Interactions between the Translation Machinery and a Translational preQ$_1$ Riboswitch

by

Paul E. Lund

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2015

Doctoral Committee:

Professor Nils G. Walter, Chair
Professor Carol A. Fierke
Professor Janine R. Maddock
Professor Patrick J. O'Brien

Acknowledgements

I am fortunate to have many great people in my life, without whose influence, actions, and support the work presented in this dissertation could not have been realized. First and foremost I must thank my advisor Nils, who permitted me the intellectual freedom and provided the lab environment in which I was able to pursue the subjects of my dissertation research, as well as the other members of my committee. I am lucky to have committee memberes that have always been engaged in the discussions of my research and whose guidance and feedback I have always appreciated. Members of the Maddock lab, and Lolita in particular, helped me get my figurative foot in the door when it came to working comfortably and competently with ribosomes, and I am grateful to the Cierpicki and Gestwicki lab members, and especially the always-helpful members of the Fierke lab, all of whose assistance was invaluable when it came to expressing and purifying protein (of which I have done I surprising amount for being in an RNA lab).

All of my mentors past and present helped me get to the point at which I find myself today, and I am very lucky to have had wonderful mentors at each stage of my academic carrier. Also, I want to express my profound appreciation for the members of the Walter lab for creating a great work environment, and in particular Gabrielle, Arlie, Mario and May with whom I've worked most closely, as well as Kamali for her tireless cheerleading.

Lastly, I wish to thank my friends both here and outside of Michigan, especially my longtime friend Shel who, being three at least three degrees separated from everything happening at Michigan during graduate school, was my often-neeed independent consultant, sounding

board, and reality check. Perhaps most importantly, I gratefully acknowledge my friends and classmates in the Program in Chemical Biology, without whom I literally could not have accomplished what I have with respect to my research here at Michigan, and who have made these last 5 years truly enjoyable.

Table of Contents

List of Figures

Appendix Figures

List of Tables

Appendix Tables

List of Appendices

List of Abbreviations

30S IC	30S initiation complex
70S IC	70S initiation complex
Bas mRNA	*B. anthracis* mRNA
BSA	bovine serum albumin
CAT	chloramphenicol acetyltransferase
CV	column volume
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
FCS	fluorescence correlation spectroscopy
FRET	fluorescence resonance energy transfer
HPLC	high performance liquid chromatography
HMM	Hidden Markov Model
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ISI	inter-spike interval
LNA	locked nucleic acid
mRNA	messenger RNA
miRNA	micrRNA
nt	nucleotide
NTP	nucleotide triphosphate
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
$preQ_1$	7-aminomethyl-7-deazaguanine
RNA	ribonucleic acid
rRNA	ribosomal RNA
RS	Rank Surprise
S1	ribosomal protein S1
SD	Shine-Dalgarno
SDS	sodium dodecylsulfate
SiM-KARTS	Single Molecule Kinetic Analysis of RNA Transient Structure
SKM	segmental k-means
smFRET	single molecule fluorescence resonance energy transfer
sRNA	small RNA
ss	single-stranded
TBE	Tris-borate EDTA buffer
TCEP-HCl	tris(2-carboxyethyl)phosphine hydrochloride
TE	10 mM Tris-HCl, 1 mM EDTA buffer
TEV	tobacco etch virus

TIR	translation initiation region
TIRFM	total internal reflection fluorescence microscopy
TLC	thin-layer chromatography
tmRNA	transfer-messenger RNA
tRNA	transfer RNA
Tte mRNA	*T. tengcongensis* mRNA
UTR	untranslated region
v/v%	volume by volume percent
w/v%	weight by volume percent

Abstract

There are many ways in which gene expression is regulated in biology, including the diversity of regulation that occurs at the level of RNA. This is due in no small part to the sheer variety of ways in which this regulation manifests, whether in *trans* by the action of miRNA and siRNA in higher organisms or sRNA in the case or prokaryotes, or in *cis,* through sequence and structure in the 5′ and 3′ UTRs, among many others. Regulation of the translation of a messenger RNA into protein in bacteria occurs both through sequence features, such as the Shine-Dalgarno (SD) sequence, and local structural features of the RNA. Translational riboswitches in bacteria provide an excellent example of just such a *cis*-acting regulatory mechanism. The work presented in this dissertation aims to answer fundamental questions about how the structural features that make up the preQ$_1$ riboswitch perform their regulatory function through interactions with the translation machinery. Broader questions about how individual components of the translation machinery, such as ribosomal protein S1 and the 30S ribosomal subunit, interact with structured RNAs are also addressed.

The basis of the putative mechanism for most translational riboswitches can be described generally as a change in the structure of the translation initiation region of the host mRNA, brought about by the binding of the respective ligand(s). However, we sought a more detailed mechanistic view of the interplay between the translational preQ$_1$ riboswitch found in the 5′ UTR of an mRNA from *T. tengcongensis,* its ligand preQ$_1$, and the accessibility of the SD sequence. To this end, we developed SiM-KARTS, a generalized strategy in which the hybridization kinetics of a short nucleic acid probe are used to interrogate the site-specific

structural dynamics of individual RNA molecules of arbitrary size. Intriguingly, we found that the expression platform of the riboswitch alternates between two conformations with different degrees of SD accessibility, which are distinguished by "bursts" of probe binding, and that the main effect of ligand is to alter the pattern of these bursts. This challenges the initial assumption that riboswitches behave in a digital manner, where response to ligand should be all or nothing as opposed to the more nuanced and complex type of attenuation that we observed, and thus has broader implications for how we think about translational riboswitch regulation.

RNA structure plays a role in influencing many other cellular processes besides translation, and the folding or unfolding of that structure is integral to the function of the RNA. Ribosomal protein S1, despite its name, has a well-established role not only in translation initiation, but also in transcription and phage replication. The part that S1 plays in these processes is related to its ability to bind and or unfold structured RNA. In order to better understand how this large protein interacts with folded RNA, we used the well-characterized preQ$_1$ riboswitch as a model pseudoknot to study how S1 interacts with defined, stable tertiary structure. We observed that S1 is able to bind and at least partially unfold this pseudoknot, and defined an apparent upper limit to the degree of structural stability in an RNA fold that S1 is able to act upon.

Lastly, we looked at the influence of S1 on translation of two preQ$_1$ riboswitch-containing mRNAs. The effects of ligand on mRNA translation are not greatly potentiated by the loss of S1; however, a dramatic effect was observed on the translational coupling of co-transcribed genes, which invokes a role of S1 in the translation of polycistronic mRNA. These results further underscore the need for better knowledge of how the 30S subunit binds to structured mRNA, and to mRNA in general. To this end, we describe initial steps to develop a single molecule assay for monitoring early 30S-mRNA interactions during the initiation phase of translation.

CHAPTER 1: An introduction to the regulatory functions of structured RNA and initiation of translation in bacteria

The translation of messenger RNA (mRNA) into protein is one of the major control points in the expression of any gene and as such, translation is highly regulated by the cell through a number of different mechanisms, including by the inherent structure of the mRNA itself. A full understanding of initiation, the first and rate-limiting phase of translation[1,2], is therefore crucial for understanding how translation is regulated and, more generally, how gene expression levels are determined[3,4].

1.1 Bacterial translation in brief

Translation of the mRNA-encoded amino acid sequence is mediated by the ribosome and can be divided into three phases: Initiation, Elongation, and Termination/Recycling[5]. Initiation is the first phase of translation, during which the ribosome assembles on the mRNA at the start codon (**Figure 1-1**). In the first step of initiation, the 30S subunit complexed with IF1 and IF3 is bound by IF2·GTP. This is followed by binding of the mRNA and fMet-tRNA^{fMet}, the order of which is thought to be unimportant and stochastic[2,6]. A number of conformational rearrangements occur that position fMet-tRNA^{fMet} at the start codon in the peptidyl-tRNA site (P-site) generating the 30S initiation complex (30S IC), to which the 50S subunit then binds. 50S subunit joining is followed by rapid hydrolysis of IF2-bound GTP, correlated with additional rearrangements of the

subunits[7] and dissociation of factors. This yields a functional 70S initiation complex (70S IC) that then proceeds into the elongation phase of translation.

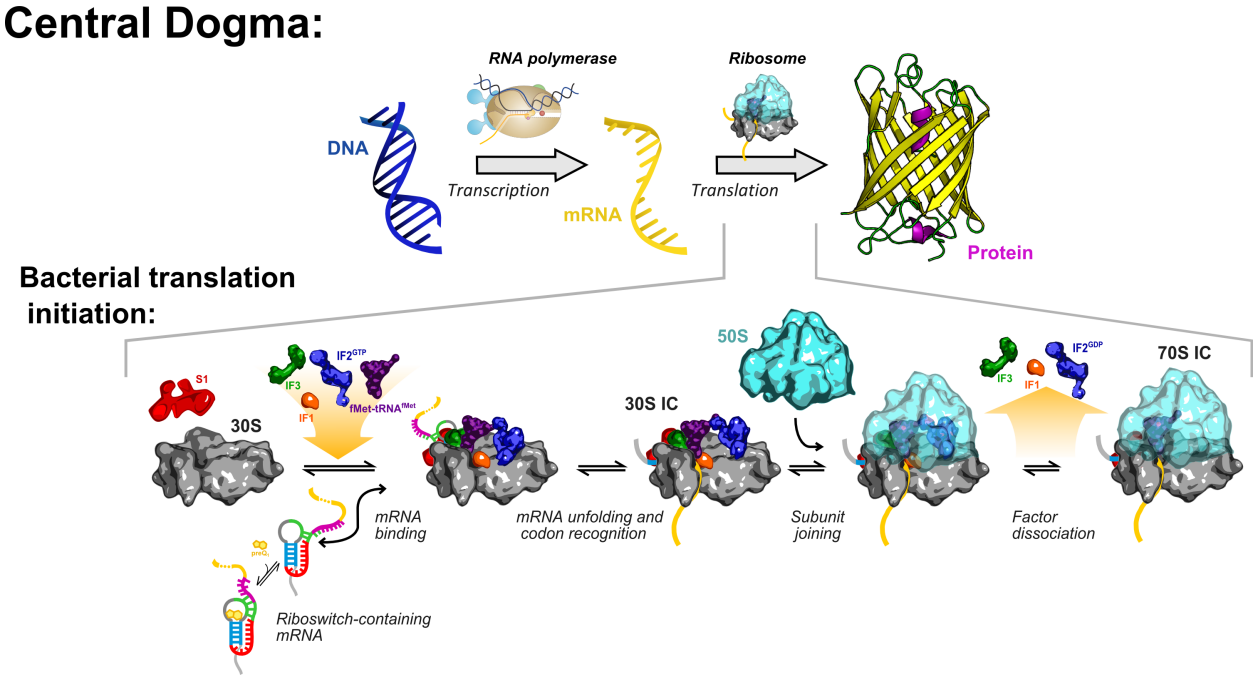


Figure 1-1 Overview of translation initiation in bacteria

In the Central Dogma of molecular biology, genetic information is encoded in an organism's DNA. The flow of this information is from DNA to messenger RNA (mRNA) through transcription of the DNA by RNA polymerase, followed by translation of the mRNA into protein by the ribosome. The initiation phase of translation in bacteria is a complicated, multistep process in which the mRNA, 30S small ribosomal subunit, initiator tRNA and initiation factors come together to form the 30S initiation complex (30S IC). This is followed by joining of the 50S large ribosomal subunit, hydrolysis of GTP by IF2, and dissociation of initiation factors to give the 70S initiation complex (70S IC). At this point, the initiation phase is complete and the ribosome is ready to being new protein synthesis. Elements adapted from Ref. 8.

Despite numerous years of study and the relatively small number of components (i.e. ribosome, initiation factors, and tRNA), much is yet unknown about the intermolecular interactions that govern initiation of translation in bacteria. Of these components, the translation initiation region (TIR), a term used to describe the entire 5′ portion of the mRNA to which the ribosome binds, arguably possesses the greatest variability and thus the mRNA itself has the greatest potential for modulating expression.

1.1.1 Role of the Shine-Dalgarno sequence

One well-known feature commonly found in the TIR of many (but not all[9-12]) mRNA is the Shine-Dalgarno (SD) sequence – a purine rich stretch of 3-8 nucleotides located a short distance upstream of the start codon[13,14]. The SD sequence is complementary to the sequence of nucleotides (i.e., the anti-SD) at the very 3′ end of the 16S ribosomal RNA (rRNA). Formation of the SD•anti-SD duplex provides translational specificity for the ribosome[15] as it aids in selection of the appropriate start site[14].

Variation in the TIR is known to play an important role in regulating expression. For example, the degree of SD/anti-SD complementarity and position relative to the start codon are known to dramatically influence ribosome affinity for the mRNA translational efficiency[9,16-18], with a higher dependence on the degree of complementarity required for efficient translation in Gram-positive bacteria, such as *Bacillus subtilis,* than for Gram-negative bacteria like *E. coli*[19].

1.1.2 RNA structure in the translation initiation region

Secondary structure in the TIR is also known to be a key determinant of expression level through its various effects on initiation[3,4]. With regard to structure in the TIR of bacterial mRNAs, there are numerous examples of regulatory motifs known as riboswitches (discussed below) whose structures change in response to stimuli. The ribosome itself has been found to possess a degree of intrinsic helicase activity[20], which presumably allows it to negotiate structure in the TIR during initiation, as well as structure during translation in general[21]. This helicase activity is attributable in part to ribosomal proteins S3, S4, and S5 which are situated near the entrance of the downstream tunnel on the 30S, through which the incoming mRNA must pass[22]. In addition to this intrinsic activity of the ribosome, a number of other mechanisms exist for dealing with structured RNA, in particular ribosomal protein S1.

1.2 Biological roles of ribosomal protein S1

Ribosomal protein S1 has a well-established role in the translation of mRNA by the *Escherichia coli* ribosome, in which S1 mediates the binding of many mRNAs by the 30S subunit[23-25], and is particularly important for translation of mRNA with a weak or no SD sequence[26,27]. S1 is a large, acidic protein that weakly associates with the 30S subunit in sub-stoichiometric amounts[28,29], and is required for cell growth and viability[23,30]. Chemical cross-linking data[31] and recent cryo-EM reconstructions[32] showed clearly that S1 makes extensive contacts with the solvent face of the 30S subunit, as well as interacts with ~11 nts of the mRNA immediately upstream of the SD sequence. This positioning is consistent with studies identifying S1 as playing a key role in the initial binding of mRNA to the ribosome through its affinity for A/U-rich sequences that often occur near the start site in mRNA[24,33-35]. Importantly, S1 is known to be involved in the translation of mRNA with highly structured TIR[36-39] and that this role is related to the ability of S1 to unwind RNA secondary and tertiary structure[30,40].

Interestingly, the gene that encodes S1, *rpsA*, itself lacks a traditional SD sequence. Binding of, and thus protein production from, the *rpsA* mRNA by the 30S ribosomal subunit is also mediated by 30S-bound S1. S1 expression is also autoregulated: binding of free S1 to the TIR of the *rpsA* mRNA[10,41-43] prevents binding by 30S-bound S1 and thus S1 production is regulated by a negative-feedback mechanism.

Despite its name, S1 plays other cellular roles[44] beyond its mRNA binding activity on the 30S subunit. For example, S1 interacts with tmRNA independently of the ribosome and is required for the rescue of stalled ribosomes and tagging of proteins for degradation by trans-translation[45]. S1 also has a better known role as part of Qβ replicase[46], in which it aids in initiating replication of the Qβ viral RNA genome[47,48] as well as promoting product release at the

termination step[49]. In a similar vein, S1 has been shown to enhance transcription by RNA polymerase, potentially by promoting transcript release and preventing the nascent mRNA from having inhibitory interactions with the transcription complex[50]. S1 also stimulates the activity of T4 endoribonuclease RegB through its interactions with the substrate RNA[51]. In each of these capacities, the RNA binding activity of S1 is essential for carrying out the respective function.

1.3 Riboswitches: Regulation of RNA, by RNA, realized though structural changes

The impact of structural dynamics on function has become increasingly clear, particularly for RNA molecules[52], and this is exemplified by riboswitch regulation of gene expression in bacteria[53-55].

Riboswitches are a category of conserved, *cis*-acting RNA structural motifs present it untranslated regions of many bacterial messenger RNAs (mRNAs)[56,57]. These structural motifs can either up- or down-regulate the expression of the downstream open reading frame(s) (ORF), either at the transcriptional or translation level. This regulation is effected, either directly or indirectly, as a result of structural changes that are induced by the binding of a ligand. Riboswitches are composed of an aptamer domain, which is the minimal sequence required to bind ligand, and an expression platform, through which the structural changes induced by ligand binding by the aptamer domain are telegraphed into modulation of gene expression[58]. This typically entails formation of a transcription terminator or anti-terminator stem loop, in the case of transcriptional riboswitches, or changes in the accessibility of the SD sequence and start codon of the adjacent ORF, in the case of translational riboswitches (**Figure 1-2**).

Within the last decade, three classes riboswitches have been described[59-62] whose ligand was identified as preQ$_1$, a metabolic precursor of the hyper-modified base Queuosine (Q). These

families of riboswitches were identified in the 5′-UTRs from bacterial genes thought to be involved in the biosynthesis of Q, which is often present in wobble position of the anticodons for Asparagine, Aspartic acid, Histidine, and Tyrosine tRNAs in bacteria[59,62]. Members of the class-I preQ$_1$ riboswitch family have some of the smallest known aptamer domains (as few as ~33 nt)[59] and the binding of ligand is accomplished through the use of various H-type pseudoknot folds[63-66]. While reporter assays have validated the *in vivo* activity of the class-I preQ$_1$ riboswitch in a transcriptional context[59], less is known about the *in vivo* activity of the translational variants. One member of this family in particular, the translational preQ$_1$ riboswitch found in *Thermoanaerobacter tengcongensis* features prominently in the dissertation work described here.

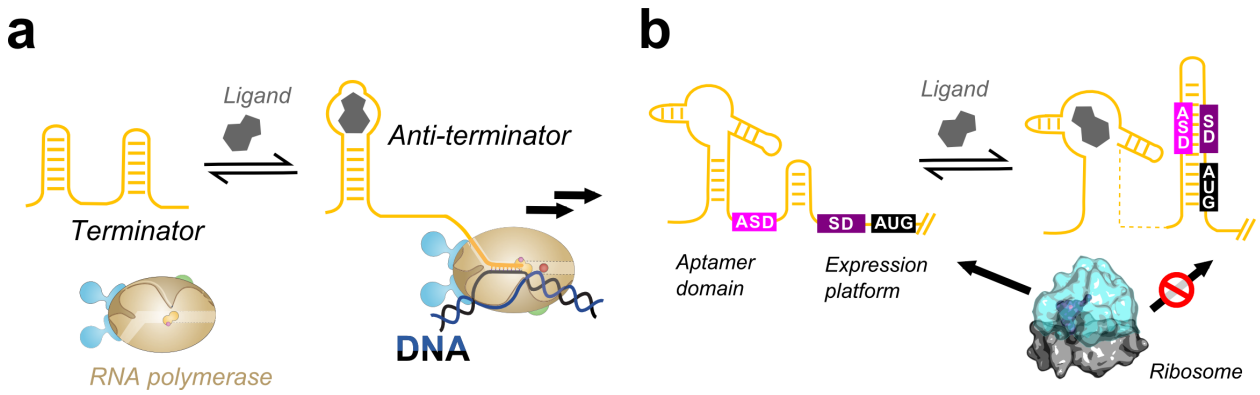


Figure 1-2 Generalized mechanisms for riboswitch-mediated transcriptional and translational attenuation brought about by ligand binding.

Riboswitches regulate expression at the transcriptional (**a**) or translational (**b**) level as a result of changes in conformation brought about by binding of the riboswitch ligand. In the example in **a**, binding of ligand induces formation of an anti-terminator stem, promoting continued transcription of the gene by RNA polymerase, demonstrating how ligand may promote expression in an ON type mechanism. In the example in **b**, ligand binding results in sequestration of the ribosome binding site (SD-AUG), providing an example of an OFF type mechanism. Elements adapted from Ref. 8.

This dissertation contains five chapters. The first chapter provides an introduction to the importance of RNA structure in regulation of expression, with a special focus on translational riboswitches, and an overview of the initiation of translation in bacteria. Chapter 2 is adapted from a manuscript currently submitted for consideration at *Nature Communications* entitled "The SD sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts" for which I am co-first author. In this chapter, the development of the SiM-KARTS technique is described, as well as its application to the study of changes in the accessibility of the SD sequence as a function of ligand concentration. Chapter 3 presents a mechanistic study of the RNA unfolding properties of ribosomal protein S1 from *E. coli*, using a series of model RNA pseudoknots based on the pseudoknot present in the class-I preQ$_1$ riboswitch. Chapter 4 summarizes work done to identify and develop a new preQ$_1$-riboswitch candidate mRNA from *Bacillus anthracis*, as well as the motivation for and development of a single molecule assay for studying the early steps in the initiation stage of translation. The final chapter recapitulates the general conclusions from each chapter with an additional discussion of future directions for my dissertation research.

CHAPTER 2: The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts[1]

2.1 Overview

In response to intracellular signals in Gram-negative bacteria, translational riboswitches – commonly embedded in messenger RNAs (mRNAs) – regulate gene expression through promotion or inhibition of translation initiation. In the case of translation inhibition, it is generally thought that this regulation originates from occlusion of the Shine-Dalgarno (SD) sequence upon ligand binding, but little direct evidence exists. Here, we develop Single Molecule Kinetic Analysis of RNA Transient Structure (SiM-KARTS) to investigate the ligand dependent accessibility of the SD sequence of an mRNA hosting the 7-aminomethyl-7-deazaguanine ($preQ_1$) sensing riboswitch. Spike train analysis reveals that individual mRNA molecules alternate between two conformational states, distinguished by "bursts" of probe binding associated with increased SD sequence accessibility. Addition of $preQ_1$ decreases the lifetime of the SD's high-accessibility (bursting) state, and prolongs the time between bursts. In addition, ligand-jump experiments reveal imperfect riboswitching of single mRNA molecules.

[1] The contents of this chapter are adapted from a manuscript of the same title currently submitted for consideration at *Nature Communications*, by Arlie J. Rinaldi, Paul E. Lund, Mario R. Blanco, and Nils G. Walter. A.J.R. and P.E.L. are co-first authors. P.E.L. designed, conducted, and analyzed the *in vitro* translation experiments and SiM-KARTS experiments done in the presence of blocking strand. A.J.R and N.G.W conceived of the project. A.J.R. designed, conducted, and analyzed the SiM-KARTS experiments. M.R.B. wrote and implemented the scripts for burst analysis. M.RB. and P.E.L. wrote and implemented scripts for Fano factor calculations. All authors were involved in interpreting and discussing the results, and participated in writing the paper.

Such complex ligand sensing by individual mRNA molecules rationalizes the nuanced ligand response observed during bulk mRNA translation.

2.2 Introduction

Riboswitches are non-coding structural elements most commonly embedded in the 5′ untranslated region (UTR) of bacterial messenger RNAs (mRNAs) that regulate the expression of a downstream gene through the binding of an intracellular signal[58,67-70]. These signals include nucleobases[71,72], amino acids[73,74], cofactors of metabolic enzymes[75,76], and metal ions[77-79], among others. Genetic regulation is achieved through a multitude of mechanisms; however, the two most common modes are transcriptional attenuation and inhibition of translation initiation. In translational riboswitches, ligand binding sequesters the Shine-Dalgarno (SD) sequence of the mRNA through alternative base-pairing, resulting in occlusion of the ribosomal binding site needed for efficient initiation of translation. In recent years, a plethora of biophysical techniques has been employed to understand the genetic regulation mechanism utilized by translational riboswitches[57,63,80-84]; however, these techniques have rarely involved the entire mRNA and have largely failed to provide direct mechanistic insight at a molecular level into the coupling of ligand-induced conformational changes with downstream regulatory effects.

The preQ$_1$ riboswitch from *Thermoanaerobacter tengcongensis* (*Tte*) is a strikingly small translational riboswitch that responds to the presence of the modified nucleobase preQ$_1$[85]. Crystal structures[63], in combination with single molecule fluorescence resonance energy transfer (smFRET), NMR and computational studies[83], have indicated that this riboswitch achieves genetic regulation by the formation of a pseudoknot, in which the P2 helix containing the first two nucleotides of the SD sequence is formed. In the absence of ligand, this helix is only transiently closed[83], presumably revealing the SD sequence in its entirety as part of the

downstream expression platform, and thus permitting hybridization with the anti-Shine-Dalgarno (anti-SD) sequence of the 16S ribosomal RNA (rRNA, **Figure 2-1a, b**). These and many similar studies[57,85] have provided valuable insight into the conformational sampling of the preQ$_1$ as well as other translational riboswitches as a function of ligand concentration. They have left largely unresolved, however, the molecular mechanism by which sequestration of just a small fraction of the SD sequence brings about the proposed ON/OFF gene regulatory control of an entire mRNA through coupling between the aptamer and expression platform.

We developed a technique termed Single Molecule Kinetic Analysis of RNA Transient Structure (SiM-KARTS), wherein a short, fluorescently labeled RNA probe, whose sequence is complementary to a particular region of interest, is used to probe changes in structure of a longer target RNA through repeated binding and dissociation events. In the current implementation, an RNA probe corresponding to the 3′ sequence of *T. tengcongensis* (*Tte*) 16S rRNA (i.e., an anti-SD probe) binds to preQ$_1$ riboswitch-containing, surface-immobilized, single *Tte* mRNA molecules, and thus directly reports on the accessibility of the SD sequence. Our results reveal unexpected complexities of ligand-induced riboswitching during translation initiation and add a new dimension to an emerging model wherein stochastic single molecule events contribute to fine-tuned temporal gene expression fluctuations in bacteria. We anticipate that SiM-KARTS will find broad application in probing dynamic RNA structural elements at the single molecule level.

