism (8) . In this assay, a DNA oligonucleotide is used to probe the relative accessibility of a region in a given RNA molecule. While a single-stranded region can be targeted by RNase H cleavage, a double-stranded or highly structured region does not allow DNA binding and is thus protected from cleavage. Because RNase H is active in a large range of experimental conditions (ref), it can be used to probe RNA structural changes as a function of various stimuli. For example, RNase H was recently

used to characterize the folding pathway of the B12 riboswitch where it was determined that transcriptional pausing is important for riboswitch activity (9) . In this chapter, we describe how RNase H can be used to characterize the structure of native riboswitch molecules and how this can be employed to deduce biologically-relevant information.

2. MATERIALS

2.1 DNA templates required for in vitro transcription

The DNA template was made by PCR amplification on the genomic DNA from *Escherichia coli* strain MG1655-K12 and consisting of a lacUV5 promoter fused to a riboswitch sequence (*thiM* riboswitch, Figure 1) followed by a portion of the open reading frame (11 codons). According to the model, the *thiM* riboswitch allows ribosome translation in absence of ligand. However, ligand binding causes a structural reorganization that sequesters the AUG start codon, thus inhibiting riboswitch translation initiation. The riboswitch DNA template sequence is the following (sequence written 5' to 3' polarity) :

gggcaccccaggctttacactttatgcttccggctcgtataatgtgtggCTGCGATTTATCATCGCAACCAA
ACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTG
ATCTGGATAATGCCAGCGTAGGGAAGTCACGGACCACCAGGTCATTGCTTCT
TCACGTTATGGCAGGAGCAAACTATGCAAGTCGACCTGCTGGGTTCAGCGCA
ATCT

The promoter is shown in lowercase and the ATG start codon is underlined.

2.2 In vitro transcription of RNA strands under single-round conditions

1. 5X transcription buffer (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 100 mM $MgCl_2$ and 500 µM EDTA, 70 mM ß-Mercaptoethanol) (see **Note 1**).
2. 100 µM RNA tetranucleotide (5'-CUGC-3').
3. 2.5 µM ATP and GTP.
4. 1 mM rNTP.
5. 2 µCi [alpha-^{32}P] UTP.
6. 300 fmol/µL DNA template.
7. 6mg/mL heparin, RNA polymerase holoenzyme from Escherichia coli (Epicentre).

8. 200 µM thiamin pyrophosphate (TPP).

2.3. RNase H cleavage assays

1. 1X cleavage buffer (5 mM Tris-HCl pH 8, 20 mM $MgCl_2$, 100 mM KCl, 50 µM EDTA and 9 mM ß-Mercaptoethanol) (see **Note 1**).

2. RNase H enzyme (Ambion-10U/µL).

3. The DNA oligonucleotide (100 µM) should target a region that is informative for the study. In the present case, the oligonucleotide is targeting the aptamer region of the *thiM* riboswitch that should be protected from RNase H cleavage upon TPP binding (**Figure 1**). The sequence of the DNA oligonucleotide is the following (sequence written 5' to 3' polarity) : CCGAGTCGTT.

4. 2X stop solution : 95% formamide, 18 mM EDTA and 0.02% SDS.

Figure 1

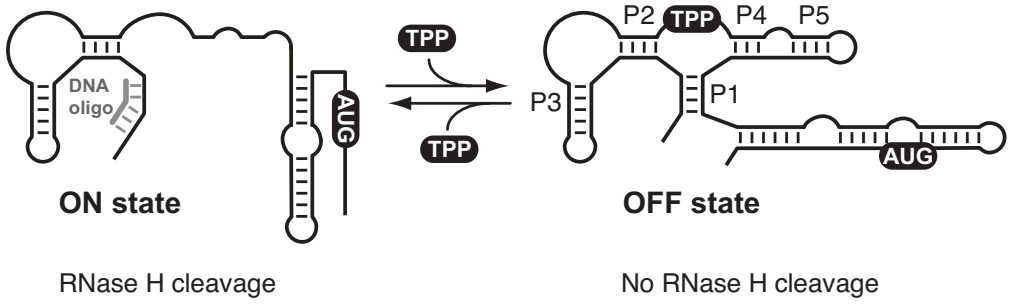


Figure 1. **Schematic representation of the E. coli *thiM* riboswitch.** In this model, the riboswitch ON state in adopted in absence of TPP and allows ribosome translation due to the relative accessibility of the AUG start codon. However, in presence of TPP, the riboswitch is reorganized and sequesters the AUG start codon, ultimately leading to translation repression. P1 to P5 helical domains are indicated on the riboswitch. According to this model, a DNA oligonucleotide targeting the P1 stem would only be able to hybridize in absence of TPP.