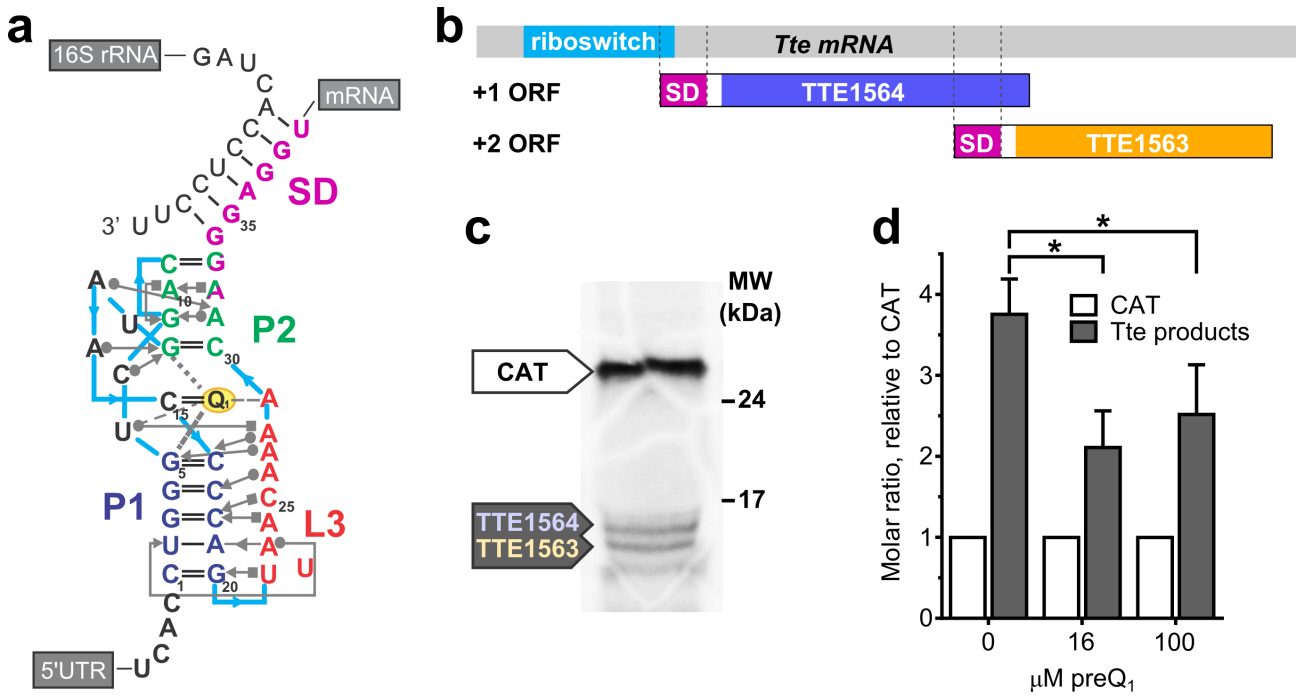


Figure 2-1 *In vitro* translation of *Tte* mRNA.

(**a**) Structural map of the *Tte* preQ$_1$ translational riboswitch displayed with Leontis-Westhof notation[86]. The Shine-Dalgarno sequence (SD, purple) partially overlaps the P2 stem nucleotides (green). Formation of the P2 stem interferes with proper base-pairing between the anti-SD sequence at the 3' end of the 16S rRNA and the SD sequence of the *Tte* mRNA. (**b**) Schematic of *Tte* mRNA used for *in vitro* translation assays. The putative mRNA transcript is bicistronic, containing the overlapping reading frames for TTE1564 and TTE1563. The preQ$_1$ riboswitch aptamer (light blue) overlaps with a portion of the Shine-Dalgarno sequence (SD, purple) of TTE1564. (**c**) Example autoradiograph of *in vitro* translation products. *Tte* mRNA was translated using L-[^{35}S]-Cys in the presence of the control mRNA encoding chloramphenicol acetyltransferase (CAT) at a 4:1 ratio of *Tte*:CAT mRNA. Molecular weight markers are indicated on the right (full-length gel is shown in **Figure A.1-1**). (**d**) Quantification of *in vitro* translation products as a function of preQ$_1$ concentration. The total protein produced from the *Tte* mRNA (sum of TTE1564 and TTE1563 bands, gray bar) is reported relative to the intensity of the CAT product in the same lane, after normalizing for the cysteine content of each protein (5, 1, 1 for CAT, TTE1564, and TTE1563, respectively). The individual contributions from each protein are presented in more detail in **Figure 2-3b** and **c**. The results represent the mean ± standard deviation (s.d.) of three replicates (*P < 0.05).

2.3 Results

2.3.1 preQ$_1$ regulates translation of the Tte *mRNA* in vitro

The expression platform of translational riboswitches contains the SD sequence, a short (3-8 nucleotide, nt), purine-rich sequence located approximately 5-9 nt upstream of the start codon of bacterial mRNAs[13]. It hybridizes with the 3′ end of 16S rRNA (**Figure 2-1b**), and this interaction is important for canonical initiation and proper start codon selection by the 30S ribosomal subunit[1]. To date, a majority of riboswitch studies have focused on the properties of only the isolated aptamer domain[57]. Studies incorporating the full riboswitch including the downstream expression platform have mostly involved reporter gene assays replacing the mRNA's native gene(s) with the exogenous reporter. Here, we instead opted to test the functional activity of the preQ$_1$ riboswitch in the context of its native mRNA. In *T. tengcongensis*, the riboswitch is located in the 5′ UTR of a putative bicistronic operon as the *Tte* mRNA encodes two genes (**Figure 2-1b**). *In vitro* translation using purified *E. coli* translation factors and ribosomes, which share the anti-SD sequence of *T. tengcongensis* 16S rRNA with the exception of an additional 3′ single nucleotide overhang (**Figure 2-2a**), produced the corresponding two proteins, TTE_RS07450 and TTE_RS07445 (subsequently referred to by their former locus tags TTE1564 and TTE1563, respectively), as expected (**Figure 2-1c**, **Figure 2-2b, c**). We next performed competition experiments using a 4:1 molar ratio of *Tte* to CAT control mRNA, where the latter encodes chloramphenicol acetyltransferase (CAT), does not contain the preQ$_1$ riboswitch and thus is not expected to be modulated in its translation by preQ$_1$ (**Figure 2-3a**). We observed an mRNA-specific, ~40% decrease in translation of the two *Tte* mRNA genes upon addition of saturating concentrations (16 and 100 μM, see below) of preQ$_1$ (**Figure 2-1d**; note that the quantification accounts for the higher number of labeled cysteines in CAT, see **Materials and**

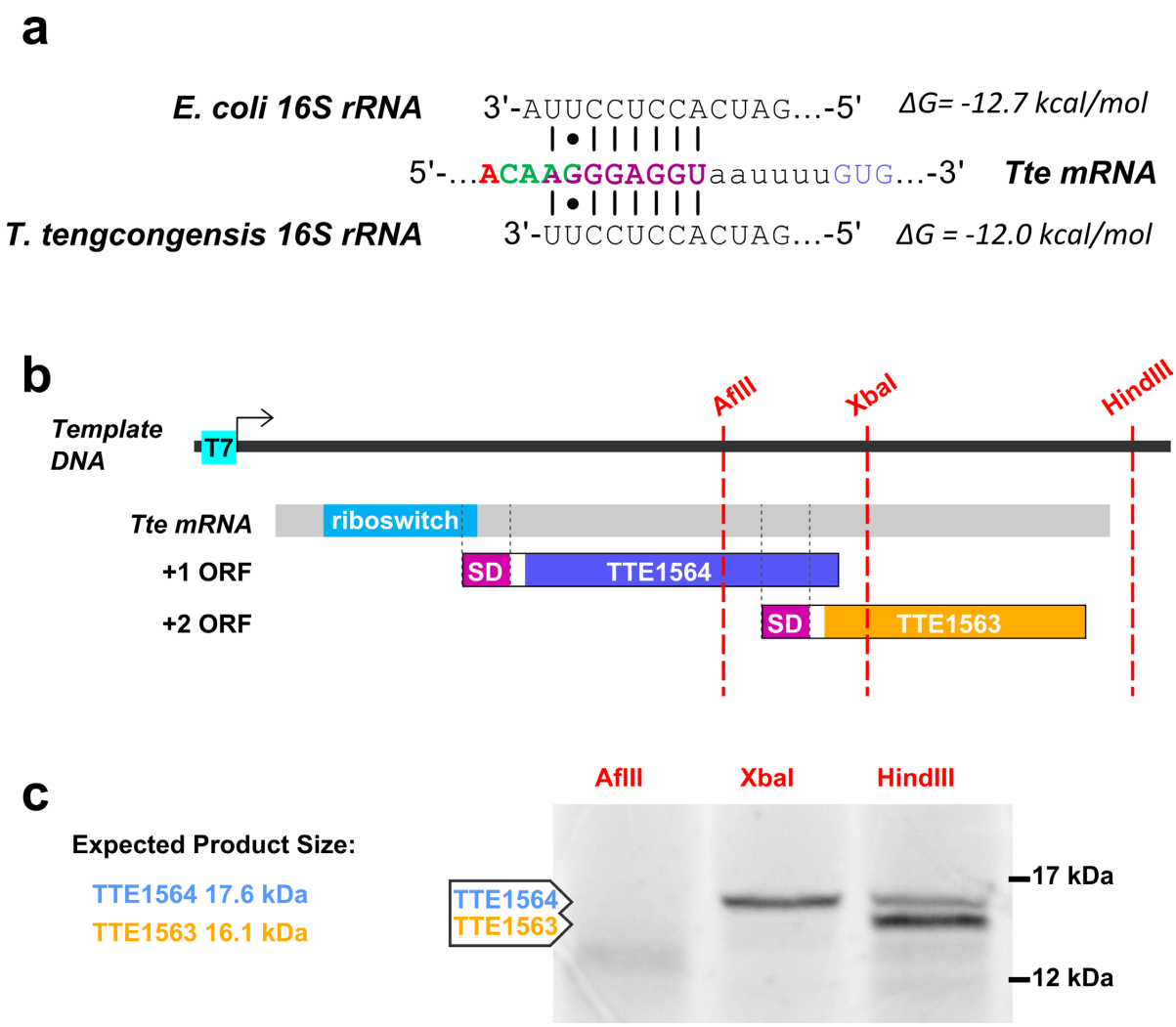


Figure 2-2 Specifics of the *in vitro* translation of *Tte* mRNA.

(**a**) Comparison of Shine-Dalgarno/anti-Shine-Dalgarno pairings between the *Tte* mRNA and the 16S rRNA from *E. coli* ribosomes, used for *in vitro* translation assays, and from *T. tengcongensis*. The Shine-Dalgarno sequence (purple) partially overlaps the P2 stem nucleotides (green). The last A of L3 is shown in red. The GUG start codon of TTE1564 is shown in dark blue. Gibbs free energies of folding were calculated in RNAstructure v5.6 (Mathews Lab). The 12 nucleotides shown for the *T. tengcongensis* 16S rRNA were used as the sequence for the anti-SD probe in SiM-KARTS experiments. (**b**) Schematic of truncated *Tte* mRNA transcripts used to positively assign bands to TTE1564 and TTE1563 products. Truncated transcripts were prepared by digesting the DNA template with specific restriction enzymes (red). Only complete ORFs with a stop codon will generate a protein product, allowing for unambiguous identification of bands. (**c**) Autoradiograph of *in vitro* translation using L-[^{35}S]-Cys of various truncations of the *Tte* mRNA. Molecular weight markers are indicated on the right (full-length gel is shown in **Figure A.1-2**).

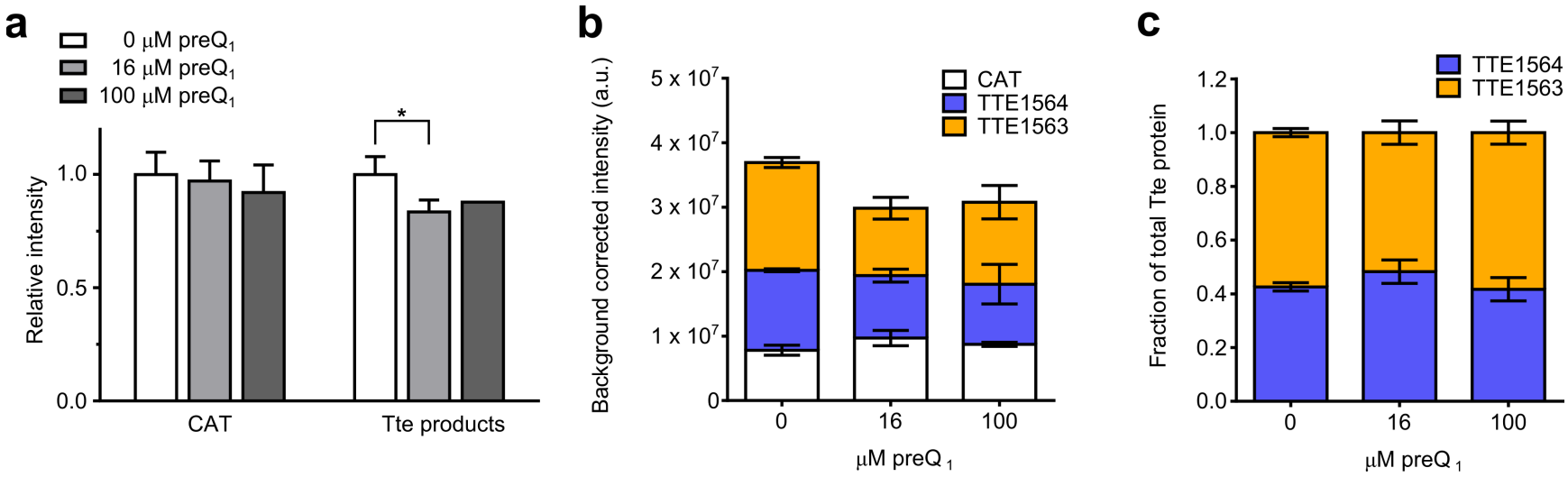


Figure 2-3 Quantification of the *in vitro* translation of CAT and *Tte* mRNAs.

(**a**) *In vitro* translation reactions containing only CAT or *Tte* mRNA in the presence or absence of saturating concentrations of $preQ_1$. Background corrected band intensities were normalized to the total intensity in the lane and reported relative to the mean of the zero $preQ_1$ reactions (full-length gel is shown in **Figure A.1-3**). The results represent the mean ± s.d. of three replicates, except for the 100 μM $preQ_1$ *Tte* reaction (dark gray bar), which represents a single measurement. (*$P < 0.05$). (**b**) Background-corrected band intensities for the same competition *in vitro* translation experiment shown in **Figure 2-1c, d** after normalizing for the cysteine content of each protein (5, 1, 1 for CAT, TTE1564, and TTE1563, respectively). The results represent the mean ± s.d. of a single experiment with three replicates (*$P < 0.05$). See **Figure A.1-1** for full-length gel. (**c**) Relative ratio of TTE1564 and TTE1563 proteins produced as a function of $preQ_1$ concentration from the same experiment shown in **Figure 2-1d**. The results represent the mean ± s.d. of the ratio of TTE1564 and TTE1563 bands from three replicates. No significant changes in this ratio were observed with increasing concentrations of $preQ_1$, consistent with a tight coupling in expression of the two genes in the operon (full-length gel shown in **Figure A.1-1**).

Methods 2.5.7). This result suggests that preQ$_1$ decreases the translational efficiency of *Tte* mRNA, and that the native mRNA is thus responsive to ligand-induced structural changes.

2.3.2 Careful choice of SiM-KARTS probe sequence is important for studying the preQ$_1$ riboswitch in its native context

To observe changes in SD sequence accessibility as a function of ligand concentration, we developed SiM-KARTS (**Figure 2-4**), utilizing a short, fluorescently (Cy5) labeled RNA anti-SD probe with the sequence of the 12 nt at the very 3′ end of *T. tengcongensis* 16S rRNA (**Figure 2-1a**). The SiM-KARTS technique exploits the transient and repeated binding of a short, fluorescently labeled probe oligonucleotide to interrogate the structure of a conformationally dynamic site of interest in an RNA of arbitrary size, and report on structural changes at that site through changes in the probe's binding and dissociation kinetics.

While we could have chosen to use a probe of different length with perfect complementarity, we chose instead to use the anti-SD sequence present at the 3′ end of the *T. tengcongensis* 16S rRNA. Hybridization of the anti-SD sequence with the SD sequences of bacterial mRNAs aids the ribosome in correctly locating the start sites in mRNA and initiating translation. By choosing this sequence, we effectively created a highly simplified *in vitro* mimic of the bacterial ribosome. The use of the 16S rRNA sequence, which is highly, though not exactly, complementary to the riboswitch expression platform, allows us to recapitulate the interaction between the ribosome and the mRNA. This alleviates the concern that if we were to use a different probe sequence with greater length or perfect complementary and possibly slightly elevated site-specificity, we might inadvertently alter the nature of what is likely a carefully balanced interplay between the SD and anti-SD sequences, which evolved together to

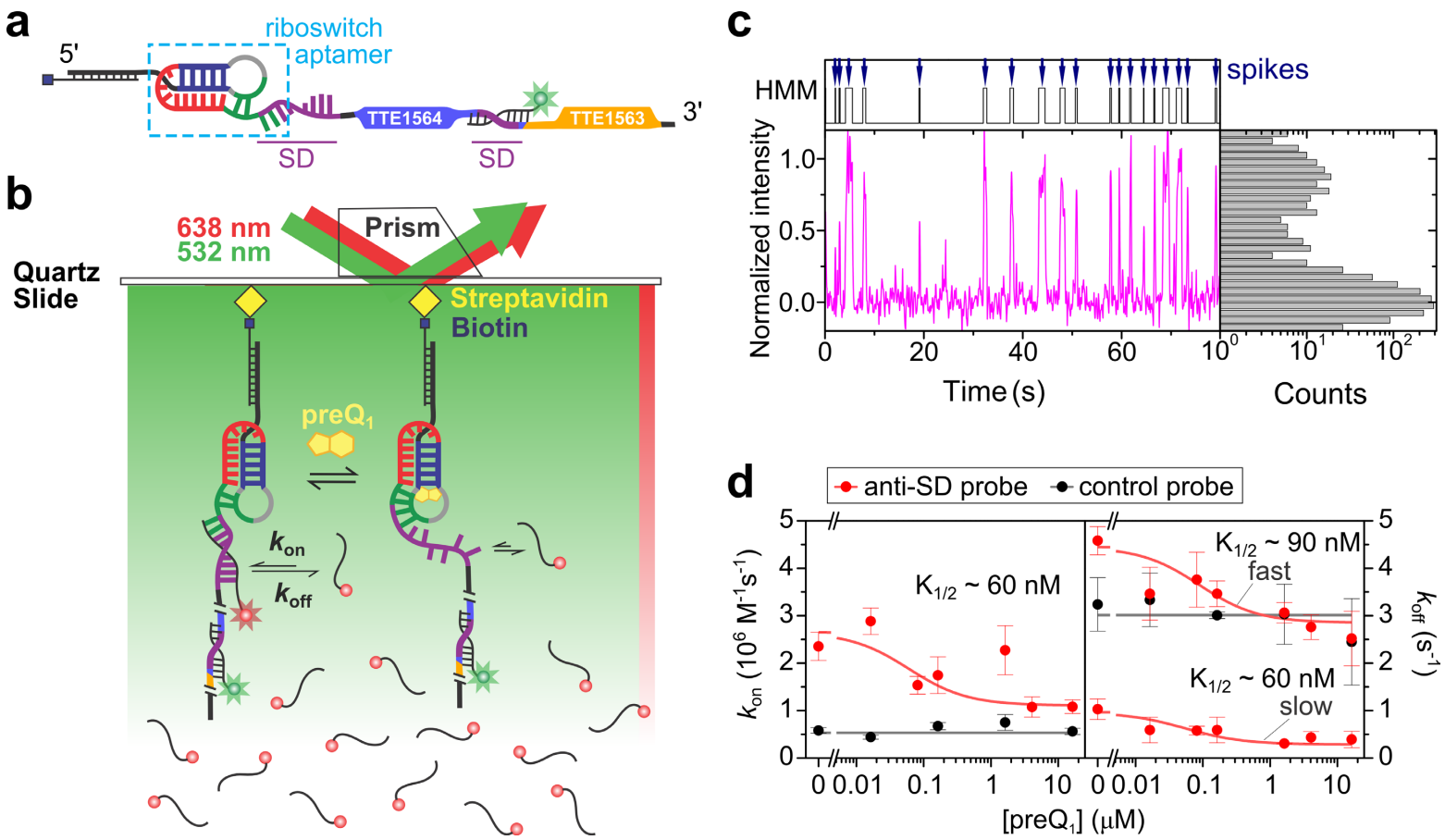


Figure 2-4 SiM-KARTS measurements of preQ$_1$-dependent anti-SD binding kinetics.

(**a**) *Tte* mRNA complex used in SiM-KARTS experiments. Full-length *Tte* mRNA molecules are immobilized to the slide surface via a biotinylated-capture strand that is hybridized to the 5' end of the mRNA. Features of the riboswitch and associated reading frames are colored as in **Figure 2-1a** and **b**, respectively. A TYE563-LNA (with green star) is hybridized to the start of the downstream open-reading frame to occlude this second SD sequence and to locate mRNAs on the slide surface. (**b**) Experimental prism-based TIRFM setup. The *Tte* mRNA complexes shown in **a** are immobilized to a slide surface that has been passivated with biotinylated-BSA (omitted for clarity). Repeated binding and dissociation of the anti-SD probe labeled with Cy5 (red sphere, red star) is monitored through co-localization of TYE563 and Cy5 fluorescence.

Figure 2-4 SiM-KARTS measurements of preQ$_1$-dependent anti-SD binding kinetics (continued)

(**c**) Representative anti-SD probe binding fluorescence versus time trajectory and corresponding fluorescence intensity histogram for a single *Tte* mRNA molecule in the absence of preQ$_1$. Cy5 intensity from the anti-SD probe (magenta) and Hidden Markov idealization to a two-state model (HMM, gray) are plotted as a function of time. The TYE563 fluorescence trace used to identify and localize the *Tte* mRNA has been omitted for clarity. (**d**) Anti-SD (red) and control (black) probe binding and dissociation rate constants (k_{on}, left plot; k_{off}, right plot) were determined from exponential fits of dwell times in the unbound and bound states, respectively, as a function of preQ$_1$ concentration. Binding and dissociation rate constants for the control probe are unaffected by preQ$_1$ concentration. The corresponding $K_{1/2}$ value from the saturation curve fit of the anti-SD probe binding is indicated. The results represent the average ± standard error of the mean of at least three independent experiments.

bring about the regulatory control needed by the bacterium.

Target mRNA molecules were hybridized with a high-melting temperature TYE563-labeled locked nucleic acid (TYE563-LNA) for visualization, immobilized on a quartz slide at low density via a biotinylated capture strand, and imaged with single molecule sensitivity by total internal reflection fluorescence microscopy (TIRFM, **Figure 2-4a, b**). To simplify our analysis in the context of the full-length mRNA, we chose the TYE563-LNA marker to also block the distinct SD sequence and start codon of the TTE1563 ORF, preventing the anti-SD probe from binding to the downstream TTE1563 SD (**Figure 2-4a** and **Figure A.2-1**). TYE563 fluorescence could only be observed once all three components (biotinylated capture strand, *Tte* mRNA and TYE563-LNA) were assembled on the surface (**Figure 2-4b** and **Figure A.2-2**), attesting to the high specificity of the experiment.

2.3.3 SiM-KARTS allows for detection of the binding and dissociation events of single anti-SD probe molecules

Because the interaction between the *Tte* mRNA and the anti-SD probe is limited to seven Watson-Crick base pairs and one wobble pair(**Figure 2-1a**), binding of the probe to a single

mRNA molecule under equilibrium conditions is reversible and transient (**Figure 2-4b**). Additionally, since the experiment is performed using TIRFM, only probe molecules transiently immobilized to the slide surface via the mRNA target will be observed within the evanescent field and co-localized with TYE563 in a diffraction-limited spot, whereas probes diffusing freely in solution will instead contribute to a modest background fluorescence. Repeated, transient diffraction-limited co-localization of Cy5 and TYE563 fluorescence therefore unambiguously characterizes individual binding events of the anti-SD probe to a single target mRNA molecule (**Figure 2-4b**). Since changes in the probe binding and dissociation time constants can be sensitively monitored over an arbitrarily long time window with high precision, these characteristic repeat signals are expected to quantitatively report on the accessibility of the SD sequence and thus secondary structure of individual mRNA molecules.

Previous SAXS[63], smFRET and MD simulation data[83] on the *Tte* preQ$_1$ riboswitch aptamer have found that helix P2 of the riboswitch pseudoknot is partially open in the absence of preQ$_1$, leaving the SD sequence more exposed than in the presence of ligand (**Figure 2-1a**). To ask whether SiM-KARTS detects the expected difference in accessibility of the SD sequence under equilibrium conditions in the absence and presence of preQ$_1$, Cy5-labeled anti-SD probe was flowed onto a slide with immobilized and TYE563-LNA-bound *Tte* mRNA first in the absence of ligand. Demonstrating the highly parallel nature of SiM-KARTS, thousands of transient binding events were observed in over 100 mRNA molecules per experiment. In addition, the transient increases and decreases in Cy5 fluorescence intensity occured on a much faster timescale than is expected for photoblinking or photobleaching, allowing us to confidently attribute these changes in Cy5 intensity to binding and dissociation of the Cy5-labeled anti-SD probe.

The resulting Cy5 emission trajectories were fit using a two-state Hidden Markov Model (HMM) to extract dwell times of the probe in the bound and unbound states, τ_{bound} and $\tau_{unbound}$, respectively (**Figure 2-4c**). The use of HMMs allows us to objectively identify binding events in single molecule fluorescence time trajectories that are inherently noisy due to the background of excess free Cy5-labeled probe in solution. HMMs filter out this noise and can detect binding events as short as a single camera integration time. HMMs also offer an advantage over simple signal thresholding or image spot finding algorithms because HMMs inherently consider the kinetics of binding events that are the key information sought from SiM-KARTS. Using HMM analysis, we found that a lower-than-average signal-to-noise ratio in some trajectories sometimes could lead to misidentification of the molecule's state (bound or unbound). To determine what, if any, effect this has on the results of these experiments, we undertook a more systematic examination of the influence of user intervention in the identification of binding events during HMM fitting.

2.3.4 Slight variability in trace idealization is well tolerated in SiM-KARTS

Initial HMM fitting following preprocessing of the trace data as described in **Materials and Methods 2.5.4** produces a reasonable idealization for the majority of molecules; however, additional adjustment of individual traces is frequently necessary. We refer to the degree to which the idealization for each molecule is scrutinized as supervision. To examine the effects of variations in idealization, a dataset comprised of 288 single molecule fluorescence time trajectories was independently idealized three times, exercising a different level of supervision each time (None, Moderate, or Heavy). The initial idealization of traces, as described above without any further adjustment or intervention on the part of the user, is referred to as unsupervised idealization. As seen in the example trace in **Figure 2-5a,** the initial unsupervised

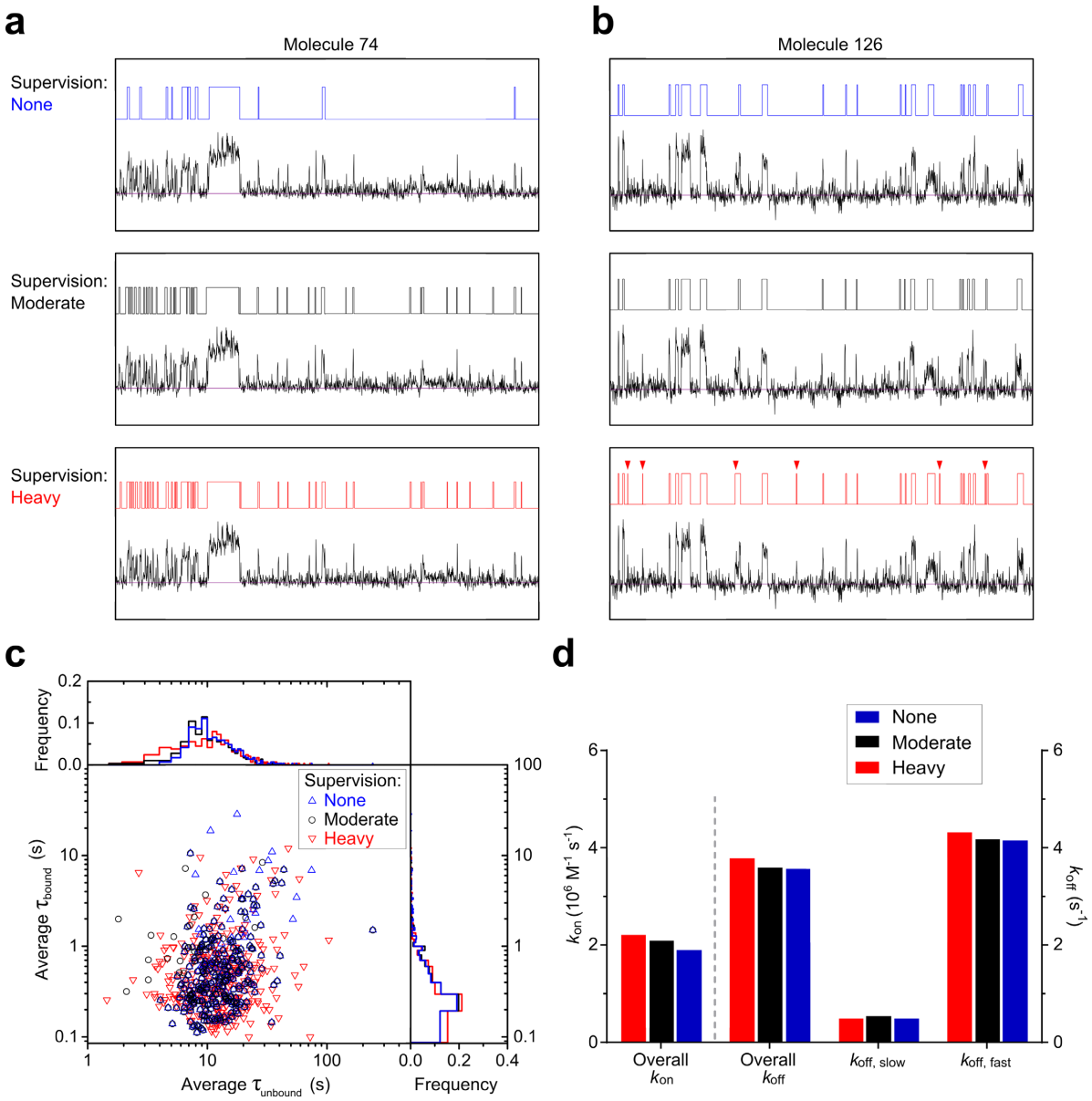


Figure 2-5 Influence of variation in idealization on SiM-KARTS results.

Fluorescence time trajectories for the 288 molecules in the 'No blocking strand' dataset (described in **Results 2.3.11** and **Figure 2-10c**) were idealized to a two-state model using the QuB suite software as described in **Results 2.3.4** and **Materials and Methods 2.5.4.** (**a**) Example single molecule trace and Hidden Markov idealizations (HMM) from QuB with varying degrees of supervision. The 'Heavy' and 'Moderate' supervision HMMs are identical. (**b**) Example single molecule trace as in **a**. Red arrowheads highlight differences between HMMs generated with 'Heavy' and 'Moderate' supervision. (**c**) Average dwell time correlation plot for the same set of molecules employing different levels of supervision during idealization. Heavy or Moderate adjustment of individual trace idealizations (Heavy, red triangles; Moderate, black circles) increases the range of observed values for average $\tau_{unbound}$ and τ_{bound}, but the center of the distribution is unchanged. (**d**) Comparison of rate constants for the same molecules idealized with different levels of supervision. Idealization-dependent changes in rate constant are modest.

idealization (None, blue HMM) sometimes inadequately fits genuine binding events. In these instances, the idealization for the specific molecule is repeated after adjusting the model's starting parameter estimates (Moderate or Heavy supervision). We found that re-estimating the mean amplitudes and standard deviations of each state for the specific trace using the Amps function in the QuB suite software, or simply decreasing (or, in the case of over-fitting, increasing) the standard deviation of the unbound state, and then repeating the idealization is sufficient to achieve an accurate HMM fitting (**Figure 2-5a**, black and red HMMs).

Adjusting the idealization for a given molecule only when there were obvious deficiencies in the HMM fitting as found in the example in **Figure 2-5a** was considered as "exercising Moderate supervision" over the idealization. As a result, the HMM idealization for the majority of molecules was often unchanged from the initial unsupervised idealization (**Figure 2-5b**; compare None, blue, and Moderate, black, HMM fits). By contrast, when exercising Heavy supervision, very close attention was paid to the idealization for each trace, which led to frequent adjustment in an effort to capture every increase in fluorescence intensity that could reasonably be interpreted as a binding event (**Figure 2-5b**; red arrowheads highlight the differences between idealizations in which Heavy, red, or Moderate, black, supervision was employed). This extent of refinement of a given trace's idealization occasionally required changing additional parameters such as the maximum number of iterations used in the fitting algorithm, adjusting the initial rate estimates used by the model, or examining the histogram of fluorescence intensities to visually estimate the center of the bound state amplitude (mean intensity) and possibly employing a large, fixed value for the standard deviation for the bound state intensity. The latter was often required if a trace exhibited a relatively small number of short-lived binding events with dramatically different intensities, leading to a very broad distribution of bound state intensities that was too

sparsely populated within the observed time window to be fit well by the algorithm.

From the plot of average dwell times, one can clearly see that Moderate and Heavy supervision of the idealization process results in slightly greater dispersion of average unbound dwell times and a slight compaction of average bound state dwell times, as would be expected after correction of under- or over-fitting of binding events, and under-fitting of short bound dwell times (**Figure 2-5c**). However, the overall distributions remain largely unchanged. This fact is perhaps best reflected by the calculated rate constants: while there are slight differences in the calculated rate constants depending on the degree of supervision used, overall the calculated rates are little affected (**Figure 2-5d**).

With the obvious exception of **Figure 2-5**, all of the SiM-KARTS data in the current study were idealized employing what we describe above as Moderate supervision. Completely unsupervised idealization is clearly insufficient in the case of some molecules, as demonstrated above (**Figure 2-5a**). While it is expected that some genuine binding events will be missed with only Moderate supervision, the use of Heavy supervision during idealization introduces an undesirable degree of subjectivity into the analysis, and thus will likely be more prone to bias and over-fitting. Given that the level of supervision ultimately had little impact on the final rate constant analysis (**Figure 2-5d**), we conclude that the analysis of SiM-KARTS data is robust, provided there is a sufficient number of molecules in each dataset, and that slight variability in idealization (e.g., the occasional missed binding event) is well tolerated.

2.3.5 The anti-SD probe binding rate is not diffusion-limited

Key to simplifying the interpretation of binding event data is establishing experimental conditions under which measurements of probe dissociation is not limited by photobleaching and

probe binding is not limited by diffusion. Given the relatively fast rates of dissociation and the concentration of anti-SD probe (dynamically refreshed from the bulk solution), Cy5 fluorophore bleaching does not affect the observed off rates as described above[87,88]. Following a similar method to that described by Dupuis *et al.*[89], we measured the diffusion coefficient of the anti-SD probe (D_{probe}) in SiM-KARTS buffer to be $0.20 \pm 0.03 \times 10^{-6}$ cm^2 s^{-1} using fluorescence correlation spectroscopy. The diffusion limited rate-constant k_{diff} in terms of M^{-1} s^{-1} can be estimated using **Eq. 1**:

$$\boldsymbol{k_{\mathrm{diff}} = 4\pi r_{RNA} D_{probe} \frac{N_A}{1000\ \mathrm{cm}^3}} \qquad \textbf{(1)}$$

where N_A is Avogadro's number, and the radius of the 12 nt anti-SD probe, r_{RNA}, is taken to be approximately 24 Å, a value that is slightly larger than expected for a duplex of the same general length (20.5 Å for an 8-mer or 12-mer helix[90])[89]. This predicts a diffusion-limited rate $\approx 4 \times 10^8$ M^{-1}s^{-1}, which is significantly faster than the values we measured for k_{on} (that are on the order of 10^6 M^{-1} s^{-1}, see below) demonstrating that the hybridization of the anti-SD probe is not limited by diffusion.

2.3.6 Anti-SD probe hybridization kinetics are in line with previously reported values

After fitting the single molecule traces with HMMs, dwell times in each state (bound and unbound) were extracted from the idealized data. The resulting cumulative τ_{bound} and $\tau_{unbound}$ dwell time distributions were fit with a single-exponential function to calculate the binding rate constant k_{on} and a double-exponential function to extract a fast and a slow dissociation rate constant k_{off}, respectively, based on an analysis of the residuals (**Figure A.3-1**). In the absence of ligand, the anti-SD probe binds with a bimolecular rate constant k_{on} of $2.4 \pm 0.3 \times 10^6$ M^{-1} s^{-1} and dissociates with two unimolecular rate constants k_{off} of 4.6 ± 0.2 s^{-1} (relative amplitude = 76%)

and 1.0 ± 0.2 s^{-1} (relative amplitude = 24%) (**Figure 2-4d, Table 2-1**). As our further analysis will demonstrate, the apparent biphasic nature of k_{off} is simply the result of the shortcomings of a phenomenological fit to a set of complex, kinetically broadly distributed molecular behaviors. Another important aspect of our SiM-KARTS approach is that its true power is derived from the relative changes in binding kinetics of the probe, which is used to report on the conformational state of a larger RNA (as discussed below in **Results 2.3.8**). Although we do not measure the concentration dependence required to rigorously determine the bimolecular rate constant k_{on} according to the linear relationship:

$$\frac{1}{\tau_{\text{unbound}}} = k_{\text{on,observed}} = k_{\text{on}} c \qquad \textbf{(2)}$$

where c is the concentration of anti-SD probe, there is a strong precedent in the literature demonstrating this relationship for related experimental systems involving short nucleic acid duplexes[88,89]; because our k_{on} values (presented in **Figure 2-4d** and **Table 2-1**), determined from experiments at a single concentration of anti-SD probe, are in excellent agreement with k_{on} rate constants determined previously under similar ionic conditions, (e.g., Jungmann *et al.*[88] 2.3×10^6 $M^{-1}s^{-1}$ for a 9 bp duplex with 600 mM NaCl), we conclude that our values are a good approximation of the true bimolecular k_{on} in our system.

2.3.7 SiM-KARTS detects ligand-induced secondary structure changes in single mRNA molecules

To begin examining the ligand-dependent effects on the accessibility of the SD sequence, mRNA molecules were folded in the presence of varying concentrations of $preQ_1$ and subjected to equilibrium SiM-KARTS. The value of k_{on} of the anti-SD probe decreased as the concentration of $preQ_1$ increased, with a half-saturation point $K_{1/2}$ of approximately 60 nM

preQ$_1$ (**Figure 2-4d**). Such a decrease in the binding rate indicates an occlusion of this target sequence in the presence of preQ$_1$, as expected (see **Results 2.3.8**). It is important to note that this $K_{1/2}$ is related to, but is not a direct measure of, ligand affinity, which is known to be in the low nanomolar range[63,83]. Instead it reflects preQ$_1$-induced structural changes in the expression platform, in contrast to previous studies performed only in the context of the minimal aptamer[83]. The interplay between the expression platform and aptamer domain of a riboswitch is complex and it is therefore reasonable to assume that additional aspects of the RNA structure contribute to the apparent $K_{1/2}$ for changes in the expression platform, beyond simple binding of the ligand. Unexpectedly, increasing preQ$_1$ concentrations also resulted in a decrease in both the fast and slow k_{off} rate constants (**Figure 2-4d**), indicating that high preQ$_1$ concentrations stabilize the SD:anti-SD interaction once formed. A plausible explanation for such an effect is the potential for preQ$_1$ to stabilize co-axial stacking of the anti-SD probe onto a more fully formed P2 helix (**Figure 2-1a**).

2.3.8 Ligand-dependent changes in anti-SD probe binding kinetics are consistent with expectations based on previous studies of short duplex annealing kinetics

Integral to the regulation exerted by the riboswitch is the inherent nuance in differences of hybridization kinetics between a SD:anti-SD duplex with 6 versus 8 possible base pairing interactions (**Figure 2-6a**). As such, it is useful to discuss recent work examining the kinetics of such short complementary oligonucleotides.

In a study by Dupuis *et al.*, hybridization kinetics were observed by smFRET for a series of short, fully complementary duplex DNAs, revealing that from a 6-bp to an 8-bp duplex the τ_{bound} increased by approximately 100-fold[89]. In addition, the authors observed a linear relationship between increasing duplex length and decreasing $\Delta G°$ of hybridization, where $\Delta G°$ decreased by

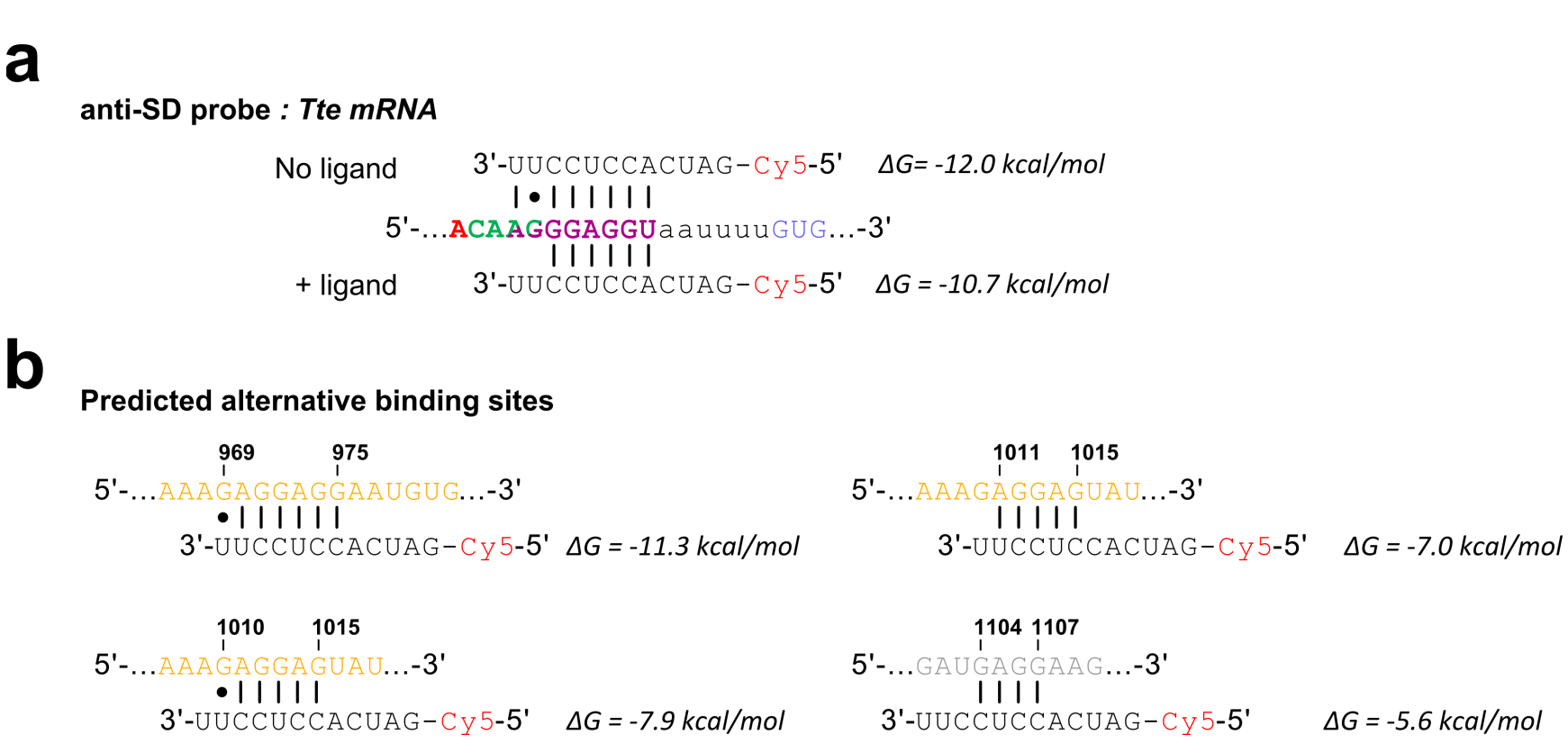


Figure 2-6 Base-pairing interactions between the anti-SD probe and *Tte* mRNA.

(**a**) Schematic showing differences in putative binding modes between the anti-SD probe and the TTE1564 SD sequence in the *Tte* mRNA in the presence and absence of bound preQ$_1$ ligand. Features of the riboswitch and mRNA are colored as in **Figure 2-2a**. The fluorophore Cy5 (red) is covalently attached at the 5' end via an aminohexyl linker. (**b**) Alternative binding sites for the anti-SD probe and associated binding free energies predicted using RNAstructure v5.7. Nucleotide numbering for the *Tte* mRNA (top lines) is relative to the start of the transcript used in SiM-KARTS experiments (**Appendix A.4**). Nucleotides in orange are part of the TTE1563 ORF; nucleotides in gray are part of the putative transcript's 3' UTR.

~1 kcal/mol/bp as one might intuitively expect – formation of a longer duplex is more favorable than formation of a shorter one. Their examination of perfectly complementary duplexes thus demonstrates that a difference of just two base pairs has an outsized impact on duplexes of this length. Applying this observation to our own study, it is reasonable to expect ΔG to change by ~ +2 kcal/mol as two base pairs in the SD sequence become unavailable due to ligand binding.

Interestingly, the prior study found that the effect of changing duplex length was primarily on k_{off} (140 s^{-1} versus 0.40 s^{-1} at relatively low ionic strength for perfect 6 and 8 bp DNA duplexes, respectively) and showed that k_{on} only slightly decreased with increasing duplex length (from 5.0×10^6 $M^{-1}s^{-1}$ to 3.5×10^6 $M^{-1}s^{-1}$). However, such specifics are undoubtedly context dependent since it has been known since the 1970's that successful binding events are initiated by a few metastable basepairing interactions, followed by zippering of the remaining base pairs (as summarized in the recent Ref. 91), and they are influenced by probe and target secondary structures (see, e.g., Ref. 92).

We can draw closer parallels to the single molecule FRET studies of Cisse *et al.*[93] examining the position-dependent effects of internal mismatches on the hybridization and dissociation kinetics of a 9-bp duplex. In a 9-bp DNA duplex with similar pattern of weak and strong base pair interactions as our anti-SD probe, internal mismatches resulting in less than 7 contiguous base pairs showed a 30-fold increased k_{off} and 100-fold decreased k_{on}. The authors also found the same to be true for duplex RNA. Their findings highlight the sensitivity of the binding and dissociation kinetics to changes in the number of basepairing interactions in this length regime, leading them to postulate that the observed 7-bp complementarity may play a role in target discrimination by the seed sequence of microRNA.

In summary, our measured rate constants are well within the expected range observed in previous studies[88,89,92,93 and ref. therein], and it is very reasonable to expect that there will be robust and measurable changes in the hybridization and dissociation kinetics for duplexes that when annealed are 6 or 8 bp, respectively. The exact direction and magnitude of a change, however, will be difficult to predict. Our system is more complex than earlier studies examining short duplex binding kinetics. Of course, formation of helix P2 will dynamically and competitively exclude part of the SD sequence our anti-SD probe binds, and there is the possibility of interactions between the probe and adjacent secondary and tertiary structure in the mRNA, which may further modulate the on- and off-rates (for example, we discuss the possibility for co-axial stacking of P2 with the anti-SD probe-SD helix leading to a slower than expected k_{off}, see above). Additionally, other factors such the sequence dependence on the opening and closing rates of the helix's closing base pair[94], and the sometimes non-intuitive stabilization afforded by different combinations of 3′ dangling nucleotides in RNA helices[95] can also influence the final observed rate constants.

2.3.9 Periods of high SD sequence accessibility occur in bursts

Further inspection of individual probe binding trajectories revealed that single molecules interconvert between periods of frequent probe binding events and periods of more sporadic events, which can be interpreted as periods of high and low SD accessibility, respectively (**Figure 2-7a)**. Traditional analysis methods of single molecules in aggregate failed to detect these changes. For example, common scatter plots of the mean τ_{bound} and τ_{unbound} dwell times for individual molecules[87,96-100] in the presence of saturating ligand concentration revealed a shift towards longer unbound times compared to the absence of ligand; however, all molecules generally fit within a single broad distribution (**Figure A.3-2a**). This observation suggests that

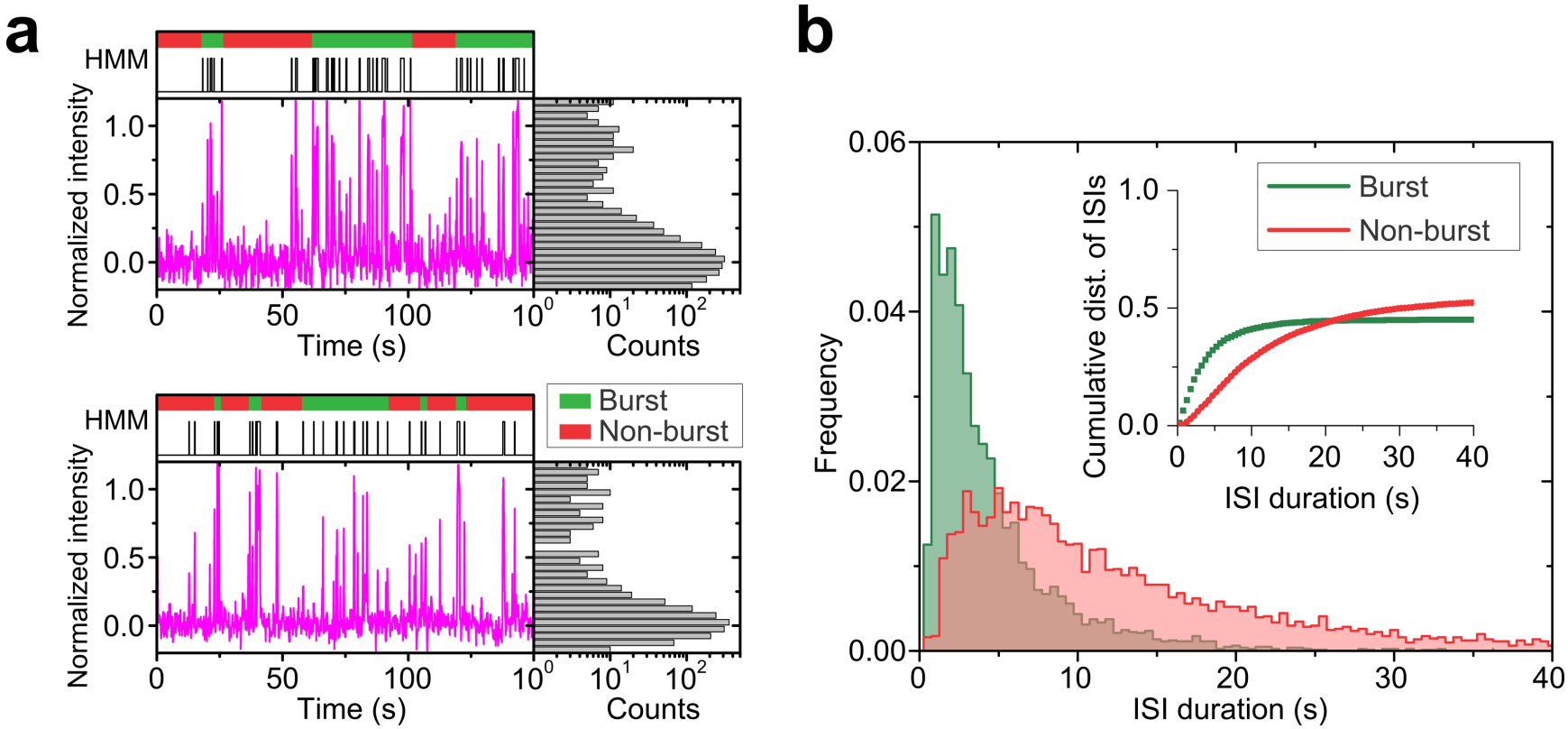


Figure 2-7 Detection of burst behavior through spike train analysis.

(**a**) Representative trajectories as in **Figure 2-4c** for two single *Tte* mRNA molecules in the absence of preQ$_1$, annotated with bursts (green bars) and non-burst periods (red bars) detected through spike train analysis. (**b**) Cumulative histogram displaying the distribution of interspike intervals during burst (green) and non-burst (red) periods in the absence of preQ$_1$.

calculating average rate constants for each mRNA molecule, while revealing heterogeneity among a population of molecules[87,96-100], largely fails to detect time evolution in the equilibrium behavior of a single molecule.