3. METHODS

Because RNases are prevalent and very hard to inactivate, it is thus important to employ RNase free solutions. Moreover, except for the polymerase, ensure that all reagents have been equilibrated at room temperature before mixing together.

3.1 In vitro transcription of RNA strands under single-round conditions

1. In a total volume of 7.5 µL, add 1.25 µL of ATP/GTP mixture, 0.85 µL CUGC tetranucleotide, 2 µL 5X transcription buffer, 1 µL DNA template and 1 µL of 200 µM TPP when necessary.
2. Incubate the sample 5 min at 37°C (see **Note 2**).
4. In the reaction mixture, add a volume of 2.5 µL containing 10 µCi [alpha-^{32}P] UTP and 0.2 U of E. coli RNA polymerase (homemade or commercially available).
5. Incubate the sample 15 min at 37°C.
6. In the reaction mixture, add a 10 µL solution containing 2 µL transcription buffer, 1 µL rNTPs, 1.5 µL heparin and 5.5 µL water (see Note 3).
7. Incubate the sample 15 min at 37°C.
8. At this point, the presence of heparin is ensuring that transcription is stopped after complete RNA strand synthesis. It is important to avoid adding denaturing agents since this would disrupt native RNA interactions and therefore mislead conclusions.

3.2 RNase H cleavage assays

1. Dilute RNase H just prior to use in ice-cold cleavage buffer to 0.12 U/µL (see **Note 4**).
2. In a fresh tube, incubate 8 µL of the transcription reaction with 2 µL of DNA oligonucleotide.
4. Incubate the reaction mixture for 5 min at 37°C.
3. Add 10 µL of RNase H dilution and incubate for 5 min at 37°C.
4. Quench reaction with 20 µL of 2X stop solution.
5. Load on a denaturing gel to separate uncleaved from cleaved riboswitch transcripts.
6. The result of a typical reaction is shown in **Figure 2**. In this experiment, it can be observed that RNase H cleavage is only observed in absence of TPP, as expected from the riboswitch model in **Figure 1**. The absence of cleavage in presence of TPP indicates

that TPP binding to the aptamer domain resulted in a stable structure that does not allow DNA binding and thus RNase H cleavage.

Figure 2

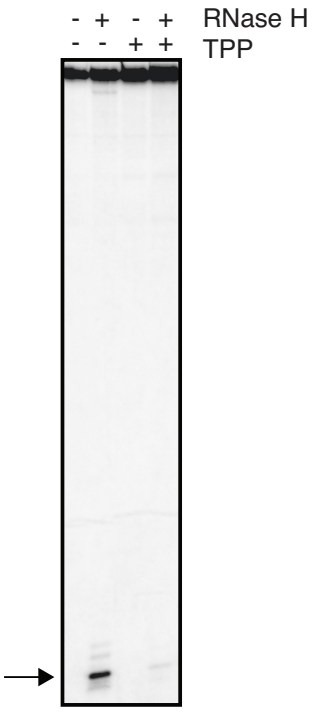


Figure 2. **RNase H cleavage of the *thiM* riboswitch as a function of TPP binding.** The arrow indicates the production of the 5' product due to RNase H cleavage activity in absence of TPP. The cleavage is highly reduced in presence of TPP, indicating that the DNA oligonucleotide does not bind to the riboswitch.

4. NOTES

1. The ß-Mercaptoethanol has to be added just before setting up the transcription reaction.

2. Pre-incubate the mixture at 37°C to allow temperature equilibration.

3. Prepare the mixture and equilibrate at 37°C before using.

4. As the RNase H enzyme is sensitive to freeze-thaw cycles, it is best to prepare a dilution and to equilibrate it at room temperature before using it.

5. ACKNOWLEDGEMENTS

We thank members of the Lafontaine laboratory for critical reading of the manuscript and the National Sciences and Engineering Research Council of Canada (NSERC) for financial support. D.A.L. is a Fonds de Recherche Santé Québec Senior Scholar.

REFERENCES

1. Breaker,R.R. (2012) Riboswitches and the RNA world. Cold Spring Harb Perspect. Biol., 4, 10.1101/cshperspect.a003566.

2. Serganov,A. and Nudler,E. (2013) A decade of riboswitches. Cell, 152, 17-24.

3. Lemay,J.F., Penedo,J.C., Tremblay,R., Lilley,D.M. and Lafontaine,D.A. (2006) Folding of the adenine riboswitch. Chem Biol, 13, 857-68.

4. Lemay,J.F., Desnoyers,G., Blouin,S., Heppell,B., Bastet,L., St-Pierre,P., Masse,E. and Lafontaine,D.A. (2011) Comparative study between transcriptionally- and translationallyacting
adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. PLoS Genet., 7, e1001278.

5. Rieder,R., Lang,K., Graber,D. and Micura,R. (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. Chembiochem, 8, 896-902.

6. Koduvayur,S.P. and Woodson,S.A. (2004) Intracellular folding of the tetrahymena group I intron depends on exon sequence and promoter choice. RNA, 10, 1526-1532.

7. Pan,T., Artsimovitch,I., Fang,X.W., Landick,R. and Sosnick,T.R. (1999) 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.

8. Zarrinkar,P.P. and Williamson,J.R. (1994) Kinetic intermediates in RNA folding. Science, 265, 918-924.

9. Perdrizet,G.A.,2nd, Artsimovitch,I., Furman,R., Sosnick,T.R. and Pan,T. (2012) Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. Proc. Natl. Acad. Sci. U. S. A., 109, 3323-3328.

BIBLIOGRAPHIE

Agard, N.J., Prescher, J.A., and Bertozzi, C.R. (2004). A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. J. Am. Chem. Soc. *126*, 15046–15047.

Al-Hashimi, H.M., and Walter, N.G. (2008). RNA dynamics: it is about time. Curr. Opin. Struct. Biol. *18*, 321–329.

Ames, T.D., and Breaker, R.R. (2011). Bacterial aptamers that selectively bind glutamine. RNA Biol *8*, 82–89.

Ames, T.D., Rodionov, D.A., Weinberg, Z., and Breaker, R.R. (2010). A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. Chem. Biol. *17*, 681–685.
Artsimovitch, I., and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc. Natl. Acad. Sci. U.S.A. *97*, 7090–7095.

Bailey, M.J., Hughes, C., and Koronakis, V. (1996). Increased distal gene transcription by the elongation factor RfaH, a specialized homologue of NusG. Mol. Microbiol. *22*, 729–737.

Baker, J.L., Sudarsan, N., Weinberg, Z., Roth, A., Stockbridge, R.B., and Breaker, R.R. (2012). Widespread genetic switches and toxicity resistance proteins for fluoride. Science *335*, 233–235.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183–193.

Batey, R.T. (2015). Riboswitches: still a lot of undiscovered country. RNA *21*, 560–563.

Belogurov, G.A., Mooney, R.A., Svetlov, V., Landick, R., and Artsimovitch, I. (2009). Functional specialization of transcription elongation factors. EMBO J. *28*, 112–122.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space.

Nat. Rev. Genet. *15*, 163–175.

Bevilacqua, P.C., Ritchey, L.E., Su, Z., and Assmann, S.M. (2016). Genome-Wide Analysis of RNA Secondary Structure. Annu. Rev. Genet. *50*, 235–266.

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., and Heumann, H. (2004). The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. Nat. Struct. Mol. Biol. *11*, 551–557.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. Nature *221*, 43–46.

Burmann, B.M., Schweimer, K., Luo, X., Wahl, M.C., Stitt, B.L., Gottesman, M.E., and Rösch, P. (2010). A NusE:NusG complex links transcription and translation. Science *328*, 501–504.

Burns, C.M., Richardson, L.V., and Richardson, J.P. (1998). Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. J. Mol. Biol. *278*, 307–316.

Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. Proc. Natl. Acad. Sci. U.S.A.

Cellai, S., Mangiarotti, L., Vannini, N., Naryshkin, N., Kortkhonjia, E., Ebright, R.H., and Rivetti, C. (2007). Upstream promoter sequences and alphaCTD mediate stable DNA wrapping within the RNA polymerase-promoter open complex. EMBO Rep. *8*, 271–278.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., et al. (2012). Opening and closing of the bacterial RNA polymerase clamp. Science *337*, 591–595.

Chan, C.L., and Landick, R. (1993). Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. J. Mol. Biol. *233*, 25–42.

Chan, C.L., Wang, D., and Landick, R. (1997). Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3' end spacing distinguishes pause and termination pathways. J. Mol. Biol. *268*, 54–68.