The probe binding events detected via SiM-KARTS strongly resemble neuronal spike trains, where neuronal firing is monitored and detected as sharp, transient increases (or "spikes") in electrical activity in response to external stimuli. A common feature of these spike trains is short intervals of high firing activity, or "bursts", separated by periods of relative inactivity (non-bursts)[101]. This type of analysis previously has been used to describe transcription time series in *E. coli*[102,103], leading us to ask whether spike train analysis could detect and separate in unbiased fashion the periods of high and low frequency of probe binding events observed within single molecules. To specifically justify its use in the context of SiM-KARTS, we first calculated the Fano factor[104] of the number of spikes within a certain time window (**Figure 2-8**). The Fano Factor, F, is defined as

$$\boldsymbol{F} = \frac{\boldsymbol{\sigma_w^2}}{\boldsymbol{\mu_w}} \qquad \textbf{(3)}$$

where the σ_w^2 is the variance of the number of spikes within a certain timeframe, w, and μ_w is the mean number of spikes in that time frame. For a random Poisson distribution, i.e., one that is completely independent of the time window, the Fano factor is consistently equal to one (simulated data points in **Figure 2-8**). Our SiM-KARTS data clearly deviate from this expectation (**Figure 2-8**), indicating spike train analysis is an appropriate tool. Next, exploiting its nonparametric approach we applied the Rank Surprise (RS) method of burst detection, which has been utilized to detect regions of high spike activity)[101] that, in our case, represent periods of high SD accessibility. The RS method does not make any assumptions about the distribution of

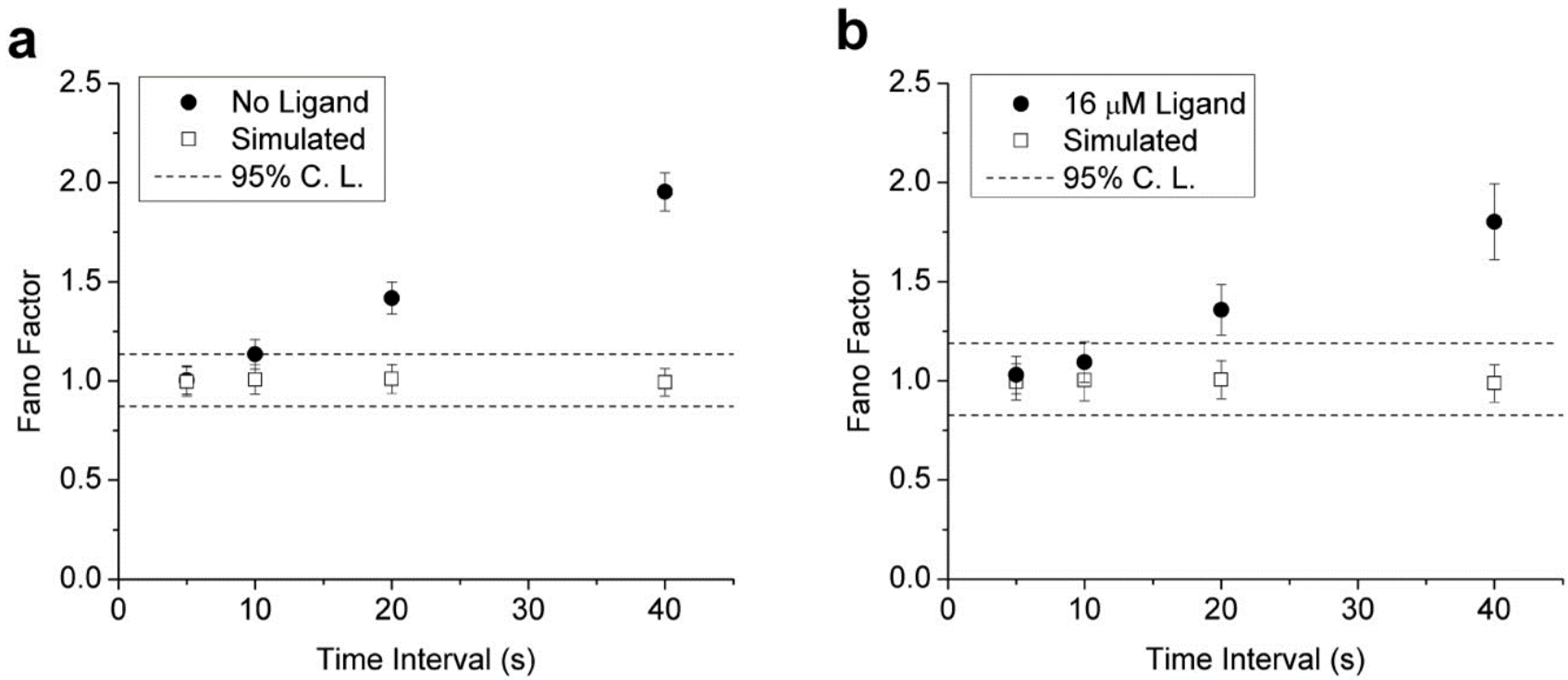


Figure 2-8 Fano factor for SiM-KARTS experiments.

The Fano factor was calculated across various time intervals for the no ligand (**a**) and saturating (**b**, 16 μM) preQ$_1$ datasets. The Fano factor for a simulated dataset with the same overall rate constant as the corresponding experimental condition but derived from a purely Poisson distribution was also calculated and plotted. The dashed lines indicate the 95% confidence level for a Poisson-like process with a specific number of trajectories (N = 445 for no ligand, and N = 234 for high ligand). The SiM-KARTS Fano factor values deviate from 1.0, indicating a non-random underlying distribution, while the simulated Poisson data remain close to 1.0.

spikes and is based solely on the definition of bursts as representing many spikes in a comparably short amount of time)[101]. Global burst analysis across the various concentrations of ligand tested found that individual molecules displayed detectable bursts of anti-SD probe binding, which were separated by non-bursting periods characterized by areas of low average binding activity (**Figure 2-7a**). When the duration of inter-spike intervals (ISIs), equivalent to $\tau_{unbound}$ dwell times, in the bursting and non-bursting periods were plotted, two distinct, previously hidden intramolecular behaviors became evident (**Figure 2-7b**). We found that single molecules typically interconvert between periods of bursting and non-bursting behavior, rather than segregating into separate subpopulations of highly- and poorly-accessible molecules, with bursts of high SD accessibility identified even at saturating ligand concentrations (**Figure 2-9a**). This finding suggests that *Tte* mRNA switches between (at least) two distinct conformational states: a bursting state with overall high SD accessibility and frequent binding events of the anti-SD probe (i.e., shorter ISIs); and a non-bursting state characterized by low SD accessibility where the SD sequence is more sequestered away from the probe (longer ISIs). The latter non-bursting state is adopted even in the absence of ligand (**Figure 2-9a**), in accord with previous studies indicating that the P2 helix partially sequestering the SD sequence can form without ligand present[63,83].

One important consideration is whether the act of observation itself influences the observed behavior. In the context of SiM-KARTS, the pertinent question is whether binding of the probe strand significantly alters the dynamics of the local structure it is intended to investigate (in this case, the SD sequence). The accessibility of the SD sequence is presumed to be in part a function of the structural dynamics of P2 formation. Based on previous single molecule studies of the aptamer domain[83], lifetimes of the partially and fully formed conformations of P2 are expected to

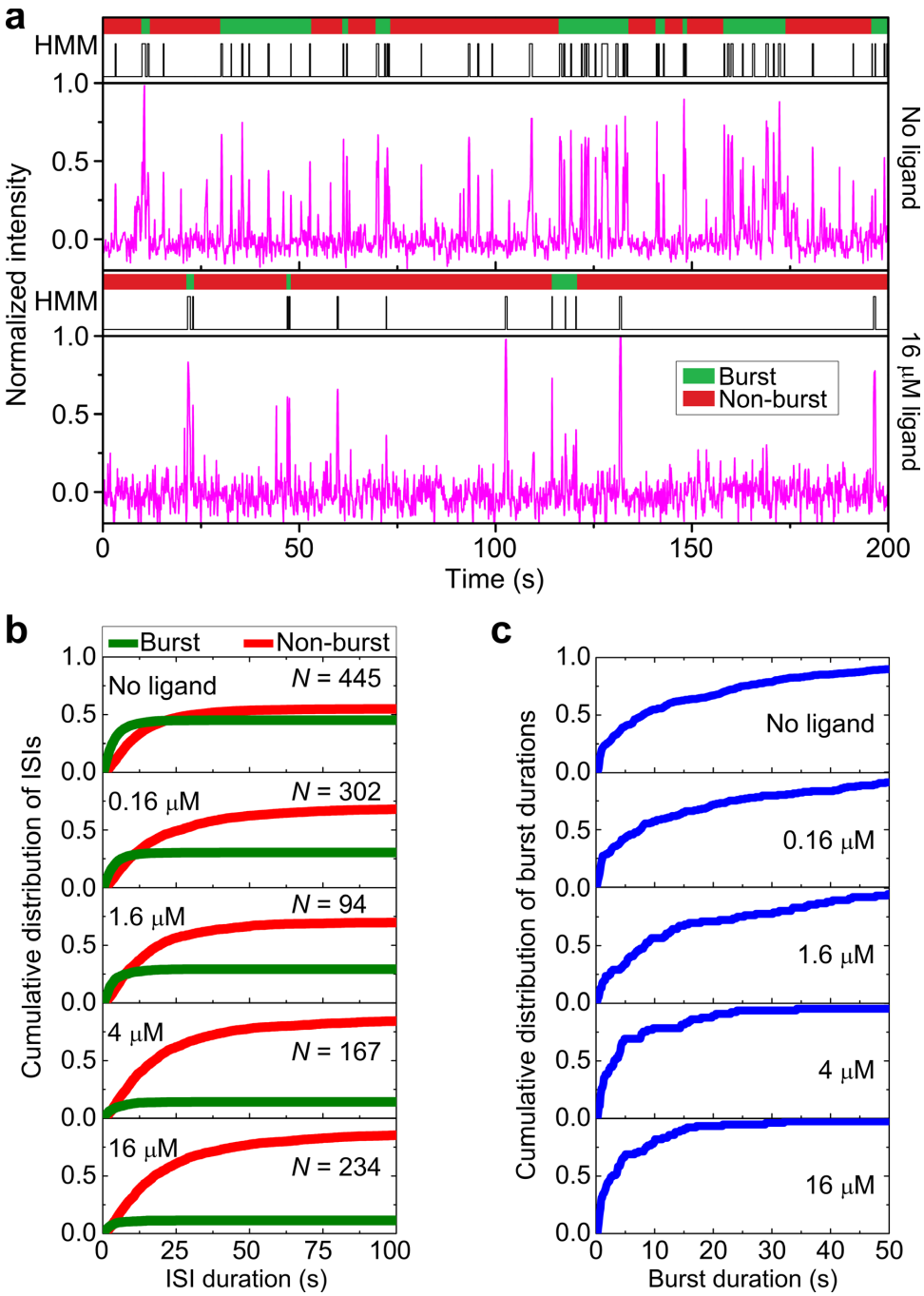


Figure 2-9 Ligand-dependent changes in bursting behavior of single riboswitches.

(**a**) Single molecule trajectories as in **Figure 2-7a** but in the absence and presence of saturating preQ$_1$ (top and bottom, respectively). (**b**) Cumulative distribution plots indicating the distribution of interspike intervals (ISIs) during burst (green) and non-burst (red) periods at varying preQ$_1$ concentrations, where *N* is number of molecules included in the analysis. (**c**) Cumulative distribution plots of burst duration for the molecules in **b** as a function of preQ$_1$ concentration. As preQ$_1$ concentration increases, the average burst duration decreases.

be relatively long under the buffer conditions used here for SiM-KARTS – on the order of seconds or, in the presence of ligand, perhaps tens of seconds. By contrast, the average probe binding duration (τ_{bound}) is only fractions of a second (**Figure A.3-2a**) and thus is much shorter by comparison. Because of the significant difference between the timescales of probe binding and dissociation and of conformational changes in the riboswitch fold, the probe is not expected to remain bound long enough to significantly interfere with the intrinsic structural dynamics of the *Tte* mRNA.

2.3.10 preQ$_1$ decreases specifically the number of bursts and the burst duration

Visual inspection of the trajectories at high ligand concentration suggested that the bursting state is shorter lived, and non-bursting periods longer lived, compared to low ligand concentration conditions (**Figure 2-9a**). In global spike train analysis, this observation is reflected in an increasing bias towards non-burst associated ISIs with increasing preQ$_1$ concentration (**Figure 2-9b**, **Table 2-1**). Furthermore, the cumulative burst duration distribution shifts towards shorter values (**Figure 2-9c**), indicating that the conformation in which the SD sequence is more accessible becomes shorter-lived and destabilized by ligand. However, the duration of ISIs, and thus the binding rate constant k_{on}, within bursting states is largely unaffected by ligand, indicating that the bursting state conformation is similar in the absence and presence of ligand, just that its lifetime is shorter (**Table 2-1**). Overall, our results suggest that the *Tte* mRNA with embedded preQ$_1$ riboswitch transitions between two distinguishable equilibrium conformational states: a bursting state conformation with an exposed SD sequence that is available for frequent binding of the anti-SD sequence, and a non-bursting conformation with a less accessible SD sequence. Both of these states co-exist and interconvert in both the presence and absence of ligand. As more ligand is added, transitions to the bursting state become less frequent and shorter

Table 2-1 Kinetic parameters extracted from SiM-KARTS and burst analysis.

The cumulative distribution plots corresponding to the ISI times inside and outside of the bursts shown in **Figure 2-9b** were fit with an exponential function from which the ISI half-lives ($t_{1/2}$) were calculated. No significant change is observed in $t_{1/2}$ values of the ISIs inside the burst; by contrast, a notable increase is observed in the ISIs outside the bursts as ligand concentration is increased. Similarly, the burst duration decreases with increasing ligand concentration.

Condition	Median burst duration (s)		k_{on} (10^6 M^{-1} s^{-1})	ISI $t_{1/2}$ (s)	k_{off} (s^{-1})[c]
No Ligand	7.95	Burst [a]	4.87 ± 0.02	2.85 ± 0.01	3.82 ± 0.07
		Non-burst [a]	1.47 ± 0.01	9.45 ± 0.03	3.74 ± 0.05
		Overall [b]	*2.4 ± 0.3*		*3.7 ± 0.2* (76 ± 5%)
0.16 μM	7.5	Burst [a]	5.23 ± 0.03	2.65 ± 0.01	2.83 ± 0.02
		Non-burst [a]	0.974 ± 0.002	14.2 ± 0.1	2.40 ± 0.02
		Overall [b]	*1.74 ± 0.4*		*2.8 ± 0.4* (79 ± 4%)
1.6 μM	8.3	Burst [a]	5.18 ± 0.04	2.67 ± 0.02	2.22 ± 0.04
		Non-burst [a]	1.22 ± 0.01	11.4 ± 0.1	2.43 ± 0.03
		Overall [b]	*2.3 ± 0.5*		*2.3 ± 0.2* (70 ± 7%)
4 μM	3.1	Burst [a]	4.65 ± 0.02	2.98 ± 0.01	2.45 ± 0.04
		Non-burst [a]	0.985 ± 0.002	14.1 ± 0.1	2.84 ± 0.03
		Overall [b]	*1.1 ± 0.2*		*2.2 ± 0.2* (76 ± 1%)
16 μM	2.8	Burst [a]	6.73 ± 0.05	2.06 ± 0.02	2.9 ± 0.1
		Non-burst [a]	0.888 ± 0.002	15.6 ± 0.1	2.46 ± 0.06
		Overall [b]	*1.1 ± 0.1*		*2.1 ± 0.4* (80 ± 6%)

a. Values were calculated from single exponential fits of the pooled data from all experiments for a given condition. The reported error is the standard error of the fit.

b. Values represent the average ± the standard error of the mean of three independent experiments.

c. Values represent the weighted average of the fast and slow rate constants derived from a double exponential fit. Percentages in parentheses indicate the contribution of the fast dissociation rate constant to the overall k_{off}.

lived, yet remain a persistent feature of the mRNA-embedded riboswitch.

2.3.11 Anti-SD probe binding frequency and duration are greatly decreased in the presence of a blocking strand, demonstrating specificity

We are not able to entirely rule out the possibility that some of the anti-SD probe binding events we observed are due to binding at sites other than the TTE1564 SD, the expression platform of the riboswitch. In fact, the presence of fluorescence traces exhibiting multistep TYE563 photobleaching indicates that it is possible for the TYE563-LNA to improperly hybridize at a site other than the TTE1563 SD and start codon, thus allowing binding by the anti-SD probe at the downstream TTE1563 SD for that subset of *Tte* mRNA molecules. However, only a low number (5-15%) of fluorescence traces exhibit such multistep TYE563 photobleaching, even in the presence of a stoichiometric excess of TYE563-LNA. This strongly suggests that although off-target sites for the TYE563-LNA exist, misannealing of the TYE563-LNA is rare and thus anti-SD probe binding at the TTE1563 SD is likely negligible.

Because the Cy5-labeled anti-SD probe has the same sequence as the 3′ end of the 16S rRNA, it has the potential to bind transiently to true SD sequences, as well as at SD-like sequences in the mRNA (**Figure 2-6b**). Indeed, recent work by Li and Weissman[105] has indicated a biologically important role for binding of the 16S rRNA to SD-like sequences in the open reading frame during translation. To test whether these changes are due to conformational rearrangements near the SD sequence of the riboswitch, and to assess the potential for the Cy5-labeled anti-SD probe to bind at other sites in the *Tte* mRNA, we performed equilibrium SiM-KARTS experiments as with an additional Cy3-labeled blocking strand present during initial complex heat annealing and dilution before immobilization on the slide surface. The Cy3-blocking strand hybridizes to a 35 nt region, sequestering the expression platform as well as the

initial five nucleotides in the TTE1564 ORF and disrupting the riboswitch aptamer domain, effectively preventing binding by the anti-SD probe at the TTE1564 SD (**Figure 2-10a**).

The fluorophores TYE563 and Cy3 attached to the LNA and blocking strand, respectively, have similar fluorescence emission profiles and thus traces that, in this experiment, exhibited two-step photobleaching indicate the presence of both the Cy3-blocking strand and the TYE563-LNA (**Figure 2-10b**). For these mRNAs, both the TTE1564 SD and TTE1563 SD are sequestered and so any observed binding events are due to binding at other sites on the mRNA. In SiM-KARTS experiments where the Cy3-blocking strand is present, the frequency of anti-SD probe binding events dramatically decreases for the majority of molecules, and the anti-SD probe stays bound for shorter periods; this is reflected in a marked shift towards longer average unbound dwell times, and a decrease in the average bound dwell time (**Figure 2-10c**). Importantly, analysis of the Fano factor for different time intervals reveals that, unlike the 16 μM data (**Figure 2-8b**), anti-SD probe binding to *Tte* mRNA heat-annealed in the presence of Cy3-blocking strand is a Poisson process and does not occur in bursts (or, at minimum, not on a comparable timescale).

Analysis of the binding and dissociation rate constants (k_{on} and k_{off}) shows that the on-rate is decreased in the presence of the Cy3-blocking strand, and the off-rate is significantly increased, approaching the limit of the time resolution used in our experiments (**Table 2-2**). Taken together, these data indicate that under the conditions of the SiM-KARTS experiments, binding of the anti-SD probe at sites in the mRNA other than TTE1564 SD occurs infrequently, and that the probe is weakly bound and dissociates quickly. In contrast, the k_{off} values measured in the in the presence of saturating (16 μM) preQ$_1$ and absence of the Cy3-blocking strand are several fold slower than would be expected if all observed binding events were due to probe

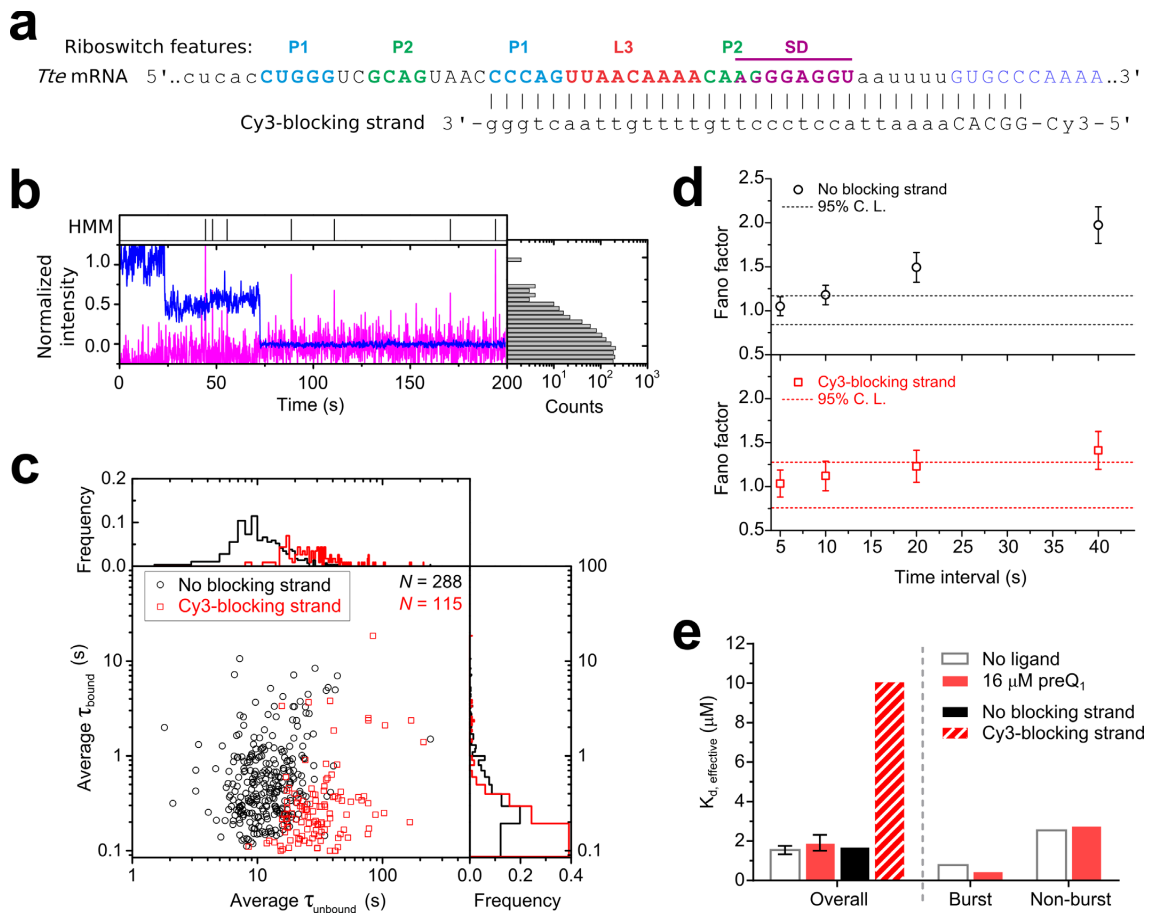


Figure 2-10 SiM-KARTS experiments on *Tte* mRNA with a blocked expression platform to examine anti-SD probe binding specificity.

(**a**) Schematic representation of the binding site of the Cy3-labeled blocking strand. Nucleotides comprising the various structural features of the riboswitch are colored as in **Figure 2-1a** and **Figure 2-2a**. (**b**) Example fluorescence-time trace, Hidden-Markov idealization (HMM), and Cy5 intensity histogram from SiM-KARTS experiments performed with blocking strand. Only molecules that displayed two-step photobleaching in the TYE563/Cy3 channel (blue trace), indicating the presence of both the TYE-563 LNA and the Cy3-blocking strand, were selected for analysis. (**c**) Average dwell time correlation plot for mRNA molecules with and without annealed Cy3-blocking strand. *N* is the number of molecules in the dataset for the respective condition. (**d**) Plot of Fano factors calculated for the anti-SD probe binding data shown in **c.** Dashed lines indicate the expected 95% confidence level for a Fano factor corresponding to a Poisson process, given the number of molecules in the dataset. Error bars indicate the standard deviation of Fano factor values calculated from 100 samplings of the data (see Methods). (**e**) Comparison of the dissociation equilibrium constants ($K_{d,\ effective}$) derived from the measured k_{on} and k_{off} rates (**Table 2-1** and **Table 2-2**) for the anti-SD probe and *Tte* mRNA in the absence (No ligand) or presence of saturating (16 μM) $preQ_1$, and for *Tte* mRNA heat annealed in the absence (No blocking strand) or presence of the Cy3-blocking strand. Note that experimental conditions for the "No ligand" and "No blocking strand" were identical, but represent independently collected datasets. Error bars represent the propagated uncertainty from the overall rate constant values reported in **Table 2-1**.

Table 2-2 Kinetic parameters in the presence and absence of a blocking strand.

The cumulative frequency distributions of all unbound and bound dwell times for SiM-KARTS experiments shown in **Figure 2-10** were fit with single (k_{on}) or double (k_{off}) exponential association functions. In the presence of a Cy3-blocking strand that occludes the riboswitch expression platform (containing the TTE1564 SD, **Figure 2-10a**), anti-SD probe binding events are infrequent and short-lived. The $K_{d,\ effective}$ is derived using the measured rate constants (k_{off}/k_{on}) for probe binding. When the expression platform is occluded, the affinity of the anti-SD probe is solely due to any remaining binding sites elsewhere in the mRNA, and is greatly reduced.

Condition		k_{on} (10^6 M^{-1} s^{-1})	k_{off} (s^{-1})	$K_{d,\ effective}$ (μM)
No blocking strand	*Overall*	*2.1*	*3.6 (84%)*[a]	*1.7*
	Fast	--	4.2	
	Slow	--	0.54	
Heat anneal with Cy3-blocking strand	*Overall*	*0.82*	*8.3 (84%)*[a]	*10*
	Fast	--	9.6	
	Slow	--	1.1	

[a.] Values represent the weighted average of the fast and slow rate constants derived from a double-exponential fit. Percentages in parentheses indicate the contribution of the fast dissociation rate constant to the overall k_{off}.

binding at sites other than the TTE1564 SD (**Table 2-1** and **Table 2-2**), and is similarly evidenced by a population shift towards shorter average τ_{bound} times (**Figure A.3-2c**). Both of these observations lend support to the assertion that the anti-SD probe primarily reports on accessibility of TTE1564 SD and that binding at other sites would contribute to only a slight underestimation of k_{on} and overestimation of k_{off}. This finding is perhaps captured most clearly by the change in $K_{d,\ effective}$ for *Tte* mRNA heat-annealed in the presence of Cy3-blocking strand (**Figure 2-10e, Table 2-2**), where the affinity of the anti-SD probe for the *Tte* mRNA decreases significantly when the expression platform is blocked, to a level lower than what is seen in the presence of saturating ligand, in both Burst and Non-burst periods.