Chauvier, A., and Lafontaine, D.A. (2015). Probing of Nascent Riboswitch Transcripts. Methods Mol. Biol. *1334*, 109–114.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.-C., Bastet, L., Naghdi, M.R., Dubé, A., Turcotte, P., Perreault, J., and Lafontaine, D.A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. Nat Commun *8*, 13892.

Cheah, M.T., Wachter, A., Sudarsan, N., and Breaker, R.R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature *447*, 497–500.

Cheng, B., and Price, D.H. (2009). Isolation and functional analysis of RNA polymerase II elongation complexes. Methods *48*, 346–352.

Ciampi, M.S. (2006). Rho-dependent terminators and transcription termination. Microbiology (Reading, Engl.) *152*, 2515–2528.

Ciszak, E.M., Korotchkina, L.G., Dominiak, P.M., Sidhu, S., and Patel, M.S. (2003). Structural Basis for Flip-Flop Action of Thiamin Pyrophosphate-dependent Enzymes Revealed by Human Pyruvate Dehydrogenase. J. Biol. Chem. *278*, 21240–21246.

Coppins, R.L., Hall, K.B., and Groisman, E.A. (2007). The intricate world of riboswitches. Curr. Opin. Microbiol. *10*, 176–181.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Dambach, M., Sandoval, M., Updegrove, T.B., Anantharaman, V., Aravind, L., Waters, L.S., and Storz, G. (2015). The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. Mol. Cell *57*, 1099–1109.

Deana, A., and Belasco, J.G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. Genes Dev. *19*, 2526–2533.

Doherty, E.A., and Doudna, J.A. (2001). Ribozyme structures and mechanisms. Annu Rev Biophys Biomol Struct *30*, 457–475.

Dorywalska, M., Blanchard, S.C., Gonzalez, R.L., Kim, H.D., Chu, S., and Puglisi, J.D. (2005). Site-specific labeling of the ribosome for single-molecule spectroscopy. Nucleic Acids Res. *33*, 182–189.

Dutta, D., Chalissery, J., and Sen, R. (2008). Transcription termination factor rho prefers catalytically active elongation complexes for releasing RNA. J. Biol. Chem. *283*, 20243–20251.

Ebright, R.H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J. Mol. Biol. *304*, 687–698.

Engel, C., Sainsbury, S., Cheung, A.C., Kostrewa, D., and Cramer, P. (2013). RNA polymerase I structure and transcription regulation. Nature *502*, 650–655.

Epshtein, V., Mironov, A.S., and Nudler, E. (2003a). The riboswitch-mediated control of sulfur metabolism in bacteria. Proc Natl Acad Sci U S A *100*, 5052–5056.

Epshtein, V., Toulmé, F., Rahmouni, A.R., Borukhov, S., and Nudler, E. (2003b). Transcription through the roadblocks: the role of RNA polymerase cooperation. EMBO J. *22*, 4719–4727.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An allosteric path to transcription termination. Mol. Cell *28*, 991–1001.

Epshtein, V., Dutta, D., Wade, J., and Nudler, E. (2010). An allosteric mechanism of Rho-dependent transcription termination. Nature *463*, 245–249.

Friedman, D.I., and Baron, L.S. (1974). Genetic characterization of a bacterial locus involved in the activity of the N function of phage lambda. Virology *58*, 141–148.

Furukawa, K., Ramesh, A., Zhou, Z., Weinberg, Z., Vallery, T., Winkler, W.C., and Breaker, R.R. (2015). Bacterial riboswitches cooperatively bind Ni(2+) or Co(2+) ions and control expression of heavy metal transporters. Mol. Cell *57*, 1088–1098.

Galluppi, G.R., and Richardson, J.P. (1980). ATP-induced changes in the binding of RNA synthesis termination protein Rho to RNA. J. Mol. Biol. *138*, 513–539.

Gilbert, W. (1986). Origin of life: The RNA world. Nature *319*, 618–618.

Gorski SA, Vogel J, Doudna JA (2017) RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol* **18**: 215–228.

Groisman, E.A., Cromie, M.J., Shi, Y., and Latifi, T. (2006). A Mg2+-responding RNA that controls the expression of a Mg2+ transporter. Cold Spring Harb. Symp. Quant. Biol. *71*, 251–258.

Grundy, F.J., and Henkin, T.M. (2003). The T box and S box transcription termination control systems. Front. Biosci. *8*, d20-31.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849–857.