This observation supports the assertion that the vast majority of binding events observed, even under saturating ligand conditions, are indeed genuine. For example, one can assume that the bound dwell time distribution observed in the presence of 16 μM ligand is in fact the sum of "genuine" binding events and "non-specific" binding events (**Figure A.3-2c**, red histogram). One may also assume that the bound dwell time distribution for binding events observed for *Tte* mRNA annealed with the blocking strand represents the expected contribution of probe binding at sites other than the riboswitch expression platform i.e., "non-specific" binding (**Figure A.3-2c**, black histogram). Because relative distribution of bound dwell times for the blocked dataset is skewed towards binding events lasting only a single frame, if this dwell time distribution is scaled such that the number of single-frame binding events is equal to that observed in the 16 μM dataset, then subtracting the scaled, blocked dataset distribution from the 16 μM dataset distribution should remove the "non-specific" binding events (**Figure A.3-2c**, blue histogram). After doing this, a majority (> 60%) of binding events in the original distribution still remain. This is in fact highly-conservative because it assumes that all binding events lasting only one

frame in the 16 μM dataset are non-specific, which is unlikely to be the case, and thus overcorrects for the amount of non-specific binding.

Additionally, while the k_{on} of the anti-SD probe in the presence of the blocking strand (0.82 × 10^6 M^{-1} s^{-1}, **Table 2-2**) is comparable to that of the probe in the presence of high ligand concentration (1.1 × 10^6 M^{-1} s^{-1}, **Table 2-1**), the values of k_{off} differ substantially (8.3 s^{-1} versus 2.1 s^{-1}, respectively), again indicating that any spurious, non-specific events are characterized by much faster probe dissociation (**Figure A.3-2b**). This suggests that while both conditions (blocking strand or high ligand) induce a conformation in which the SD sequence is (partially) blocked, the ligand induces a distinctly different kinetic profile. This unequivocally demonstrates that most binding events we observe in our equilibrium SiM-KARTS experiments are localized at the upstream SD sequence.

2.3.12 Binding kinetics of a control probe are not sensitive to the presence of preQ$_1$

Finally, to further confirm the specificity of our assay, we performed equilibrium SiM-KARTS experiments using a control probe complementary to a region distal to the riboswitch and within the open reading frame of the mRNA (**Figure 2-4d**; **Figure A.2-1, Appendix A.4**), whose structure, and thus, accessibility to control probe binding should be unaffected by preQ$_1$. Indeed, the kinetics of the binding of this probe showed little change in response to preQ$_1$ concentration (**Figure 2-4d**), indicating that the ligand-induced conformational changes are localized to the SD region of the riboswitch.

2.3.13 Single mRNA molecules incompletely adapt to in situ *changes in preQ$_1$ concentration*

To assess the response of individual mRNAs and their SD accessibility to changing (non-equilibrium) ligand concentrations, as may occur in the bacterial cell, we devised a ligand-jump

experiment that allowed us to apply SiM-KARTS to molecules tracked throughout a transition from no ligand, to saturating ligand, and back to no ligand (*Minus*, *Plus*, *Minus′* segments, respectively) in a set of contiguous fluorescence-time trajectories (**Figure 2-11a**). We then applied global spike train analysis to all molecules we were able to track through this non-equilibrium ligand-jump experiment. To further characterize the evolution of the SD accessibility through the changes in ligand concentration, each of the three segments for a given molecule was ranked by time spent in the bursting state. An individual molecule's *Minus*, *Plus*, or *Minus′* segment with the highest density of bursts was ranked as High (H), the next highest as Mid (M), and the lowest as Low (L). We then quantified the overall distribution of burst ranks in the three different segments (**Figure 2-11b**). For the majority of molecules, the highest burst density occurred in the first segment where ligand is absent (*Minus*), as expected. By contrast, the saturating ligand segment (*Plus*) exhibited mostly Low- and Mid-burst density rankings, again as expected. Finally, the *Minus′* segment, where the ligand-containing buffer had been extensively washed out, exhibited a fairly equal distribution of ranks. These measures indicate that, as an ensemble, single mRNA molecules respond to ligand concentration with the expected modulation in SD accessibility. We further plotted the bursting behavior of each of 97 molecules as a rastergram and organized them into six groups according to their per-segment burst density ranks (**Figure 2-11c**), several of which were of particular interest. This comparison showed that ~24% of all molecules responded to the addition and removal of ligand with reduction and recovery of bursting behavior (H-L-M in the *Minus, Plus, Minus′* sequence; molecules 75-97 in **Figure 2-11c** and upper trace **Figure 2-11a**), respectively, as expected. However, ~30% of molecules seemed to not revert to High burst density on the timescale of the experiment after the ligand was washed out (H-M-L, molecules 46-74 in **Figure 2-11c**). This population of mRNAs

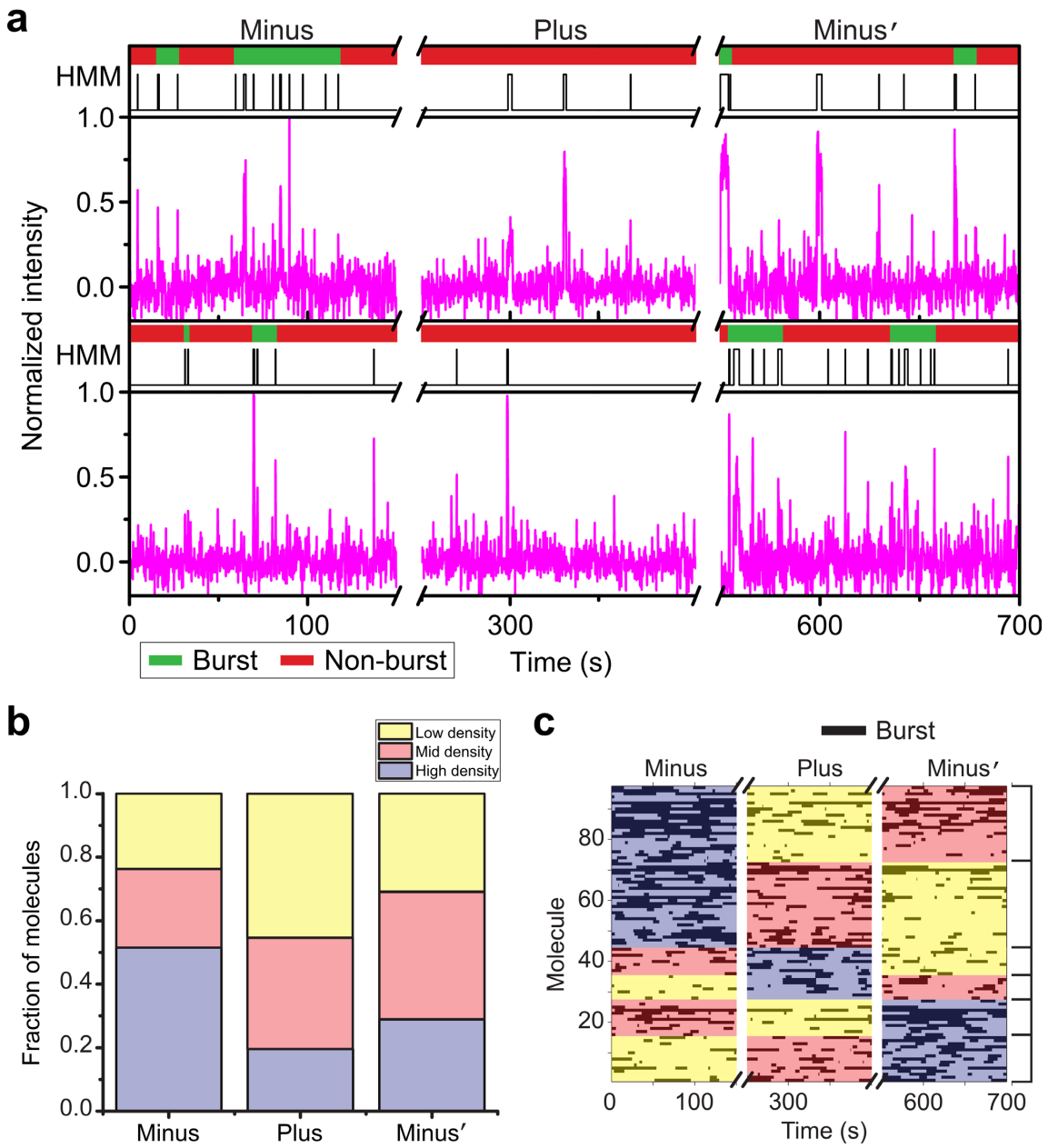


Figure 2-11 Single mRNA molecules can undergo conformational switching depending on their environment.

(**a**) Exemplary single molecule trajectories from ligand-jump SiM-KARTS experiments composed of three time segments. Anti-SD probe binding to the same set of individual *Tte* mRNA molecules is monitored first in the absence of preQ$_1$ (Minus), then in the presence of 16 μM preQ$_1$ (Plus), and again in the absence of preQ$_1$ (Minus'). Each axis break represents a 50 second dark period between segments during which buffer was exchanged. (**b**) Distribution of burst density rankings for each segment of an individual molecule's trajectory, for all molecules in the ligand-jump SiM-KARTS experiment (N = 97). (**c**) Rastergram displaying the bursting behavior of the *Tte* mRNA molecules in **b**. Bursts are displayed as black bars. Individual probe binding events (spikes) are omitted for clarity.

reacts to initial ligand binding by SD sequence occlusion, but appears to remain in a ligand-bound conformation even after the preQ$_1$-containing buffer is removed, consistent with the known slow rate of preQ$_1$ dissociation[63]. Interestingly, we also observed that ~16% of molecules displayed the opposite behavior, i.e., displayed their highest bursting density in the *Minus′* segment (L-M-H, molecules 1-16 in **Figure 2-11b**, lower trace in **Figure 2-11a**), after the introduction and removal of ligand. This behavior suggests that in some cases the ligand may help promote refolding of an mRNA in which the SD sequence was occluded prior to the addition of ligand.

2.4 Discussion

It is generally thought that translational riboswitches achieve gene regulation through a ligand-mediated conformational change in the aptamer domain that is then transduced into the downstream expression platform to actuate an ON/OFF switch in gene expression[58,67-70]. However, the molecular underpinnings of this transduction, especially in the context of the native gene, are still poorly studied and understood. Here, we present SiM-KARTS as a cost-effective and relatively non-invasive technique alternative to FRET that sensitively probes site-specific changes in secondary structure of arbitrarily large or complex single RNA molecules in real-time and without the need for covalently modifying the targeted RNA (**Figure 2-4b**), which can be an inefficient and expensive undertaking. Utilizing an anti-SD probe mimicking the 3′ end of the corresponding 16S rRNA, we used SiM-KARTS to quantify the accessibility of the SD sequence in the expression platform of an mRNA hosting a small preQ$_1$ riboswitch over a significantly longer time window than a similar FRET experiment, since SiM-KARTS is inherently not limited by photobleaching. The resulting extended observation of single molecules is critical for long RNAs that demonstrate relatively slow changes in structure, such as the

accessibility changes in the SD sequence of the mRNA studied here. As expected, we detected a decrease in SD accessibility upon addition of preQ$_1$ as a marked decrease in the binding rate constant k_{on} of the probe (**Figure 2-4d**). Less expectedly, however, probe binding events to a single mRNA (spikes) typically occurred in bursts (**Figure 2-4c, Figure 2-7a, Figure 2-11a**). Consequently, spike train analysis of these bursts provided evidence for two conformational states repeatedly interconverting within single molecules that are characterized by periods of high and low SD accessibility (**Figure 2-7a**). Unexpectedly, these two states were observed not only in the absence of preQ$_1$, but also at saturating ligand concentration, indicating that the riboswitch continues to occasionally sample a conformation with high SD accessibility (**Figure 2-9**). Furthermore, the main effect of increasing ligand is a clear change in the burst behavior of the *Tte* mRNA, and is reflective of the intrinsic, unimolecular folding of the *Tte* mRNA; the *identification and analysis* of these relative trends do not rely on, and are not sensitive to, the absolute values of the rate constants. Finally, non-equilibrium ligand-jump experiments indicated that single mRNA molecules only imperfectly switch between high and low SD accessibility (**Figure 2-11**). These findings rationalize the significant, but relatively modest impact of saturating preQ$_1$ concentrations on the *in vitro* translation output of the mRNA (**Figure 2-1d**) and lead us to propose the model for stochastic riboswitch-controlled gene expression depicted in **Figure 2-12**.

Notably, previous smFRET studies of the isolated *Tte* riboswitch also found two conformations populated from zero to saturating ligand concentration, identified as pre-folded and folded, that both respond to (i.e., "sense") ligand[83]. It is tempting to speculate that these two local conformations of the aptamer give rise to the two mRNA conformations that differentially bind the anti-SD during SiM-KARTS; the pre-folded state appears to have an only partially

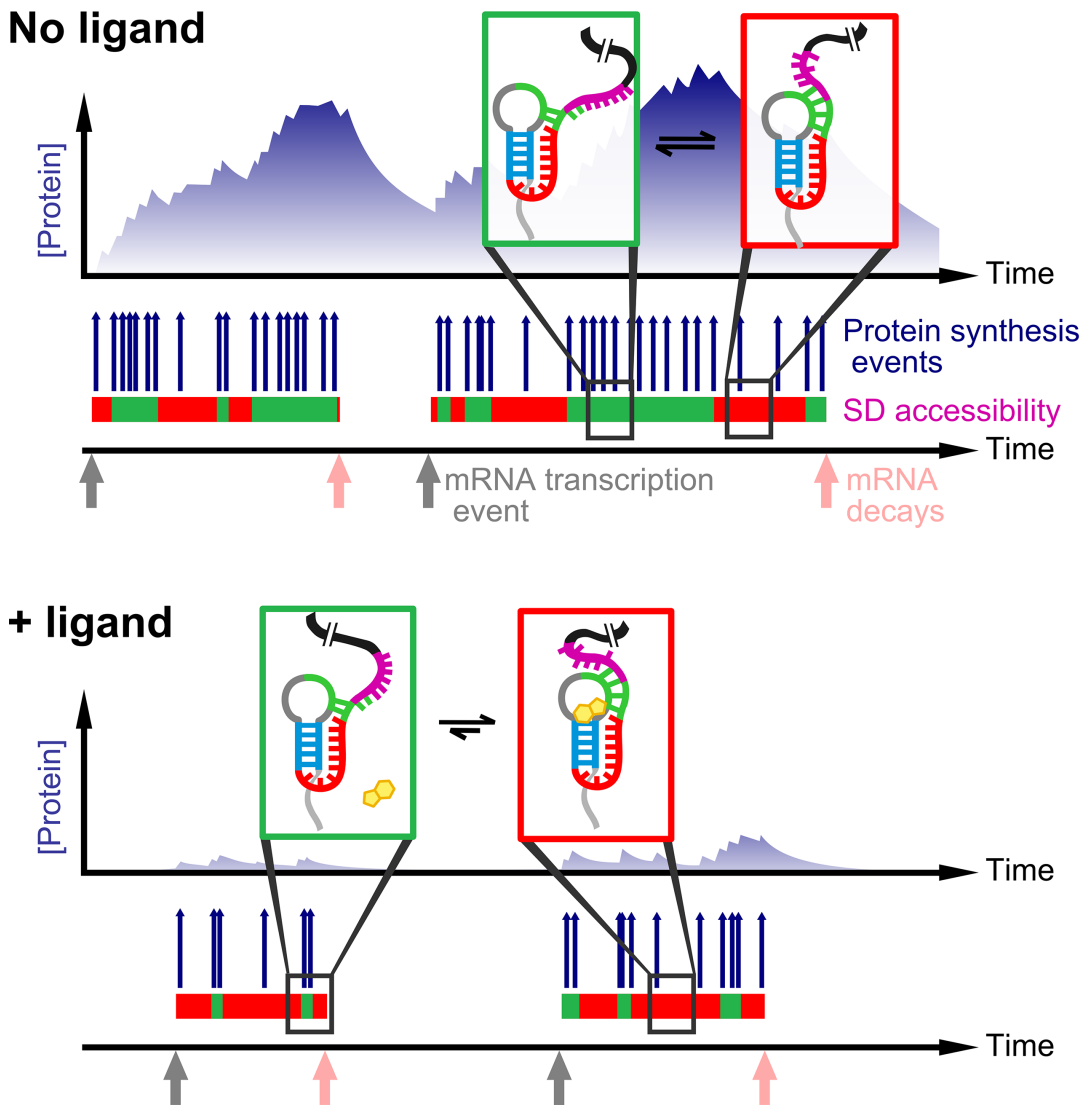


Figure 2-12 A stochastic burst model for preQ$_1$-dependent expression of the *Tte* mRNA.

In the absence of ligand (top), a transcription event (gray arrow) gives rise to an individual riboswitch-containing mRNA molecule that persists in the cell until it is degraded (light red arrow). During its lifetime, the mRNA transitions between burst (green bar segments) and non-burst (red bar segments) conformational states during which its SD sequence (purple nucleotides) is highly accessible and largely occluded, respectively. As reflected by increased anti-SD probe binding during SiM-KARTS, there are more potential opportunities for translation initiation events (blue arrows) in the burst state, leading to bursts of protein biosynthesis, than in the non-burst state. In the presence of the ligand (bottom), preQ$_1$ favors formation of the full P2 stem (green nucleotides) and occlusion of the SD sequence, resulting in shorter excursions to the burst state and, consequently, fewer opportunities for translation initiation. This may be accentuated by a decrease in the mRNA's lifetime (depicted as overall shorter burst/non-burst bars) resulting from its scarce occupancy with actively translating ribosomes, leading to significant down-regulation of protein expression. Coloring of the riboswitch cartoon is as in **Figure 2-1a** and **Figure 2-4a**.

formed P2 helix[83] that is expected to bind the probe more readily (resulting in a burst of spikes), whereas the folded state features a fully formed P2[83] that disfavors probe binding and thus leads to only sparse spikes (non-bursts, **Figure 2-12**). Consistent with the relatively modest binding free energy of -11 kcal/mol available from this small-molecule ligand[83], the mode of action of preQ$_1$ then is to subtly remodel the aptamer and only modestly reduce the hosting mRNA's translation initiation frequency, consistent with the observed reduction of *in vitro* translation product by ~40%. This seemingly moderate (~2-fold) change *in vitro* that can be directly attributed to the ligand suggests that regulation by preQ$_1$ is more nuanced than previously appreciated and that it may be potentiated by other forces at work in the cell, including: co-transcriptional folding of the mRNA from 5′ to 3′ end[106,107], competition between ribosome-catalyzed translation and RNase-mediated mRNA decay[108,109] (leading to a shorter lifetime of a sparsely translated mRNA, **Figure 2-12**), and repeated unfolding of the RNA resulting from close spacing between the transcription complex and the leading and sequentially loaded ribosomes[110].

The spike trains detected by SiM-KARTS resemble the transcriptional bursting that has been suggested as an underlying cause of genetic "noise"[103,111,112]. This generally stochastic nature of biological systems results in cell-to-cell variability and has been shown to be beneficial to organisms, particularly during times of environmental stress[113-115]. Each time a single mRNA molecule is transcribed, it gives rise to a few to tens of protein molecules[103]. We here show that translational bursting of an mRNA, which can be modulated by ligand binding to an embedded riboswitch, appears to add another layer of stochasticity to the gene expression cascade (**Figure 2-12**). We anticipate that such translational bursting will turn out to be a widespread phenomenon among mRNAs that generally present structurally dynamic ribosome substrates

whose SD region is known to significantly impact translation efficiency[116], and that our SiM-KARTS approach can detect changes in the secondary structure not just of single riboswitch-hosting mRNA molecules but of virtually any RNA under a wide range of conditions, poising it to find broad application.

2.5 Materials and Methods

2.5.1 Ligand and oligonucleotides

The preQ$_1$ ligand used in this study was synthesized as described previously[59] and was generously provided by Prof. George Garcia at the University of Michigan. DNA and LNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. (IDT) and Exiqon, respectively. Fluorophore-labeled DNA and LNA oligonucleotides were purified by reverse-phase HPLC by the respective manufacturer. The control and anti-SD probe RNAs were purchased from IDT with a 5′ aminohexyl-linker modification, and labeled with Cy5-NHS ester (GE Healthcare) as described previously[83]. The sequences of all oligonucleotides and cloning primers used in this study are listed in **Appendix A.4**.

2.5.2 RNA preparation

Reference genomic sequences for *Thermoanaerobacter tengcongensis* were downloaded from the National Center for Biotechnology Information (NC_003869.1, http://www.ncbi.nlm.nih.gov). The complete mRNA transcript, including the TTE_RS07450 and TTE_RS07445 (TTE1564 and TTE1563, respectively) ORFs, and its 3′ UTR as predicted from the FindTerm algorithm (SoftBerry), was amplified by PCR from *T. tengcongensis* genomic DNA, which was purchased from the NITE Biological Resource Center. The amplified region was cloned into the pUC19 plasmid between the BamHI and HindIII sites with an engineered

upstream T7 promoter. For use as a control for *in vitro* translation assays, an FspI site was introduced through site-directed mutagenesis[117] into the control vector provided with the ActivePro™ In Vitro Translation Kit (Ambion) containing the chloramphenicol acetyltransferase gene under the control of a T7 promoter (pAMB CAT). All of the plasmids used in this study are available through Addgene (www.addgene.org).

mRNA was produced by *in vitro* transcription. The *Tte* pUC19 plasmid was linearized with HindIII (AflII or XbaI for *in vitro* translation assays) (New England Biolabs) for run-off transcription. Similarly, the pAMB CAT plasmid was linearized with FspI (New England Biolabs). Transcription reactions were performed in the presence of 120 mM HEPES-KOH (pH 7.5 at 25°C), 25 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, 30 mM NTPs, 0.01% (w/v) Triton X-100, 200 nM linearized plasmid, 0.01 U/µL pyrophosphatase and 0.07 mg/mL T7 RNA polymerase in a total volume of 1 mL. Transcription reactions were incubated at 37 °C for 4 h. Enzyme was removed by phenol/chloroform extraction and the resulting solution was spun in an Amicon 100 MWCO spin column to reduce the volume to ~100 µL. mRNA was purified by denaturing, 7 M urea, polyacrylamide gel electrophoresis (PAGE), detected using brief 254 nm UV radiation and gel eluted overnight. mRNAs were ethanol precipitated and resuspended in TE buffer at pH 7.0. The sequences of the *Tte* mRNAs used in this study are listed in **Appendix A.4**.

2.5.3 Equilibrium SiM-KARTS.

2 nM *Tte* mRNA, TYE563-LNA, biotin capture strand, and Cy3-blocking strand (when present, see **Figure 2-10** and **Results 2.3.11**) were heat annealed at 70 °C for 2 min in the presence of SiM-KARTS buffer containing 50 mM Tris-HCl (pH 7.5 at 25°C), 0.6 M NaCl and 20 mM $MgCl_2$, and allowed to cool to room temperature over 20 min in the presence or absence of $preQ_1$. Next, the RNA mix was diluted to 40 pM in the same buffer in the presence or absence of

preQ$_1$, with an additional 12.5-fold excess of TYE563-LNA, biotin capture strand, and Cy3-blocking strand (when present) to ensure the complex would stay intact during dilution. All sequences of mRNA, capture strand, TYE563-LNA, Cy3-blocking strand and Cy5 anti-SD probe can be found in **Appendix A.4**. The diluted complex was chilled on ice. The chilled solution was flowed over an assembled microfluidic channel on a quartz slide coated with biotinylated-BSA and streptavidin, as previously described[118,119]. 100 μL of the chilled, 40 pM RNA complex was flowed over the slide and allowed to equilibrate for 5 min. Excess RNA was washed off the slide with SiM-KARTS buffer with or without preQ$_1$. An oxygen scavenging system consisting of 5 mM protocatechuic acid and 50 nM protocatechuate-3,4-dioxygenase with or without preQ$_1$, to slow photobleaching, and 2 mM Trolox, to reduce photoblinking[120], as well as 50 nM Cy5-probe was flowed over the slide and allowed to equilibrate for 5 min. Both Cy5 and TYE563 dyes were directly and simultaneously excited using 638 nm red and 532 nm green diode lasers, respectively. Emission from both fluorophores was simultaneously recorded using an intensified CCD camera (I-Pentamax, Princeton Instruments) at 100 ms time resolution using the Micro-Manager software. Fluorescence time traces were extracted from the raw movie files using IDL (Research Systems) and analyzed using Matlab (The MathWorks) scripts. Genuine traces exhibiting binding were manually selected using the following criteria: a single photobleaching step of the TYE563 signal to localize the mRNA molecule on the slide surface, TYE563 fluorescence intensity of >200 intensity units, and at least two Cy5 co-localization signals per trajectory corresponding to anti-SD binding events with a signal to noise ratio of at least 3:1. Suitable traces were compiled. Hidden Markov Modeling analysis was performed on the Cy5 intensity using the segmental k-means algorithm in the QuB software suite as described[121]. A two-state model was used with an unbound and bound state to idealize the data (for an additional

discussion of the idealization procedure, see **Materials and Methods 2.5.4** below). Transition density plots were constructed to extract the dwell times in the bound and unbound states, as described[122]. The normalized cumulative distributions of bound dwell times were fit with a double-exponential and unbound dwell times were fit with a single-exponential association function (see **Results 2.3.7**) in OriginLab 8.5 from which on- and off-rates were calculated. Rate constants for the anti-SD probe as a function of $preQ_1$ concentration were fit with a dose-response curve for an inhibitor with a standard Hill slope of -1. Linear regression of the data for the control probe in OriginLab showed the slope to not be significantly different from zero; these data were thus fit with regression lines of zero slope.

2.5.4 Minimal preprocessing of trace data and idealization in QuB

Prior to idealization using a two-state Hidden Markov Model (HMM), single molecule fluorescence time traces were preprocessed using a custom Matlab script to provide a rough normalization of the Cy5 fluorescence intensity across all molecules in the dataset. This is necessary because an overly large range of intensities that represent the bound state for different molecules will make it difficult to assign characteristic intensity values for the bound and unbound states in HMM analysis. In the present study, a wide range of Cy5 fluorescence intensities were observed for the probe in the bound state. Modest variability in the Cy5 intensities observed for molecules in the same experiment results from uneven laser illumination and excitation across the field of view during acquisition, and variability across experiments is due to the inherently arbitrary units of a fluorescence intensity. During preprocessing, the mean (m) and standard deviation (σ) of the Cy5 fluorescence intensity was determined across all frames in a given trace. Then, the Cy5 intensity for each frame was divided by the quantity $m + n\sigma$, where the multiplying factor n is determined empirically for each experimental system. For

the experiments presented in the current study, $n = 4$. In principle, other more sophisticated methods of normalization that result in comparable intensity values for the bound and unbound state between different molecules can be used, provided that assumptions inherent in HMM analysis are not violated (e.g., a Gaussian noise profile)[121]. Additionally, in cases where data were pooled from experiments in which the observation time (i.e., movie length) was significantly different, the traces were truncated so that all molecules in the dataset were analyzed over the same observation time. Because of the behavioral heterogeneity common to single molecule studies, it is important to ensure in this way that molecules contribute equally to subsequent analyses.

After preprocessing, normalized trace data were compiled into a single, segmented file and loaded into the QuB suite (v2.0.0.22, University at Buffalo). The camera integration time (100 ms in this study) was used for the sampling time. Each trace was treated as an individual segment. A two-state model was constructed with approximate estimates of the forward (unbound → bound) and reverse (bound → unbound) rates; 0.1 and 2.5, respectively, were used as starting rate estimates in the current study. After constructing the model, the mean amplitude and standard deviation of each state were estimated over all segments in the file using the Amps function. Because the signal-to-noise ratio varies between different molecules, the standard deviation for the unbound state was fixed at a relatively high value (between 0.22 and 0.3 in the present study) to avoid initial over-fitting of the bound state. All of the segments in the file were then simultaneously idealized using a fixed standard deviation for the unbound state by the segmental k-means (SKM) algorithm.

2.5.5 Ligand-jump SiM-KARTS

Fluorescently (TYE563) labeled preQ$_1$ riboswitches were immobilized as detailed above and first imaged in SiM-KARTS buffer without ligand. Initial co-localization of the TYE563 and Cy5 signals provided for unambiguous determination of the relative locations of single-mRNA molecules on the slide, even after TYE563 photobleaching. In some cases, the TYE563 signal persisted throughout both dark periods, further confirming that the same molecule was tracked throughout the duration of the experiment. Binding of anti-SD labeled probes (Cy5) at these locations was continuously monitored to determine the accessibility of the SD sequence. SD accessibility was monitored for 150 seconds and then a new solution of anti-SD probe in SiM-KARTS buffer was introduced in conjunction with preQ$_1$ at saturating concentration (16 μM). The process was repeated but with a final SiM-KARTS buffer solution without ligand. Because the fluorescence measured is of the anti-SD probe, which is in great excess and only excited briefly while near the surface due to the TIRFM illumination conditions, we can observe binding events to the same mRNA molecule throughout the change in ligand concentrations with limited risk of photobleaching the rapidly dissociating anti-SD probe molecules.

2.5.6 Ribosome preparation

Salt-washed ribosomes and separated ribosomal subunits were prepared using a previously described protocol with several modifications[123]. Briefly, *E. coli* MRE600 was grown in LB at 37 °C to an OD_{600} of 0.6-0.8, cooled at 4 °C for 45 min, and then pelleted. All subsequent steps were performed on ice or at 4 °C. The cell pellet was resuspended in buffer A (20 mM Tris-HCl [pH 7.05 at 25°C], 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, and 6 mM β-mercaptoethanol), and the cells were lysed in a single pass using a M-110L Microfluidizer processor (Microfluidics). The lysate was cleared by centrifugation at 30,000 × *g*, and the clarified lysate was pelleted over a 35 mL sucrose cushion (1.1 M sucrose, 20 mM Tris-HCl [pH

7.05 at 25°C], 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA) in a Beckman Ti-45 rotor for at least 16 h at 32,000 rpm. The pellet was washed with 1 mL of buffer B (20 mM Tris-HCl [pH 7.05 at 25°C], 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA), resuspended in 10 mL of buffer B by gentle stirring, and brought to a final volume of 35 mL in buffer B. This material was then pelleted again over a 35 mL sucrose cushion. The resulting pellet was washed with 1 mL of storage buffer (50 mM Tris-HCl [pH 7.5 at 25°C], 70 mM NH_4Cl, 30 mM KCl, 7 mM $MgCl_2$, 6 mM β-mercaptoethanol), resuspended in 2.5 mL storage buffer by gentle stirring, and the dialyzed against 3 changes of buffer C (50 mM Tris-HCl [pH 7.05 at 25°C], 150 mM NH_4Cl, 1 mM $MgCl_2$, and 6 mM β-mercaptoethanol) to dissociate the subunits. A portion of the dialyzed sample was adjusted to a final Mg^{2+} concentration of 7 mM, flash frozen with liquid nitrogen in aliquots (salt-washed ribosomes), and stored at -80 °C. For the remaining dialyzed sample, ~100 A_{260} units of material was loaded onto each of six 10-40% sucrose gradients in buffer C, and separated by zonal centrifugation in a Beckman SW-28 Ti rotor for 18 h at 20,000 rpm. Gradient fractions containing 30S or 50S ribosomal subunit peaks were pooled separately and pelleted in a Beckman Ti-70 rotor for 12 h at 61,500 rpm. Pelleted subunits were resuspended in storage buffer and flash frozen with liquid nitrogen in aliquots (separated subunits), and stored at -80 °C.

2.5.7 In vitro *translation assays*

Salt-washed ribosomes and separated subunits were found to perform similarly (**Figure A.1-2**) and so salt-washed ribosomes were used for all *in vitro* translation assays unless otherwise noted. *In vitro* transcribed mRNAs were translated using the PURExpress® Δ Ribosome Kit (New England Biolabs). For each reaction, 3 µL of a 4 µM mRNA solution (4 µM CAT, 4 µM *Tte* mRNA, or a mixture of 0.8 µM CAT and 3.2 µM *Tte* mRNA) was re-folded in the presence of 0, 16, or 100 µM $preQ_1$ by heating to 70 °C for 2 min, followed by slow cooling to room

temperature for 20 min, and then placed on ice. The remaining components required for translation were master-mixed and aliquoted to each reaction (1.5 – 9 μCi L-[^{35}S]-Cysteine, 4 μL PURExpress Solution A, 1.2 μL Factor Mix, and 6 pmol salt-washed ribosomes or separated 30S and 50S subunits), along with additional preQ$_1$ required to maintain a final concentration of 0, 16, 100 μM preQ$_1$ in the final reaction volume of 12 μL. Reactions were incubated at 37 °C for 2 h, frozen on dry ice, and stored at -20 °C. The following day, reactions were thawed at room temperature and 2 μL of 1 M KOH was added to quench the reaction and cleave any remaining peptide from their tRNA. Protein products were precipitated by adding 5 volumes of cold acetone and pelleted by centrifugation at 14,000 × *g* for 10 min. Pellets were resuspended in 20 μL of 1X loading buffer (45 mM Tris-HCl [pH 8.45 at 25°C], 10% (v/v) glycerol, 50 mM DTT, 1% (w/v) SDS, 0.01% (w/v) bromophenol blue) and heated at 37°C for 45 min. Protein products were resolved on 16% Tris-tricine SDS-PAGE gels[124] electrophoresed at 150 V for 2.5 h. Gels were the fixed for 45 min in 5% (v/v) glycerol, 40% (v/v) methanol, and 10% (v/v) acetic acid and dried onto 3-mm Whatman paper using a Bio-Rad Model 583 Gel Dryer (**Figure A.1-4**). Dried gels were imaged using a storage phosphor screen and Typhoon 9410 Variable Mode Imager (GE Healthcare Life Sciences). Gel images were quantified using ImageQuant v 5.2 (Molecular Dynamics). Unless otherwise noted, after background correction the intensities for CAT, TTE1564, and TTE1563 bands were divided by their respective number of cysteine residues (5, 1, and 1, respectively). For each lane, the intensities for TTE1564 and TTE1563 bands were summed and then divided by the intensity of the respective CAT band. Values at each preQ$_1$ concentration were graphed in Prism (GraphPad Software) and an unpaired, two-tailed t-test was used to assess statistical significance.

2.5.8 Burst analysis

Burst analysis was carried out using the Rank Surprise (RS) method as described by Gourévitch and co-workers[101]. We utilized a modified Matlab implementation of the RS method based on the Matlab script provided in the supplement to Gourévitch *et al*.[101]. In the standard implementation each molecule's interspike intervals (ISIs) are ranked independently. In our implementation, termed global burst analysis, we have extended this so that each ISI detected in all our experimental conditions is ranked simultaneously; therefore, a burst is defined as a global property of the molecules and is not biased by the number of binding events of an individual molecule. Briefly, ISIs were determined by calculating the time in between individual binding events for each molecule. ISIs for all molecules were collected and used as input for global burst analysis, using the RS method to demarcate the start and end points of bursts, i.e., the first and last binding event, respectively, in a sequential series of binding events occurring in quick succession ("burst"). Each start and end point was then reassigned to the corresponding molecule, preserving the single molecule burst profile. Global burst analysis was carried out separately for equilibrium and ligand-jump datasets. The Rank Surprise (RS) algorithm developed by Gourévitch *et al*.[101] requires two parameters, the maximal ISI between spikes to be considered part of a burst and a Rank Surprise cutoff (alpha). These were set to 40 seconds and 3, respectively. Although the maximal ISI in a burst was set to 40 seconds, the distribution of ISIs we obtained after analysis is much lower than this suggesting that we have provided enough flexibility in the algorithm to find the true distribution of ISIs in burst.

For the ligand-jump SiM-KARTS experiments, we further characterized the time evolution of SD accessibility throughout changes in ligand concentration by examining changes in the burst density for each molecule. We ranked the three segments (*Minus*, *Plus*, *Minus'*) for each molecule in the ligand-jump SiM-KARTS experiment by their burst density. For a given

molecule, the segment with the highest density of bursts was ranked as High (H), the next highest as Mid (M), and lowest as Low (L). We then quantified the overall distribution of burst density rankings for the three segments. We plotted each individual molecule's burst behavior as a rastergram and organized them such that their per-segment ranks were the same within a group. MATLAB scripts for preprocessing of single molecule trace data and global burst analysis are provided in **Appendix A.6**.

2.5.9 Fano factor calculations

Matlab scripts were written to calculate and simulate the Fano factor from our experimental data and from a simulated Poisson distribution, respectively (see **Appendix A.6**). The Fano factor is defined as the variance in spike counts divided by the mean spike count for a given time interval, *w* (**Eq. 3**). For every molecule analyzed in a particular condition, a time interval of length *w* was randomly selected from the molecule's fluorescence time trace. The Fano factor was calculated for *w* equal to 5, 10, 20 and 40 seconds. Each time window *w* was sampled 100 times with a different random seed for each molecule to generate an average Fano factor. The average Fano factor and the standard deviation for each time window are presented in **Figure 2-8**. For the simulations, the Matlab Poisson random number generator was utilized to generate spike counts with an average firing rate equal to the average ISI in the burst from our experimental data and equal number of samplings. 95% confidence intervals were calculated in Matlab utilizing the expression:

`gaminv([.025,.975],(n-1)/2,2/(n-1))`, where `n` is the sample size[104].

2.5.10 Measurement of anti-SD probe diffusion coefficient

Fluorescence correlation spectroscopy (FCS) was used to determine the diffusion coefficient of the Cy5 anti-SD probe. Anti-SD probe was diluted to 2.5 nM in SiM-KARTS buffer and measurements were performed at room temperature on an Olympus IX81 inverted microscope with an ISS ALBA 5 confocal system.

2.5.11 In silico *prediction of anti-SD probe binding sites and free energies of hybridization*

The potential binding sites on the *Tte* mRNA for the anti-SD probe were predicted using the RNA Fold Bimolecular algorithm in the RNAstructure program (v5.6 and v5.7, Mathews lab, University of Rochester Medical Center http://rna.urmc.rochester.edu/RNAstructure.html). First, the SiM-KARTS anti-SD probe and the full length sequence of the *Tte* mRNA used for SiM-KARTS (see **Appendix A.4**) were each saved as their own sequence files. Nucleotides in the *Tte* mRNA that would be bound by the capture oligonucleotide at the 5′ end (including the additional guanosines added during *in vitro* transcription) as well as the nucleotides bound by the TYE563-LNA were prevented from participating in other base-pairing interactions with other regions of the *Tte* mRNA or the anti-SD probe during the folding predictions by forcing these nucleotides to remain single-stranded (using lower case lettering). The entire riboswitch aptamer domain, or the aptamer domain with the exception of the terminal two nucleotides of P2, was also forced single-stranded to approximate the ligand bound and ligand-free forms, respectively, of the riboswitch. Without these constraints on the riboswitch aptamer, the lowest free energy structure generated has an intramolecular base-pairings that are mutually exclusive with proper aptamer formation; this is likely due to the program's inability to correctly predict the pseudoknot fold of the riboswitch that also contains several non-canonical base pairs (**Figure 2-1a**). The anti-SD probe and *Tte* mRNA (with the modifications described above) were folded using the RNA Fold Bimolecular algorithm, which considers the free energy of competing intramolecular structures

when generating a set of lowest free energy structures. The algorithm was run using a 50% maximum energy difference, a limit of 20 structures, and window size of 15. The maximum energy difference and window size control how different the generated structures should be from each other to be considered during folding. Using a large (i.e., 50%) maximum percent energy difference allows suboptimal structures to be generated that are up to 50% less stable than the lowest free energy structure, an important consideration when trying to achieve sufficient sampling of alternative binding sties in structure space when the intramolecular interactions in the *Tte* mRNA will be the dominant contribution to the total free energy for each folding. A large window size requires that the predicted suboptimal folds must be relatively more different from one another. In most cases, fewer than 20 structures were found and decreasing the window size from 15 to 5 generated a largely identical set of structures, indicating that these parameters likely ensure adequate sampling.

The putative binding site at the riboswitch's expression platform (TTE1564 SD sequence) was highly represented among the suboptimal structures generated. In addition, four other predicted binding sites were represented among the generated structures, with two of the four differing by a single base pair.

The resulting connectivity table files (`*.ct`) for each structure were modified manually so that only the intermolecular base pairs made between the *Tte* mRNA and anti-SD probe remained. These modified `*.ct` files were then used to predict the free energies for the isolated anti-SD probe•*Tte* mRNA hybridization using the Efn2 module in RNAstructure at 37 °C (310.15 K). These free energy values were identical to the values predicted between the anti-SD probe and a truncated portion of the *Tte* mRNA, so long as the truncated sequence still contained some unpaired bases at both ends, indicating that effects of terminal unpaired nucleotides, which

better approximates a short probe binding internally to a longer mRNA, are taken into consideration. The same procedure using manually constructed `*.ct` files and the Efn2 module was also used to calculate the free energies for the *E. coli* 16S rRNA and TTE1564 SD shown in **Figure 2-2a**.

2.6 Acknowledgments

The authors thank J. Widom and K. Suddala for careful reading of the manuscript, and G. Garcia for the generous gift of preQ$_1$. The authors also wish to thank the Single Molecule Analysis in Real-Time (SMART) Center of the University of Michigan, seeded by NSF MRI-R2-ID award DBI-0959823 to N.G.W., as well as J. D. Hoff for performing the FCS measurements. This research was funded by NIH grant GM062357 and a sub-contract on GM063162 to N.G.W. (PI: J. E. Wedekind).

CHAPTER 3: S1 unties the pseudoknot: Mechanistic insights into S1-mediated unfolding of RNA tertiary structure[2]

3.1 Overview

In *E. coli*, ribosomal protein S1 is a large protein that loosely associates with the small ribosomal subunit and has a demonstrated role in mediating mRNA binding by the ribosome, as well as passive unwinding of RNA secondary structures that impede translation. Despite its name, this essential protein also plays a variety of other roles in the cell that are related to its ability to bind RNA, including transcriptional cycling and serving as a subunit of Qβ replicase. While certain RNA sequence determinants of S1 affinity and details of the interactions of S1 with simple secondary structures are known, the mechanistic details of the interactions of this protein with more complicated RNA secondary and tertiary structure are still murky. Using well-characterized H-type pseudoknots found in class-I preQ$_1$ riboswitches as a model of highly structured RNA with tunable stability, we characterized S1 binding at the single-molecule level. We observed that S1 is able to bind and at least partially unfold this pseudoknot, and found an apparent upper limit to the degree of structural stability in an RNA fold that S1 is able to act upon.

3.2 Introduction

[2] Single molecule FRET experiments presented in this chapter were performed in conjuction with May Daher. The original ASKA clones used to prepare S1 were gereously provided by Prof. Janine R. Maddock.

3.2.1 Biological roles of S1: translation and beyond

Ribosomal protein S1 has a well-established role in the translation of mRNA by the *Escherichia coli* ribosome in which S1 mediates the binding of many mRNAs by the 30S subunit, particularly those with weak SD sequences[23,25,26,33,37,125], and is required for cell growth and viability[23,30]. Despite its name, S1 is known to have other cellular activities[44] beyond its mRNA binding activity on the 30S subunit, including roles in trans-translation[45], transcriptional cycling[50], stimulation of T4 endoribonuclease RegB, and activity as a subunit of Qβ replicase[46-49]. In each of these capacities, the RNA binding activity of S1 is essential for carrying out the respective function.

3.2.2 Structure of S1 and its binding footprint

S1 is a large protein composed of six imperfect repeats an RNA binding domain known as the OB-fold[126]. The first two N-terminal domain repeats are involved in binding of the ribosome[127] and are thus indispensable for cell survival[30,128], and are also implicated in binding of Qβ replicase[49] and tmRNA[129]. The central three domain repeats are involved in binding mRNA, with the strongest evidence supporting the involvement of domains 3 and 4[28,40,128]. Lastly, domains 4 and 5 were found to be most important for the ability of S1 to stimulate transcription[50].

S1 can unwind dsRNA[130,131] and while it was once thought that the unwinding activity of S1 was not needed for its role in translation[28], the body of recent work to the contrary continues to grow[30,37,132]. For instance, the TIR of the *rpsO* mRNA contains a pseudoknot, the unfolding of which is promoted by S1 in isolation[30]. Furthermore, even when the strength of the SD is enhanced in *rpsO* mRNA, the unfolding (but not binding) of that mRNA, which is required for subsequent formation of the 30S IC, is strongly impaired when the 30S subunit has been depleted of S1[30].

Despite the importance of S1, studies of the mechanism employed by S1 in binding and unfolding RNA are still relatively few. One recent study by Qu *et al.*[40] employed optical tweezers to study the ability of S1 to unwind a long dsRNA helix. Illustrating the power of single molecule techniques, they determined that S1 passively unwinds dsRNA by binding to the terminal base pairs of the helix as they become transiently single-stranded due to thermal breathing, and in doing so prevents reannealing. This results in progressive unwinding in multiple small steps, from which they were also able to determine that a single S1 protein likely binds ~10 nucleotides. This is in good agreement with previous reports[131], including the cryo-EM reconstruction of S1 bound to the 30Swhere S1 was observed to interact with a stretch of 11 mRNA nucleotides[32].

3.2.3 Use of the preQ$_1$ riboswitch pseudoknot as a model of RNA tertiary structure

The work of Qu *et al.*[40] provides a detailed look at the interaction of S1 with secondary structure in RNA using a long model RNA helix, but there has yet to be an investigation into how S1 interacts with RNA possessing well-defined tertiary structure, such as the pseudoknot in the TIR of *rpsO*.

Pseudoknots are a common type of tertiary structural motif and are found in a wide variety of functionally diverse RNAs[66]. The class-I preQ$_1$ riboswitch, which has undergone extensive structural characterization by our lab and others[63,64,83,85], contains an aptamer domain that is comprised of a small, well-defined H-type pseudoknot. By virtue of being a riboswitch, a number of ligands and their corresponding affinities are also known including preQ$_1$, guanine, 7-deazaguanine, 2-aminopurine, and preQ$_0$ to name a few[59]. These ligands provide a convenient handle through which the stability of the pseudoknot can be easily and reversibly modulated while maintaining the same global fold, without the need for making sequence-level changes that

may have unpredictable effects on the structure, or fundamentally change the nature of the interaction under study. For example, an RNA helix can be made more stable simply by changing the GC content, however that could become problematic when studying phenomena that exhibit sequence dependence, as is the case here given the well-known preference of S1 for A/U-rich sequences. Tunable stability and a well-defined tertiary structure are qualities that make the pseudoknot from the preQ$_1$ riboswitch particularly well suited for use as a model RNA with which to elucidate the effects of structural stability on the capacity for S1 to interact with RNA possessing tertiary structure.

3.3 Results

3.3.1 Structural features of preQ$_1$ riboswitch-based pseudoknot variants can be distinguished by analysis of their melting curves

The pseudoknot present in the preQ$_1$ riboswitch has extreme sequence economy, requiring only 33 nts to form its H-type pseudoknot, and multiple crystal structures are available[63,64]. As such it is relatively straight forward to design mutations that will alter the stability of the pseudoknot fold in predictable ways. To this end, we generated a set of pseudoknot variants to investigate which features, if any, of the pseudoknot would affect the binding affinity of S1 (**Figure 3-1**). Two variants, C15U and C15A were designed to diminish the ability of the pseudoknot to be stabilized by preQ$_1$ through alteration of the basepairing interactions that the ligand has with C15, but without changing the global fold of the pseudoknot. A mutation equivalent to the C15U variant for the related preQ$_1$ riboswitch in *Bacillus subtilis* (*Bsu*, C34U), and was found to have no measurable response to preQ$_1$ in that context; it did however show some affinity towards 2,6-diaminopurine[59].

Part of the stabilization brought about by preQ$_1$ derives from a set of continuous base stacking interactions: when bound, the ligand stacks on top of the G5-C16 base pair at the top of the P1 stem[63]. The CG mutation is intended to modify the "floor" of the binding pocket upon which the ligand stacks, without altering the CG content of either stem. The AU mutant, by contrast, both weakens the P1 and P2 stems, as well as disrupts potential stacking interactions. The final two variants, termed UUCG and GAAA, abolish the pseudoknot fold by replacing a large portion of the sequence with the well characterized UNCG and GNRA tetraloops[133,134]. These variants, as well as the wild-type (WT) sequence, were prepared by *in vitro* transcription from mutagenized *Tte* plasmids as described in the **Materials and Methods 3.5.2** and **3.5.3**. The transcribed products were a total of 76 or 70 nt in length, with their 3′ ends all terminating with the +6 nt of the adjacent open reading frame; these pseudoknot constructs are therefore referred to as the +6 series. Additionally, several shorter constructs that contained only the minimal sequence required for pseudoknot formation were also prepared by *in vitro* transcription, including two variants *Tte*^{rigid} and *Bsu*^{floppy}, that have been previously studied by smFRET[83], along with the wild-type sequence *Tte*^{min} (**Figure 3-1b**). These constructs have mutations in the L3 tail that result in a more (*Tte*^{rigid}) or less (*Bsu*^{floppy}) compact pre-folded state than the respective wild-type pseudoknot.

In order to assess the relative stabilities of the pseudoknot variants in each series, melting curve analyses were performed in minimal buffer (10 mM sodium phosphate), or buffer containing 1 mM Mg^{2+} and 100 mM NH_4Cl. A very large hysteresis was observed between the heating and cooling temperature ramps (data not shown), indicating that unfolding and refolding of this pseudoknot likely proceed via different pathways, or alternatively that refolding is not being measured at equilibrium with the moderately slow ramp rates used (0.5 °C/min)[135].

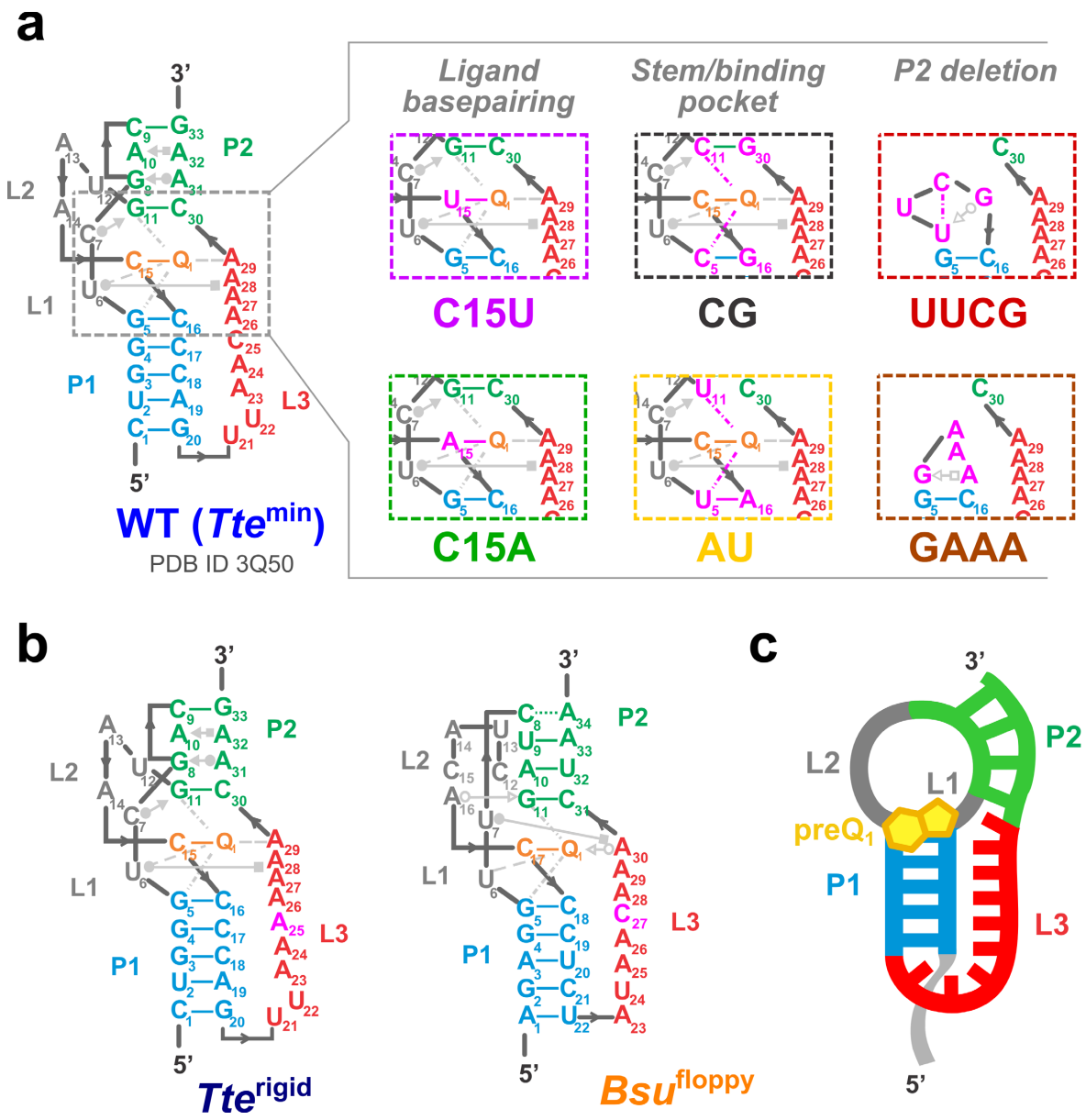


Figure 3-1 Pseudoknot variant constructs.