Haller, A., Altman, R.B., Soulière, M.F., Blanchard, S.C., and Micura, R. (2013). Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. Proc. Natl. Acad. Sci. U.S.A. *110*, 4188–4193.

Harden, T.T., Wells, C.D., Friedman, L.J., Landick, R., Hochschild, A., Kondev, J., and Gelles, J. (2016). Bacterial RNA polymerase can retain σ70 throughout transcription. Proc. Natl. Acad. Sci. U.S.A.

Herbert, K.M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., and Block, S.M. (2010). E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. J. Mol. Biol. *399*, 17–30.

von Hippel, P.H. (1998). An integrated model of the transcription complex in elongation, termination, and editing. Science *281*, 660–665.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. Nature *451*, 851–854.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E.A. (2012). Riboswitch control of Rho-dependent transcription termination. Proc. Natl. Acad. Sci. U.S.A. *109*, 5376–5381.

Hollands, K., Sevostiyanova, A., and Groisman, E.A. (2014). Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator. Proc. Natl. Acad. Sci. U.S.A.

Ingham, C.J., Dennis, J., and Furneaux, P.A. (1999). Autogenous regulation of transcription termination factor Rho and the requirement for Nus factors in Bacillus subtilis. Mol. Microbiol. *31*, 651–663.

Johnson, J.E., Reyes, F.E., Polaski, J.T., and Batey, R.T. (2012). B12 cofactors directly stabilize an mRNA regulatory switch. Nature *492*, 133–137.

Jones CP, Ferré-D'Amaré AR (2017) Long-Range Interactions in Riboswitch Control of Gene Expression. *Annu Rev Biophys* **46**: 455–481.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat. Rev. Mol. Cell Biol. *16*, 167–177.

Kainz, M., and Gourse, R.L. (1998). The C-terminal domain of the alpha subunit of Escherichia coli RNA polymerase is required for efficient rho-dependent transcription termination. J. Mol. Biol. *284*, 1379–1390.

Kashlev, M., Nudler, E., Severinov, K., Borukhov, S., Komissarova, N., and Goldfarb, A. (1996). Histidine-tagged RNA polymerase of Escherichia coli and transcription in solid phase. Meth. Enzymol. *274*, 326–334.

Kim, J.N., Roth, A., and Breaker, R.R. (2007). Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. Proc. Natl. Acad. Sci. U.S.A. *104*, 16092–16097.

Kireeva, M.L., and Kashlev, M. (2009). Mechanism of sequence-specific pausing of bacterial RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. *106*, 8900–8905.

Kohler, R., Mooney, R.A., Mills, D.J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. Science *356*, 194–197.

Komissarova, N., and Kashlev, M. (1997a). RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J. Biol. Chem. *272*, 15329–15338.

Komissarova, N., and Kashlev, M. (1997b). Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc. Natl. Acad. Sci. U.S.A. *94*, 1755–1760.

Kriner, M.A., and Groisman, E.A. (2017). RNA secondary structures regulate three steps of Rho-dependent transcription termination within a bacterial mRNA leader. Nucleic Acids Res *45*, 631–642.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147–157.

Kulshina, N., Edwards, T.E., and Ferré-D'Amaré, A.R. (2010). Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. RNA *16*, 186–196.

Landick, R., and Yanofsky, C. (1987). Isolation and structural analysis of the Escherichia coli trp leader paused transcription complex. J. Mol. Biol. *196*, 363–377.
Landick, R., Carey, J., and Yanofsky, C. (1987). Detection of transcription-pausing in vivo in the trp operon leader region. PNAS *84*, 1507–1511.

Laptenko, O., and Borukhov, S. (2003). Biochemical assays of Gre factors of Thermus thermophilus. Meth. Enzymol. *371*, 219–232.

Larson, M.H., Mooney, R.A., Peters, J.M., Windgassen, T., Nayak, D., Gross, C.A., Block, S.M., Greenleaf, W.J., Landick, R., and Weissman, J.S. (2014). A Pause Sequence Enriched at Translation Start Sites Drives Transcription Dynamics in Vivo.

Science.

Lau, L.F., Roberts, J.W., and Wu, R. (1983). RNA polymerase pausing and transcript release at the lambda tR1 terminator in vitro. J. Biol. Chem. *258*, 9391–9397.

Lauhon CT, Szostak JW (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J Am Chem Soc* **117**: 1246–1257.

Lee, D.N., Phung, L., Stewart, J., and Landick, R. (1990). Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. J. Biol. Chem. *265*, 15145–15153.