(**a**) Variants of the pseudoknot present in the *Tte* preQ$_1$ riboswitch were designed to modulate the stability of the pseudoknot through mutations that affect the capacity to directly basepair with the ligand, disrupt the stacking interactions in the binding pocket, or completely abolish the pseudoknot structure through deletion of P2. Structural features of the pseudoknot are colored as in **Figure 2-1a,** and a subset of the key interactions with preQ$_1$ are show with Leontis-Westhof[86] notations. Sequence mutations and altered interactions are shown in magenta. Other nucleotides that are present in the sequence of the final construct but not required to form the pseudoknot fold are omitted for clarity. Full sequence information for all constructs used in this study appears in **Appendix Table B.2-3**. (**b**) Pseudoknot variants based on the preQ$_1$ riboswitches from *Tte* and *Bsu* that result in a more (*Tte*^{rigid}) or less (*Bsu*^{floppy}) compact pre-folded state than the respective wild-type pseudoknot[83]. (**c**) Simplified depiction of the pseudoknot fold.

Therefore, only the unfolding curves were considered further. Illustrative example melting curves are shown in **Figure 3-2**. Several of the pseudoknot variants showed two clear transitions in the absence of ligand, the first occurring ≤ 45 °C for the minimal constructs and ≤ 60 °C for the +6 series. It appears that the additional sequence 5′ and 3′ of the pseudoknot in the +6 series constructs partially stabilizes this first transition. It should be noted that in some instances, qualitatively different melting profiles were obtained for sequential melting curves on the same sample (**Figure B.4-1**), indicating differing degrees of fold heterogeneity. This underscores the need for careful attention to consistent and thorough refolding. Melting curves from variants lacking P2 (**Figure 3-2b**, UUCG) show only a single transition > 70 °C, as is expected for the melting of highly stable tetraloop structures[133]. This allows the first, lower temperature transition to be confidently assigned to melting of the P2 stem, and the second higher temperature transition to melting of P1. Additionally, the C15A and C15U mutants in the +6 series showed two distinct transitions, similar to WT, indicating that they possess a similar global fold as intended. However, these ligand binding mutants exhibit a different response to ligand binding from the WT or minimal *Tte* variants. Whereas the WT and *Tte* variants show an increase in the transition temperature for P2 melting to the point of overlapping with the transition for P1, which suggests a more uniform and cooperative unfolding in the presence of ligand, the C15A an C15U variants show a more well defined transition for P2 melting, indicating that ligand is able to organize but not significantly stabilize P2 (**Figure 3-2b**). Melting transitions for all pseudoknot constructs are summarized in **Table 3-1** and **Table 3-2**. As expected, the majority of the +6 series mutants were not stabilized in the presence of ligand, and in several, no clear transition attributable to P2 was observed. For those that do bind ligand, only the transition temperature of P2 was affected.

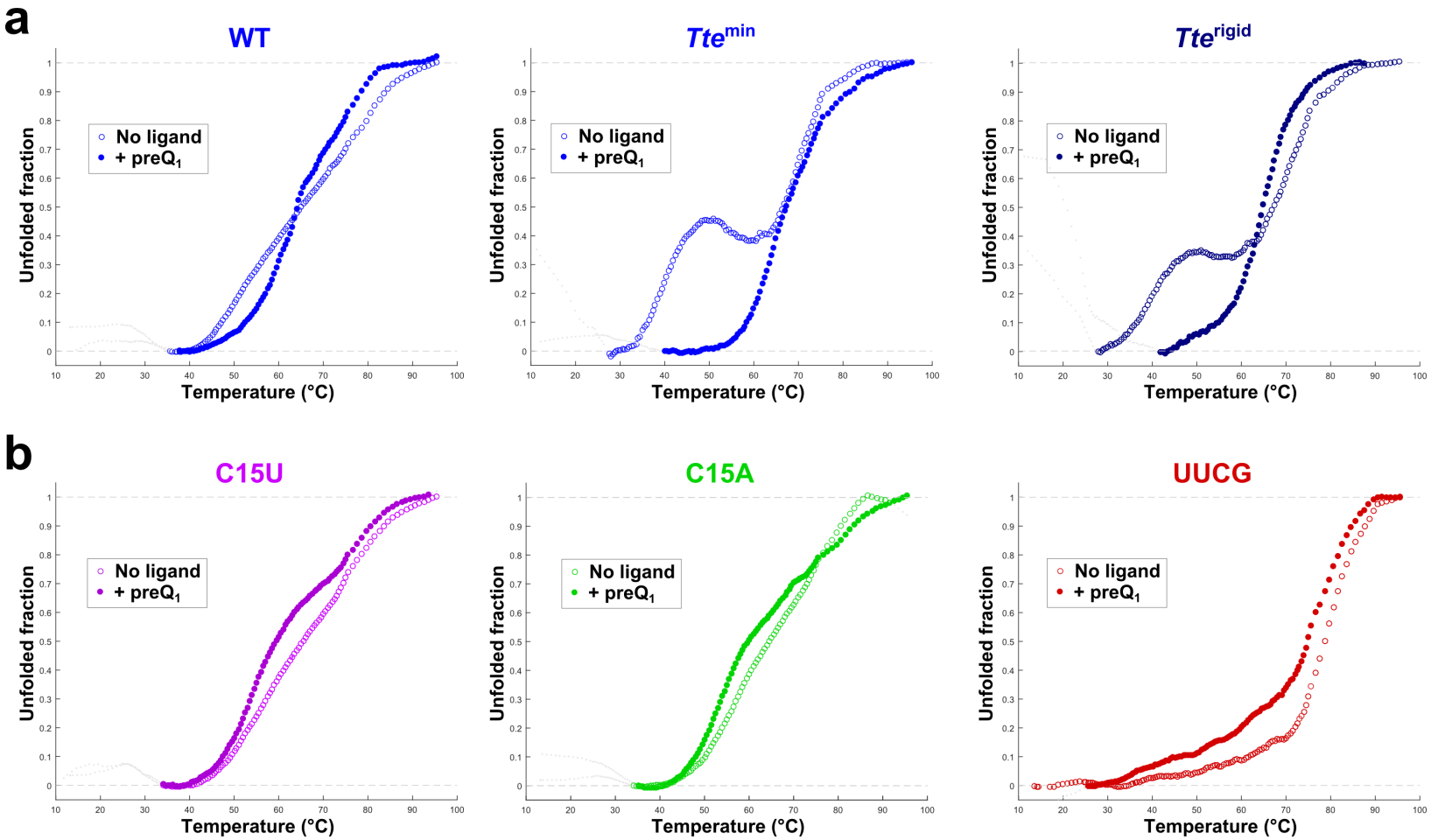


Figure 3-2 Characterization of pseudoknot variants through UV melting curves.

Illustrative melting curve data taken in the presence of 1 mM $MgCl_2$ and 100 mM NH_4Cl, after sloping baseline correction and normalization. (**a**) Melting curves of the WT pseudoknot (+6 series) and *Tte* variants (minimal series) show a clear increase in the T_m of the first transition in the presence of ligand. Note that the decrease in unfolded fraction between the first and second transition *Tte*^{min} is an artifact due to uncertainty in baseline correction and smoothing of the data. (**b**) Pseudoknot variants with the potential to bind ligand show a sharpening of the first transition in the presence of $preQ_1$. Comparison with P2 deletion mutants (e.g., UUCG) allows for the first transition to be more confidently assigned to melting of P2.

Table 3-1 Melting temperatures for the +6 series of pseudoknot constructs.

Melting transition temperatures were determined in 10 mM sodium phosphate buffer (pH 7.0 at 22 °C) in the absence and presence of ligand, with or without added mono- and divalent ions. P2 transition temperatures for variants that showed clear response to ligand are in bold face. All temperatures are in °C. Numbers in parenthesis indicate the standard error of the mean of at least two measurements (--, no apparent transition; *ND,* not determinable; *merged,* P1 and P2 melting transitions are overlapping).

Pseudoknot variant	**No ligand** Mg^{2+}, NH_4Cl	**+ preQ$_1$** Mg^{2+}, NH_4Cl	**No ligand** Phosphate only	**+ preQ$_1$** Phosphate only
P2:				
WT	**51.7 (0.3)**	**64.2 (0.2)**	**33 (2)**	**56**
C15U	**55**	**55.1 (0.7)**	**43**	**60 (2)**
C15A	58	55 (2)	*ND*	*ND*
AU	*ND*	*ND*	22	23
CG	*ND*	*ND*	*ND*	40
UUCG	53 (4)	--	--	--
GAAA	48	--	--	--
P1:				
WT	76	79.2 (0.7)	57 (1)	*merged*
C15U	77	77 (2)	58.5 (0.2)	60 (2)
C15A	78	77 (1)	58	*ND*
AU	79 (2)	78.7 (0.6)	61	59
CG	78.3 (0.7)	77 (2)	58	58
UUCG	81 (2)	79.2 (0.1)	58	57
GAAA	80 (2)	79.0 (0.2)	62	59

Table 3-2 Melting temperatures for the minimal series of pseudoknot constructs.

Melting transition temperatures were determined under the same conditions as in **Table 3-1**. All temperatures are in °C. Numbers in parenthesis indicate the standard error of the mean of at least two measurements (--, no apparent transition; *ND,* not determinable; *merged,* P1 and P2 melting transitions are overlapping).

Pseudoknot variant	**No ligand** Mg^{2+}, NH_4Cl	**+ preQ$_1$** Mg^{2+}, NH_4Cl	**No ligand** Phosphate only	**+ preQ$_1$** Phosphate only
P2:				
Tte^{min}	45 (2)	64.8 (0.9)	48 (2)	57
Tte^{rigid}	42 (2)	66.3 (0.5)		
Bsu^{floppy}	--	--		
P1:				
Tte^{min}	74 (2)	*merged*	69	*merged*
Tte^{rigid}	71.5 (0.7)	*merged*		
Bsu^{floppy}	70 (4)	61 (2)		

3.3.2 Some, but not all, pseudoknot variants are capable of stably binding the preQ$_1$ ligand

While melting curve analysis identified variants that were stabilized in the presence of ligand, we endeavored to further characterize these variants using an electrophoretic mobility gel shift assay (EMSA) using RNAs labeled with 3′ terminal Cy3 fluorophore. Under the conditions of these assays, the concentration of RNA relative to ligand in the reaction is such that the resulting data provides a titration curve (**Figure 3-3**). As expected, the WT and *Tte* variants display very high affinity for preQ$_1$, and have a sharp inflection point at approximately a 1:1 stoichiometry (**Figure B.3-1**). Somewhat unexpectedly, the C15U variant also showed clear ligand binding, evidenced by the appearance of a doublet band, although its binding of preQ$_1$ is weaker. Two other variants, AU and *Bsu*^{floppy} (**Figure 3-3b**, **c**) both showed doublet bands that were insensitive to ligand, indicating that multiple stable conformations with different degrees of compactness exist for these pseudoknots. Fitting of these titration data with a one-site binding model equation that accounts for depletion of the free ligand concentration (**Appendix B.11**) provided unsatisfactory results, likely due to the inability to account for incomplete saturation at high ligand concentration (i.e. fraction of RNA with ligand bound < 1).

The ability of S1 to bind to the pseudoknot variants was assessed using a similar EMSA. Illustrative example gels are shown in **Figure 3-4a** and **b**. For the +6 series, the formation of two S1-bound species (Complexes 1 and 2) was observed for pseudoknot variants that exhibited a global fold similar to WT, namely C15A and C15U, as indicated from melting curve analysis. Other variants in the +6 series showed no (or greatly diminished) formation of a species with similar mobility to Complex 1 seen for the WT pseudoknot. Similarly, S1 binding titrations with the shorter, minimal pseudoknot *Tte*^{min} (whose pseudoknot sequence is shared with the WT construct) showed primarily a single discernable S1-bound species with mobility similar to that

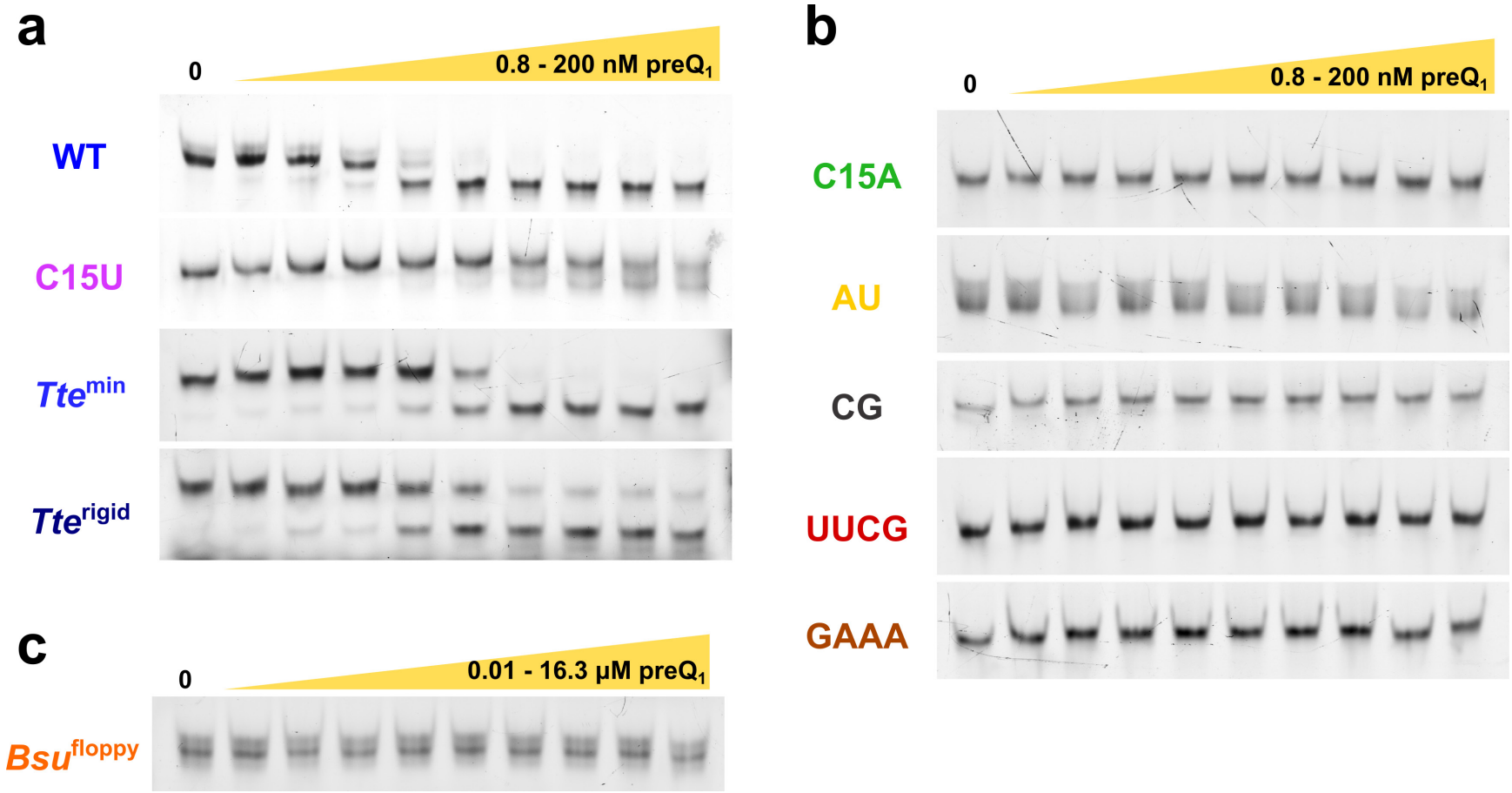


Figure 3-3 Assessment of stable preQ$_1$ binding by pseudoknot variants.

(**a**) Titration of 10 nM WT, C15U and *Tte* pseudoknot variants with preQ$_1$ brings about a clear shift in mobility towards a more compact, faster migrating species, indicating that binding of preQ$_1$ is relatively stable. Concentrations of preQ$_1$ used were 0, 0.8, 1.6, 3.1, 6.3, 12.5, 25, 50, 100, and 200 nM. (**b**) Titration with preQ$_1$ of pseudoknot variants that have greatly weakened ligand affinity, or are incapable of binding ligand. Concentrations of RNA and preQ$_1$ used were the same as in **a**. (**c**) While the *Bsu*^{floppy} variant has been shown previously to respond to ligand[83], bound preQ$_1$ is likely too labile to bring about a mobility shift. Concentrations of RNA were as in **a**, preQ$_1$ concentrations used were 0, 0.01, 0.02, 0.04, 0.08, 0.16, 0.41, 0.82, 1.63, and 16.3 μM.

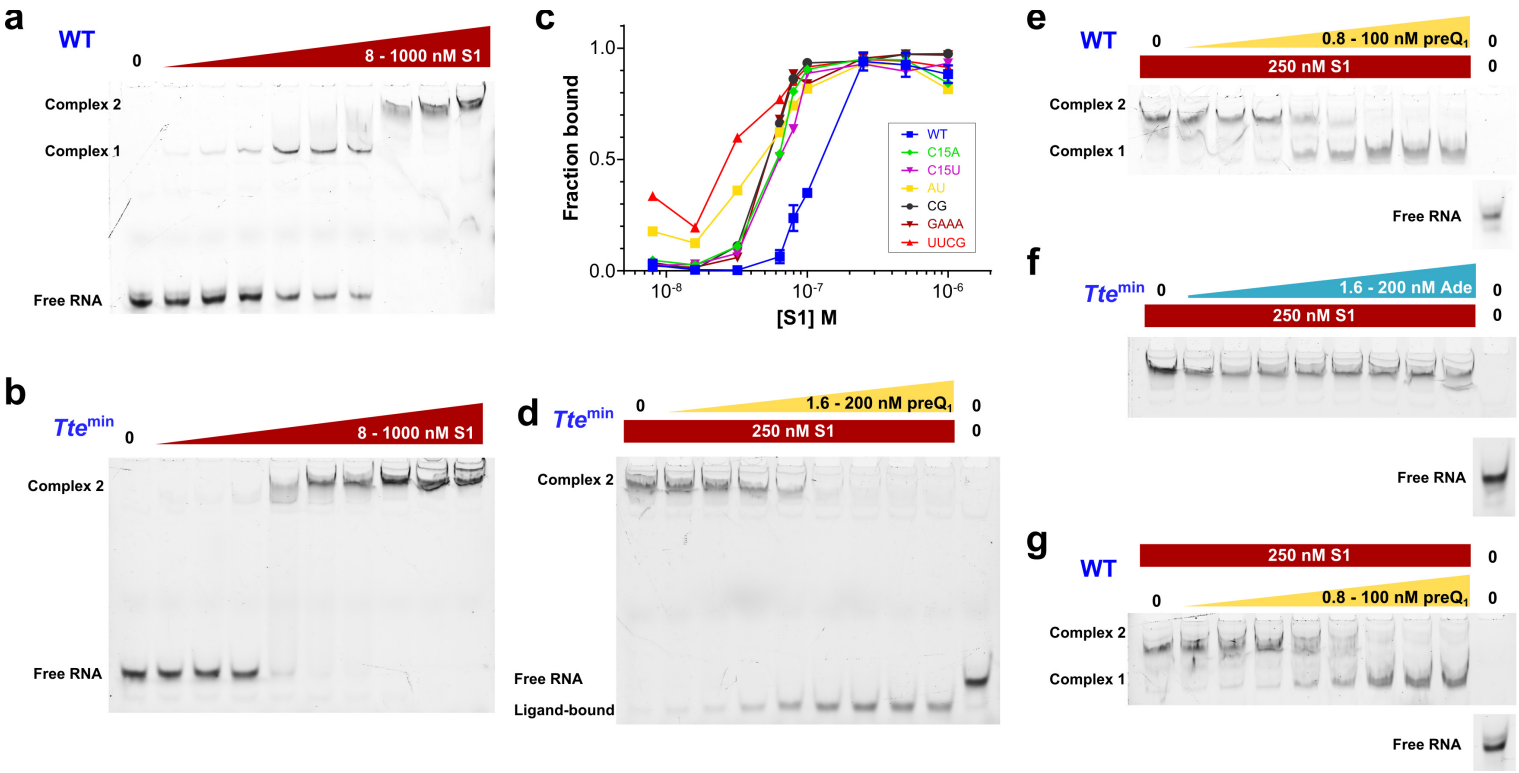


Figure 3-4 Assessment of S1 binding and its dependence on pseudoknot stability.

(**a**) Titration of WT pseudoknot with S1 shows formation of two complexes in a concentration dependent manner, suggesting two distinct binding sites. Concentrations of S1 are 0, 8, 16, 32, 64, 80, 100, 250, 500, 1000 nM. (**b**) Titration of the shorter *Tte*^{min} pseudoknot with S1 shows formation of only a single distinct complex. Concentrations of S1 are the same as in **a**. (**c**) Quantification of the fraction of RNA bound in Complex 2 for the +6 series of constructs in the absence of ligand. Binding of pseudoknot RNA by S1 to form Complex 2 occurs at lower concentrations for pseudoknot variants that have a destabilized (or no) pseudoknot structure. Error bars represent the standard deviation of 2 independent measurements. (**d**) The *Tte*^{min} RNA pseudoknot is not bound by S1 when preincubated with ligand. Concentrations of preQ$_1$ are 0, 1.6, 3.1, 6.3, 12.5, 25, 50, 100, and 200 nM. (**e**) Preincubation with ligand prevents S1 binding to form Complex 2. Concentrations of preQ$_1$ are 0, 0.8, 1.6, 3.1, 6.3, 12.5, 25, 50, and 100 nM. (**f**) Changes in S1 binding affinity are the specific to preQ$_1$ as titration with adenine (Ade) has no effect. Concentrations of adenine are the same as those used for preQ$_1$ in **d**. (**g**) Addition of preQ$_1$ after incubation with S1 shows displacement of S1 to form Complex 1. Concentrations of preQ$_1$ are the same as in **e**. RNA concentration is 10 nM in all panels.

of Complex 2 (**Figure 3-4b**). Taken together, these data suggest that two separate binding sites for S1 exist: one with higher affinity that leads to Complex 1 and is located in the additional, non-pseudoknot sequence portions of the longer +6 series constructs, and a second, lower affinity site that is shared by both series, and thus is located in the pseudoknot and gives rise to a species with further reduced mobility (Complex 2).

Quantification of the fraction of RNA present in Complex 2 (**Figure 3-4c**) shows that the apparent affinity of S1 increases as the stability of the pseudoknot fold decreases (WT ≪ C15U ≈ C15A < AU < UUCG). This observation is supported by the experiment shown in **Figure 3-4d**, in which the amount of Tte^{min} RNA present as the S1-bound complex decreases when the RNA is preincubated with $preQ_1$ before 250 nM (i.e. saturating, 25-fold excess) S1 is introduced. A similar phenomenon is observed for the WT pseudoknot construct, indicating that binding of $preQ_1$ and binding of S1 at the second site, which gives rise to Complex 2, are mutually exclusive (**Figure 3-4e**). Furthermore, this effect is specific to the binding of ligand as preincubation with adenine, a structurally similar nucleobase that the pseudoknot does not bind[59], has no effect (**Figure 3-4f**). Lastly, S1 that is already bound to pseudoknot RNA at site 2, but not site 1, can be displaced by the addition of $preQ_1$, indicating that S1 binding is at equilibrium in these assays (**Figure 3-4g**).