Lemay, J.-F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Massé, E., and Lafontaine, D.A. (2011). Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. PLoS Genet. *7*, e1001278.

Levin, J.R., Krummel, B., and Chamberlin, M.J. (1987). Isolation and properties of transcribing ternary complexes of Escherichia coli RNA polymerase positioned at a single template base. J. Mol. Biol. *196*, 85–100.

Li, J., Mason, S.W., and Greenblatt, J. (1993). Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. Genes Dev. *7*, 161–172.

Lussier, A., Bastet, L., Chauvier, A., and Lafontaine, D.A. (2015). A Kissing-Loop is Important for BtuB Riboswitch Ligand Sensing and Regulatory Control. J. Biol. Chem.

Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. *11*, 29–35.

Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell *113*, 577–586.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and

Breaker, R.R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science *306*, 275–279.

Marr, M.T., and Roberts, J.W. (2000). Function of transcription cleavage factors GreA and GreB at a regulatory pause site. Mol. Cell *6*, 1275–1285.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat. Rev. Mol. Cell Biol. *8*, 209–220.

Mattick, J.S. (2004). The hidden genetic program of complex organisms. Sci. Am. *291*, 60–67.

Miranda-Ríos, J., Navarro, M., and Soberón, M. (2001). A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc. Natl. Acad. Sci. U.S.A. *98*, 9736–9741.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell *111*, 747–756.

Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? Mol. Cell *20*, 335–345.

Mooney, R.A., Davis, S.E., Peters, J.M., Rowland, J.L., Ansari, A.Z., and Landick, R. (2009). Regulator trafficking on bacterial transcription units in vivo. Mol. Cell *33*, 97–108.

Morgan, W.D., Bear, D.G., and von Hippel, P.H. (1983). Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. J. Biol. Chem. *258*, 9565–9574.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002). Genetic control by a metabolite binding mRNA. Chem. Biol. *9*, 1043.

Nechooshtan, G., Elgrably-Weiss, M., and Altuvia, S. (2014). Changes in transcriptional pausing modify the folding dynamics of the pH-responsive RNA element. Nucleic Acids

Res. *42*, 622–630.

Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat. Chem. Biol. *9*, 834–839.

Nelson, J.W., Atilho, R.M., Sherlock, M.E., Stockbridge, R.B., and Breaker, R.R. (2017). Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. Mol. Cell *65*, 220–230.

Nudler, E., and Mironov, A.S. (2004). The riboswitch control of bacterial metabolism. Trends Biochem. Sci. *29*, 11–17.

Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. Science *273*, 211–217.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33–41.

Nudler, E., Gusarov, I., and Bar-Nahum, G. (2003). Methods of walking with the RNA polymerase. Meth. Enzymol. *371*, 160–169.

Pace, N.R., and Brown, J.W. (1995). Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. J. Bacteriol. *177*, 1919–1928.

Pan, T., and Sosnick, T. (2006). RNA folding during transcription. Annu Rev Biophys Biomol Struct *35*, 161–175.

Pan, T., Artsimovitch, I., Fang, X.W., Landick, R., and Sosnick, T.R. (1999). Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. Proc. Natl. Acad. Sci. U.S.A. *96*, 9545–9550.

Pavco, P.A., and Steege, D.A. (1990). Elongation by Escherichia coli RNA polymerase is blocked in vitro by a site-specific DNA binding protein. J. Biol. Chem. *265*, 9960–

9969.

Perdrizet, G.A., 2nd, Artsimovitch, I., Furman, R., Sosnick, T.R., and Pan, T. (2012). Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. Proc. Natl. Acad. Sci. U.S.A. *109*, 3323–3328.

Perdue, S.A., and Roberts, J.W. (2010). A backtrack-inducing sequence is an essential component of Escherichia coli σ(70)-dependent promoter-proximal pausing. Mol. Microbiol. *78*, 636–650.

Perdue, S.A., and Roberts, J.W. (2011). Σ(70)-dependent transcription pausing in Escherichia coli. J. Mol. Biol. *412*, 782–792.

Peters, J.M., Mooney, R.A., Grass, J.A., Jessen, E.D., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev. *26*, 2621–2633.

Petersen, L.A., and Downs, D.M. (1997). Identification and characterization of an operon in Salmonella typhimurium involved in thiamine biosynthesis. J. Bacteriol. *179*, 4894–4900.

Price, I.R., Gaballa, A., Ding, F., Helmann, J.D., and Ke, A. (2015). Mn(2+)-Sensing Mechanisms of yybP-ykoY Orphan Riboswitches. Mol. Cell *57*, 1110–1123.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science *328*, 504–508.

Ray-Soni, A., Bellecourt, M.J., and Landick, R. (2016). Mechanisms of Bacterial Transcription Termination: All Good Things Must End. Annu. Rev. Biochem.

Regulski, E.E., Moy, R.H., Weinberg, Z., Barrick, J.E., Yao, Z., Ruzzo, W.L., and Breaker, R.R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. Mol. Microbiol. *68*, 918–932.

Reynolds, R., Bermúdez-Cruz, R.M., and Chamberlin, M.J. (1992). Parameters affecting transcription termination by Escherichia coli RNA polymerase. I. Analysis of 13 rho-independent terminators. J. Mol. Biol. *224*, 31–51.

Richardson, J.P. (1991). Preventing the synthesis of unused transcripts by Rho factor. Cell *64*, 1047–1049.

Richardson, J.P. (2003). Loading Rho to terminate transcription. Cell *114*, 157–159.

Rinaldi, A.J., Lund, P.E., Blanco, M.R., and Walter, N.G. (2016). The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. Nat Commun *7*, 8976.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2002). Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms. J. Biol. Chem. *277*, 48949–48959.

Rosenthal, A.Z., Kim, Y., and Gralla, J.D. (2008). Poising of Escherichia coli RNA polymerase and its release from the sigma 38 C-terminal tail for osmY transcription. J. Mol. Biol. *376*, 938–949.

Roth, A., Winkler, W.C., Regulski, E.E., Lee, B.W.K., Lim, J., Jona, I., Barrick, J.E., Ritwik, A., Kim, J.N., Welz, R., et al. (2007). A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat. Struct. Mol. Biol. *14*, 308–317.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nat. Methods *5*, 507–516.

Saecker, R.M., Record, M.T., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J. Mol. Biol. *412*, 754–771.

Santangelo, T.J., and Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. Nat. Rev. Microbiol. *9*, 319–329.

Santangelo, T.J., and Roberts, J.W. (2004). Forward translocation is the natural pathway of RNA release at an intrinsic terminator. Mol. Cell *14*, 117–126.

Sasse-Dwight, S., and Gralla, J.D. (1989). KMnO4 as a probe for lac promoter DNA melting and mechanism in vivo. J. Biol. Chem. *264*, 8074–8081.

Sasse-Dwight, S., and Gralla, J.D. (1991). Footprinting protein-DNA complexes in vivo. Meth. Enzymol. *208*, 146–168.

Sedlyarova, N., Shamovsky, I., Bharati, B.K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., and Nudler, E. (2016). sRNA-Mediated Control of Transcription Termination in E. coli. Cell *167*, 111–121.e13.

Serganov, A. (2009). The long and the short of riboswitches. Curr. Opin. Struct. Biol. *19*, 251–259.

Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., and Patel, D.J. (2006). Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature *441*, 1167–1171.

Sevostyanova, A., Svetlov, V., Vassylyev, D.G., and Artsimovitch, I. (2008). The elongation factor RfaH and the initiation factor sigma bind to the same site on the transcription elongation complex. Proc. Natl. Acad. Sci. U.S.A. *105*, 865–870.

Sevostyanova, A., Belogurov, G.A., Mooney, R.A., Landick, R., and Artsimovitch, I. (2011). The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. Mol. Cell *43*, 253–262.

Shaevitz, J.W., Abbondanzieri, E.A., Landick, R., and Block, S.M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. Nature *426*, 684–687.

Shankar, S., Hatoum, A., and Roberts, J.W. (2007). A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. Mol. Cell *27*, 914–927.

Sreenivasan, R., Heitkamp, S., Chhabra, M., Saecker, R.M., Lingeman, E., Poulos, M.,

McCaslin, D.R., Capp, M.W., Artsimovitch, I., and Record, M.T. (2016). FRET Characterization of DNA Wrapping in Closed and Open Escherichia coli RNA Polymerase - λPR Promoter Complexes. Biochemistry.

Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. (2003a). An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. *17*, 2688–2697.

Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003b). Metabolite-binding RNA domains are present in the genes of eukaryotes. RNA *9*, 644–647.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science *321*, 411–413.

Takemoto, N., Tanaka, Y., and Inui, M. (2014). Rho and RNase play a central role in FMN riboswitch regulation in Corynebacterium glutamicum. Nucleic Acids Res.

Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. Science *312*, 1208–1211.

Toulokhonov, I., and Landick, R. (2003). The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. Mol. Cell *12*, 1125–1136.

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science *292*, 730–733.

Travers, A.A., and Burgessrr, null (1969). Cyclic re-use of the RNA polymerase sigma factor. Nature *222*, 537–540.

Vander Horn, P.B., Backstrom, A.D., Stewart, V., and Begley, T.P. (1993). Structural genes for thiamine biosynthetic enzymes (thiCEFGH) in Escherichia coli K-12. J. Bacteriol. *175*, 982–992.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and

Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature *417*, 712–719.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. Nature *448*, 157–162.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163–168.

Vogel, J., Bartels, V., Tang, T.H., Churakov, G., Slagter-Jäger, J.G., Hüttenhofer, A., and Wagner, E.G.H. (2003). RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. *31*, 6435–6443.

Vvedenskaya, I.O., Vahedian-Movahed, H., Bird, J.G., Knoblauch, J.G., Goldman, S.R., Zhang, Y., Ebright, R.H., and Nickels, B.E. (2014). Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. Science *344*, 1285–1289.

Vvedenskaya, I.O., Zhang, Y., Goldman, S.R., Valenti, A., Visone, V., Taylor, D.M., Ebright, R.H., and Nickels, B.E. (2015). Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. Mol. Cell *60*, 953–965.

Wachter, A., Tunc-Ozdemir, M., Grove, B.C., Green, P.J., Shintani, D.K., and Breaker, R.R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. Plant Cell *19*, 3437–3450.

Walter, G., Zillig, W., Palm, P., and Fuchs, E. (1967). Initiation of DNA-dependent RNA synthesis and the effect of heparin on RNA polymerase. Eur. J. Biochem. *3*, 194–201.

Walter, W., Kireeva, M.L., Studitsky, V.M., and Kashlev, M. (2003). Bacterial polymerase and yeast polymerase II use similar mechanisms for transcription through nucleosomes. J. Biol. Chem. *278*, 36148–36156.

Wang, J.X., Lee, E.R., Morales, D.R., Lim, J., and Breaker, R.R. (2008). Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Mol. Cell *29*, 691–702.

Watters, K.E., Strobel, E.J., Yu, A.M., Lis, J.T., and Lucks, J.B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. Nat. Struct. Mol. Biol. *23*, 1124–1131.

Weixlbaumer, A., Leon, K., Landick, R., and Darst, S.A. (2013). Structural basis of transcriptional pausing in bacteria. Cell *152*, 431–441.

Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol. Cell *18*, 49–60.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.

Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat. Struct. Biol. *10*, 701–707.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature *428*, 281–286.
Woese, C.R. (1967). The genetic code: the molecular basis for genetic expression (Harper & Row).

Wong, T., Sosnick, T.R., and Pan, T. (2005). Mechanistic insights on the folding of a large ribozyme during transcription. Biochemistry *44*, 7535–7542.

Wong, T.N., Sosnick, T.R., and Pan, T. (2007). Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. Proc Natl Acad Sci U S A *104*, 17995–18000.

Yakhnin, A.V., and Babitzke, P. (2002). NusA-stimulated RNA polymerase pausing and termination participates in the Bacillus subtilis trp operon attenuation mechanism invitro. Proc. Natl. Acad. Sci. U.S.A. *99*, 11067–11072.

Yakhnin, A.V., and Babitzke, P. (2010). Mechanism of NusG-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the Bacillus subtilis trp leader. Mol. Microbiol. *76*, 690–705.

Yakhnin, A.V., Yakhnin, H., and Babitzke, P. (2008). Function of the Bacillus subtilis transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the trp leader. Proc. Natl. Acad. Sci. U.S.A. *105*, 16131–16136.

Yu, J., and Schneiders, T. (2012). Tigecycline challenge triggers sRNA production in Salmonella enterica serovar Typhimurium. BMC Microbiol. *12*, 195.

Yuzenkova, Y., and Zenkin, N. (2010). Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. Proc. Natl. Acad. Sci. U.S.A. *107*, 10878–10883.

Zamft, B., Bintu, L., Ishibashi, T., and Bustamante, C. (2012). Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. Proc. Natl. Acad. Sci. U.S.A. *109*, 8948–8953.

Zhang, J., and Ferré-D'Amaré, A.R. (2015). Structure and mechanism of the T-box riboswitches. Wiley Interdiscip Rev RNA.

Zhang, J., and Landick, R. (2016). A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. Trends Biochem. Sci.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell *98*, 811–824.