3.3.3 S1-bound pseudoknot RNA is at least partially unfolded

In order to better characterize the S1-bound RNA species, an S1 titration assay (**Figure 3-5a**)was performed using a doubly-fluorophore labeled pseudoknot construct (Tte^{smFRET}) whose dynamics and ligand binding properties have been extensively investigated[83]. This pseudoknot construct is labeled with a FRET pair with the acceptor fluorophore Cy5 on U12 of the L2 loop and the donor fluorophore Dy547 on G35, 2 nucleotides after the end of the sequence required to form

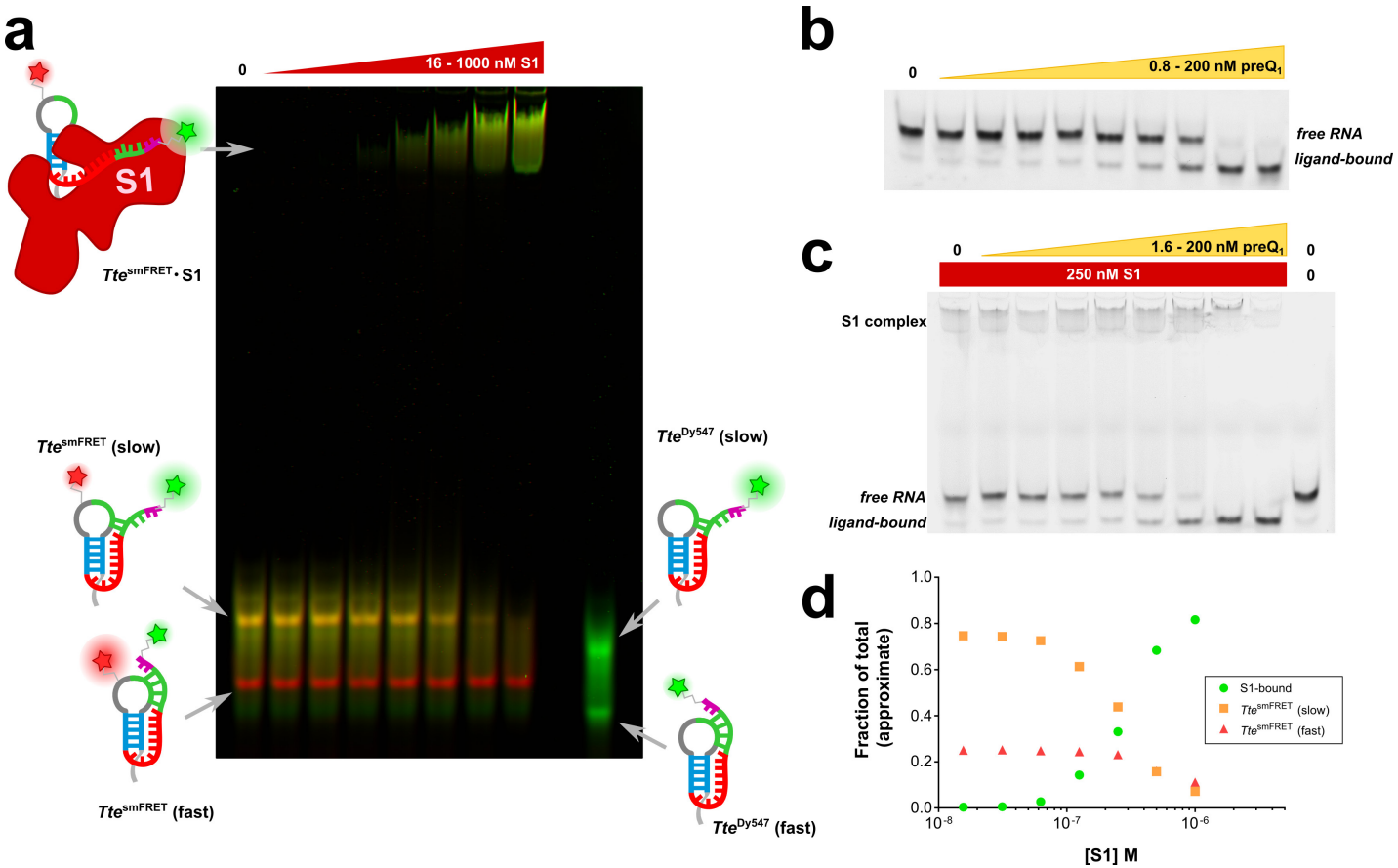


Figure 3-5 Characterizing the conformation of an S1 bound pseudoknot by in-gel FRET.

(**a**) Titration of the doubly fluorophore labeled pseudoknot used for smFRET experiments (*Tte*^{smFRET}) with S1 in the absence of preQ$_1$. Using a higher concentration of RNA (lower stoichiometry of S1:RNA) allows both slow and fast RNA bands to be clearly visualized. The S1-bound for of the pseudoknot exhibits lower FRET (green fluorescence), indicative of greater interfluorophore distance than the slow (mid-FRET, yellow) or fast (high FRET, red) free RNA forms. The singly labeled *Tte*^{Dy547} is included for reference. Pseudoknot cartoons are as in **Figure 3-1a** with 5' and 3' non-pseudoknot nucleotides in gray and purple, respectively. S1 concentrations are 0, 16, 31, 63, 125, 250, 500, and 1000 nM. (**b**) Titration with preQ$_1$ as in **Figure 3-3a** but with a 10X higher RNA concentration. (**c**) preQ$_1$ titration in the presence of S1 as in **Figure 3-4d** but with a 10X higher RNA concentration. (**d**) Approximate quantification of RNA distribution at each S1 concentration in **a**.

the pseudoknot. In the absence of S1, the *Tte*^{smFRET} construct is present in two forms: a fast moving, more compact form (high FRET, red), and a slower moving, less compact form (mid FRET, yellow). In previous studies these two conformations were also referred to as the folded and pre-folded conformations[83]. In lanes containing S1, in-gel FRET efficiency clearly shows that the RNA in complex with S1 exhibits lower FRET, indicative of greater interfluorophore distances. Although the in-gel FRET efficiencies cannot be used to calculate physical distances in the same way that is possible for single molecule data (see below), the greener color of the S1-bound complex indicates that the RNA is unfolded to a greater extent than it is in either of the free forms.

3.3.4 S1 preferentially binds the less stable, less compact form of the pseudoknot

In the S1 EMSA shown in **Figure 3-5a**, a 10-fold higher concentration of RNA was used than in **Figure 3-4**; this higher concentration of RNA allows both the fast and slow moving free forms to be easily visualized. For comparison, preQ$_1$ titration experiments in the absence and presence of 250 nM S1 were repeated as in **Figure 3-4**, but using this same 10-fold higher concentration of RNA (**Figure 3-5b**, **c**) and thus this concentration of S1 is not sufficient to completely deplete the free RNA fraction in the absence of ligand (compare to **Figure 3-4d**), allowing easier assignment of each band. Little change is observed when comparing the band intensities of the ligand-bound (faster migrating, more compact) form of the RNA in **Figure 3-5b** and **c**, whereas the free RNA form is preferentially bound into an S1 complex. This also clearly seen in the S1 titration of the *Tte*^{smFRET} construct. Quantification of the fast and slow migrating bands shows that as the concentration of added S1 increases, the more slowly migrating, less compact conformation of the pseudoknot is preferentially bound by S1, leaving the more compact and stably folded conformation.

3.3.5 Direct observation of S1-induced unfolding of pseudoknot RNA

While mobility shift assays and in-gel FRET experiments are very informative, they still lack the ability to inform on the uniformity of the S1-bound species, or provide information on binding processes which have not reached equilibrium. To more thoroughly investigate the nature of the interaction between S1 and this pseudoknot fold we designed the single molecule binding assay depicted in the **Figure 3-6b**, **c**, and **f**. Using the doubly-fluorophore labeled pseudoknot construct *Tte*^{smFRET} and a prism-TIRF microscopy setup as described previously[83], we are able to make sensitive measurements of the docking and undocking of the P2 stem nucleotides in individual, surface immobilized pseudoknot RNAs and thus monitor unfolding of this structural feature of the pseudoknot. Experiments were performed using similar concentrations and buffer conditions as were used for assays of S1 and preQ$_1$ binding. Schematic depictions of the experiments in which molecules were imaged after the stepwise addition or removal of S1 and preQ$_1$ are show in **Figure 3-6c** and **f**.

In the absence of S1, primarily two FRET states are observed – a mid-FRET (blue curve) and high FRET (red curve) states, with a greater fraction in the high FRET state when ligand is present (**Figure 3-6d** and **e**, top). These two states are in good agreement with previous studies using this RNA construct under similar, though not identical, buffer conditions[83]. In the absence of ligand with 1 mM Mg^{2+}, the low FRET state (green) was also partially occupied and molecules were relatively dynamic, transitioning between the different FRET states with some regularity. A clearer idea of the nature of the RNA conformations present under various conditions in these experiments can be determined from the observed FRET states as follows. FRET efficiency, E, is described by **Eq. 4**, where R_o is the Förster radius for the donor-acceptor fluorophore pair, and r is the interfluorophore distance.

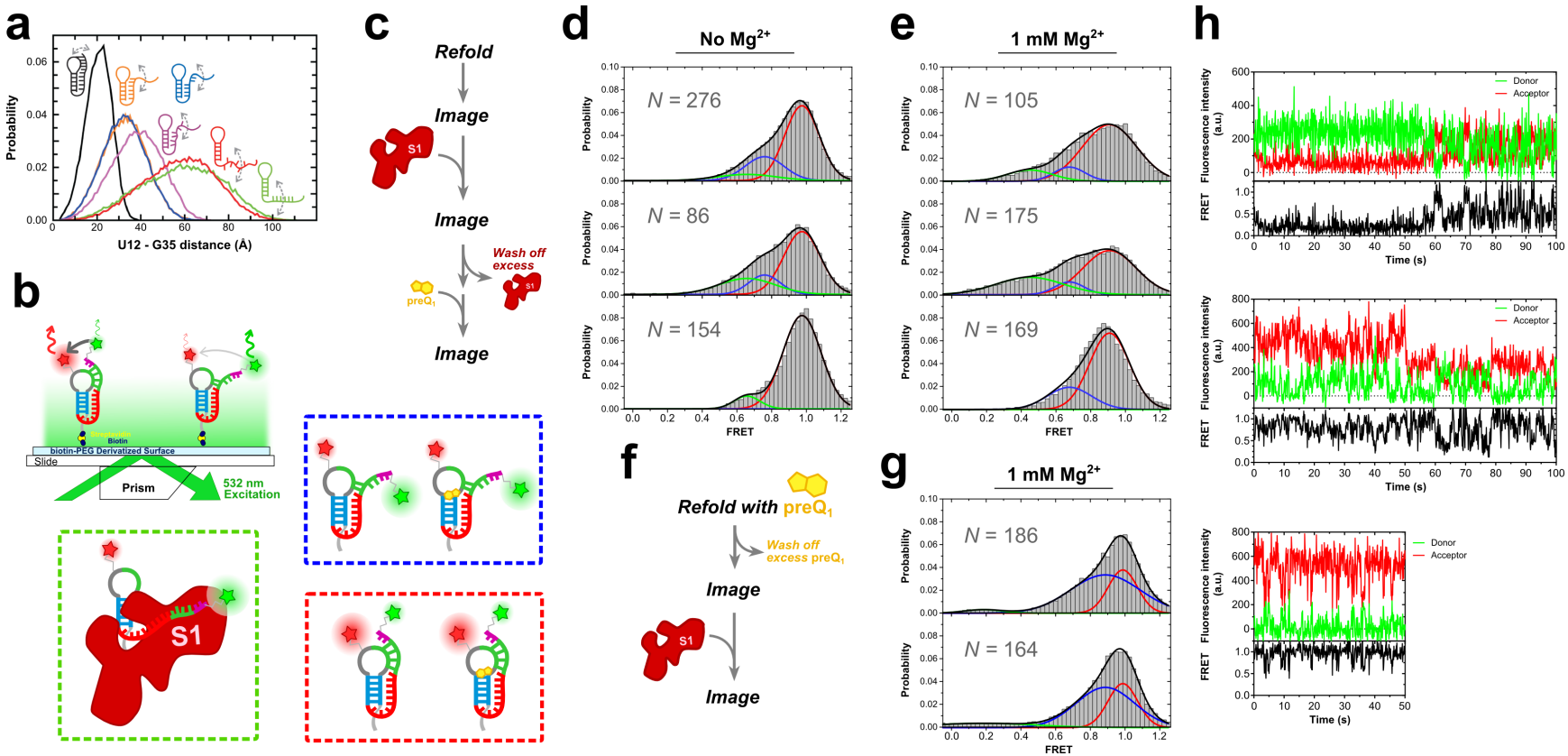


Figure 3-6 Single molecule unfolding of pseudoknot by S1.

(**a**) Expected distribution of interfluorophore distances in the *Tte*^{smFRET} pseudoknot construct from TOPRNA simulations. Reproduced from Ref. 83 under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/). (**b**) Single molecule experiment setup to detect changes in P2 of the pseudoknot by smFRET and cartoon representations of the species that can be distinguished as in **Figure 3-5a**. (**c**) Schematic of the experimental procedure for single molecule data shown in **d** and **e**. (**d**) Histograms and Gaussian fitting of apparent FRET efficiencies from single molecule experiments obtained at each of the "Image" points in depicted in **c,** performed in buffer without added $MgCl_2$. Putative assignments of the species in the underlying populations are color coded as in **b**. *N*, number of single molecule traces the respective condition. (**e**) Same as in **d**, but in buffer containing 1 mM $MgCl_2$. (**f**) Experiment schematic for single molecule data shown in **g**. (**g**) Same as in **e**, but for RNA refolded in the presence of $preQ_1$. (**h**) Example single molecule traces from the +S1 condition (middle) shown in **e**.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad \textbf{(4)}$$

The true FRET efficiency, E, in single molecule experiments such as those presented here can be readily approximated by the apparent FRET efficiency E_{PR},[136] described by **Eq. 5**:

$$E_{PR} = \frac{I_A}{I_A + I_D} \quad \textbf{(5)}$$

where I_A and I_D are the measured intensities of the acceptor and of the donor, respectively. Using a value of 50 Å for the Förster radius[137] and the apparent FRET efficiencies observed in the single molecule experiments (shown in **Figure 3-6d**, **e** and **g**, discussed below) it is possible to calculate the corresponding interfluorophore distances (**Table 3-3**).

The mid and high FRET states correspond to interfluorophore distances of ~35 – 45 Å and ~25-35 Å, respectively. TOPRNA simulations in previous studies[83] (**Figure 3-6a**) showed that the distances calculated for the mid and high FRET states correspond well with the expected distance distributions for conformations of the pseudoknot in which the P2 stem is partially, or fully formed. The low FRET state observed here in the presence of S1 corresponds to an interfluorophore distance of ~45 – 50 Å, which is consistent with a variety of partially unfolded conformations of the RNA in which the P2 stem is not present and the L3 loop is only partially docked and has lost some or all of its interactions with P1 (**Figure 3-6a**, purple and red distributions).

The addition of S1 had the most dramatic effect in experiments performed in the presence of 1 mM Mg^{2+}, but only when the RNA had not yet encountered ligand (**Figure 3-6e**, **h**). Under those conditions, incubation with S1 led to a significant increase in occupancy of the low FRET state (green), suggesting S1-mediated unfolding of the pseudoknot. Subsequent addition of

Table 3-3 Interfluorophore distances from smFRET experiments.

Approximate interfluorophore distances were calculated from the mean FRET values observed in experiments shown in **Figure 3-6** using **Eq. 4** and R_o = 50 Å.

Panel in Figure 3-6	Apparent FRET efficiency *(curve color)*		Interfluorophore distance (Å)
D *No Mg^{2+}*	0.97	*(red)*	27
	0.76	*(blue)*	41
	0.66	*(green)*	45
E *1 mM Mg^{2+}*	0.91	*(red)*	34
	0.68	*(blue)*	44
	0.45	*(green)*	51
G *Refolded with $preQ_1$*	0.99	*(red)*	24
	0.89	*(blue)*	36
	0.18	*(green)*	64

$preQ_1$, which stabilizes the pseudoknot fold and reduces the dynamics of P2[83], resulted in a loss of the low FRET state, and higher occupancy of the high FRET state, as is expected upon ligand binding. Interestingly, while the high FRET state was populated to the exclusion of the mid FRET state after the addition of $preQ_1$ in experiments without Mg^{2+}, a small fraction of molecules continued to sample the low FRET state (**Figure 3-6d**, bottom). This suggests that some molecules were prevented from binding ligand because of the presence of S1. No increase in the low FRET state was observed upon the addition of S1 for pseudoknot RNA that was previously refolded in the presence of excess $preQ_1$ (**Figure 3-6g**). Once bound, the dissociation of $preQ_1$ from the pseudoknot is known to occur only very slowly[63], suggesting that the insensitivity of molecules in **Figure 3-6g** to the presence of S1 results from the fact that the RNA is still largely bound with $preQ_1$.

3.4 Discussion

The known biological roles of S1 are all related in some fashion to the ability of this large protein to bind RNA. S1 plays an essential role in the cell and that perhaps makes it even more

surprising that still relatively few mechanistic details exist describing its capacity to bind RNA, particularly structured RNA.

To shed more light on the way in which S1 interacts with RNA possessing a strong tertiary structure, we started by developing a series of pseudoknot variants based on the well-defined pseudoknot present in the preQ$_1$ riboswitch. Analysis of the melting profiles of these variants reveals that features of the pseudoknot fold, namely P1 and P2 stems, are easily identified, and that the pseudoknot present in the riboswitch can easily be rendered less (C15U) or completely insensitive to ligand (C15A) while still maintaining the same global pseudoknot fold, as indicated by the identifiable presence, but not ligand-dependent stabilization, of the P2 stem. S1 is able to bind to all of the pseudoknot variants described in this study at a site that found to encompass at least part of the pseudoknot fold, as binding of S1 at this site was efficiently inhibited by the presence of preQ$_1$ for pseudoknots that possess the capacity to bind and be stabilized by the presence of ligand (**Figure 3-4**).

Once bound, the pseudoknot RNA is at least partially unfolded by S1 as evidenced by a decrease in FRET for the S1-bound form of a pseudoknot labeled with a FRET pair that reports on the structure in the P2 stem (*Tte*^{smFRET}, **Figure 3-5a**). This is further supported by the increased occupancy of the low FRET (~0.45) state for single molecule experiments conducted in the presence of S1 (**Figure 3-6e**). Numerous transitions between all three of the various FRET states are observed in the presence of S1, indicating that the pseudoknot remains highly dynamic even when bound by S1, or alternatively that S1 binding is highly reversible, or both. This is consistent by the observation that preQ$_1$ is able to displace already-bound S1 from the pseudoknot. In summary, the fact that S1 is able to bind to pseudoknot RNAs (some with thermal stabilities in excess of 50 °C), but not when the pseudoknot stability is further enhanced by

binding of ligand, implies that there is an upper limit to stability of tertiary structures which S1 can interrogate. Furthermore, this phenomenon is likely sequence dependent, as discussed below.

3.4.1 A proposed model for S1-binding of pseudoknot RNA

In light of the well-established preference of S1 for A/U-rich, single-stranded RNA[34,125], a likely candidate for the S1 binding site within these pseudoknots is the sequence 5′ - UUA ACA AAA CAA- 3′, which is comprised of the L3 loop (A-rich tail) and 3 nucleotides of the 3′ half of the P2 stem. The affinity of S1 for the various pseudoknots decreases as the stability of pseudoknot fold increases, and is largely inhibited when P2 is stabilized by the binding of ligand. This is apparent from the observation that S1 preferentially does not bind to the faster migrating form of the RNA pseudoknot, whose conformation is highly similar to that adopted in the presence of ligand, but rather binds preferentially to the less compact form. Taking into consideration that binding of preQ$_1$ serves to greatly decrease the kinetics of docking and undocking of nucleotides in the P2 stem (manuscript in preparation by Suddala *et al.* as well as Suddala and Rinaldi *et al.*[83]) leads to the following mechanism. It is conceivable that binding of S1 to these pseudoknot RNAs occurs through specific binding to L3 and a portion of P2 and that binding can be initiated only when there is partial melting of the P2 stem, thus granting access to the A-rich binding site that is L3. Furthermore, S1 binding to this site on the pseudoknot is reversible and thus binding of ligand, which requires (re)formation of the ligand binding pocket, results in stabilization of P2 and is sufficient to displace S1.

By extension, this suggests that in a more generalized case the binding and S1-mediated unfolding of RNA possessing strong tertiary structure requires that a sequence for which S1 has high affinity must either be exposed or present in a reasonably dynamic structure such that S1 can bind and begin to passively unwind the RNA, similar to what has been described with

respect to S1-mediated unfolding of long dsRNA[40]. In the absence of a sequence with these characteristics, S1 is precluded from binding.

3.5 Materials and Methods

3.5.1 preQ$_1$ ligand

The preQ$_1$ ligand used in this study was synthesized as described previously[59] and was generously provided by Prof. George Garcia at the University of Michigan. The concentrations of the preQ$_1$ stock solutions were measured by UV-vis spectrophotometry using a Nanodrop2000 spectrophotometer (**Figure B.1-1**). The molar extinction coefficient for 7-deazaguanine at 256 nm at pH 6.8 was used (11,200 M^{-1} cm^{-1} determined by Davoll and coworkers[138], Compound VIII with R = NH_2 and R′ = OH), as described previously for other studies involving preQ$_1$[63].

3.5.2 Preparation of Tte *mutant aptamer plasmids by site-directed mutagenesis*

The full set of plasmids encoding the putative *Tte* mRNA transcript, containing a series of mutant variants of the preQ$_1$ riboswitch, under the control of the T7 promoter was generated by site directed mutagenesis from the parent plasmid pUC19_TTE1564, which carries the wild-type (WT) sequence (see **Appendix A.5, Figure A.5-2**). A list of site-directed mutagenesis primers and the resulting *Tte* mutant aptamer plasmids appears in **Appendix Table B.2-1**. Mutagenesis primer sequences were designed with the aid of the QuikChange Primer Design online tool from Agilent (https://www.genomics.agilent.com/primerDesignProgram.jsp). Candidate primer sequences were further optimized using the OligoAnalyzer online tool v3.1 from Integrated DNA Technologies (IDT; http://www.idtdna.com/calc/analyzer) with the default settings (50 mM Na^+, 0 mM Mg^{2+}, 25 °C) to minimize the stability of self-dimers (typically predicted $\Delta G > -$

10 kcal/mol, or <10 – 20% of the ΔG of hybridization to a fully complementary sequence), and the propensity to form internal hairpins (typically T_m < 55 °C).

Mutagenesis conditions for site-directed mutagenesis were based on protocols provided by the lab of Prof. Roger Sunahara, adapted from Sawano and Miyawaki[139]. Because only one mutagenesis primer is required for this method, a primer can be designed to mutagenize either strand. Briefly, primers were purchased from Life Technologies or IDT and 5′ phosphorylated for 30 min at 37 °C at 6 μM final primer concentration in 1X NEBuffer 2.1 supplemented with 1 mM ATP, using T4 Polynucleotide Kinase (New England Biolabs, M0201), which was then heat-inactivated at 65 °C for 20 min. The mutagenesis reaction was performed in a final volume of 50 μL containing 1 nM of the parent plasmid DNA as the template, 0.25 μM of mutagenesis primer, 1 mM of each dNTP, 2.5 U of *Pfu* Turbo DNA polymerase (Agilent), 20 U of *Taq* DNA Ligase (New England Biolabs, M0208), 0.5X *Pfu* Ultra HF Buffer (Agilent), and 0.5X *Taq* DNA Ligase Buffer (New England Biolabs). The reaction conditions were: initial denaturation at 95 °C for 30 s, followed by 20 cycles of denaturation (95 °C for 30 s), annealing (55 °C for 60 s), and extension (68 °C for 8 min, or approximately 2 min per kb of total plasmid length), and then chilled to 4 °C. When the reaction was complete, 20 U of DpnI (New England Biolabs, R0176) were added to the reaction and incubated at 37 °C for 1 hr, and then chilled to 4 °C. 25 μL of the JM109 strain of competent *E. coli* cells (Promega) were transformed with 1 μL of the reaction mixture following the transformation protocol provided in the QuikChange Multi Site-Directed Mutagenesis Kit manual (Agilent) using appropriately scaled volumes, and plated onto LB-agar plates containing 100 μg/mL ampicillin. Clones carrying the plasmids with the desired mutation(s) were identified by Sanger sequencing (**Figure B.2-1**) with the following sequencing primers: 5′-TGTGGAATTGTGAGCGGA-3′ and 5′-TGTAAAACGACGGCCAGT-3′.

3.5.3 Preparation of DNA templates for in vitro *transcription*

DNA primers were purchased from Life Technologies and gel-purified to remove truncated sequences before use. Primers were electrophoresed on a 20% denaturing polyacrylamide gel containing 8 M urea in 1X TBE (Urea-PAGE). The primer band was visualized by UV shadowing briefly using a 312 nm lamp, cut out from the gel, and extruded through a needle-less 3 mL syringe. Primer was eluted from the gel pieces overnight at 4 °C in 3 mL of elution buffer (500 mM NH_4OAc, 0.1% [w/v] SDS and 0.1 mM EDTA). The eluted solution was extracted twice with an equal volume of chloroform saturated with TE (pH 6.5) and then primers were precipitated from the aqueous phase by adding 2 volumes of absolute ethanol, stored overnight at -20 °C, and collected by centrifugation at 12,800 × *g* for 45 min at 4 °C. The pelleted material was washed once with 1 mL of cold 70% (v/v) ethanol and then dried under vacuum. Dried primers pellets were then resuspended in milliQ water and desalted using Illustra MicroSpin G-25 columns (GE Healthcare); the concentration of the final primer solution was calculated from the A_{260} measured using a Nanodrop2000 spectrophotometer, using the manufacturer provided extinction coefficient. Sequences and properties of the primers used appear in **Appendix Table B.2-2**.

DNA oligonucleotides designed to give a partially double-stranded template for transcription (**Appendix Table B.2-2**) were gel purified in a similar manner to that described above, with slight modifications. DNA oligonucleotides were purchased from IDT and 200 – 400 µg of oligonucleotide was electrophoresed on an 18 cm 20% Urea-PAGE gel. The oligonucleotide band was visualized by UV shadowing briefly using a 312 nm lamp. The top half of the band was cut out from the gel, and eluted, precipitated and resuspended (without desalting) as described above. Extinction coefficients estimated using OligoCalc[140] were used to

calculate the concentration of purified oligonucleotide from the A_{260} as described. Successful removal of truncated sequences was confirmed by electrophoresing a 50 ng sample on a 20% Urea-PAGE gel, followed by staining with a 1:10,000 dilution of SYBR Gold nucleic acid stain (Life Technologies) in 1X TBE, and visualizing on a UV transilluminator. The partially double-stranded duplex template was assembled using conditions similar to those described in the MEGAshortscript T7 transcription kit manual (Life Technologies): 10 µM each of the (+) strand and respective (-) strand were combined in a final volume of 50 µL with 10 mM Tris-HCl (pH 8.0 at 22 °C) and 100 mM NaCl, heated for 10 min a 90 °C copper bead bath, and then allowed to cool to room temperature over 15 min.

Templates for *in vitro* transcription of the +6 series of pseudoknot RNA constructs (**Figure 3-1a**) were generated by PCR using the gel-purified primers pUC19_Univ_01F and Tte1564_+6_01R (**Appendix Table B.2-2**). The PCR reaction was performed using 10 ng of the respective plasmid DNA as the template, 0.5 µM of each DNA primer, 1 U Phusion High-Fidelity DNA polymerase (New England Biolabs, M0530), 1X Phusion HF Buffer, 200 µM of each dNTP in a 50 µL final volume. The PCR reaction conditions were: initial denaturation at 98 °C for 10 s, followed by 30 cycles of denaturation (98 °C for 10 s), annealing (57 °C for 15 s), and extension (72 °C for 7 s). When the reaction was complete 10 µL of 3 M NaOAc (pH 5.2) was added to each reaction and the PCR product was purified using the QIAquick PCR purification kit (Qiagen), and the concentration measured using a Nanodrop2000 spectrophotometer. The quality of the PCR product was assessed on a 10% non-denaturing PAGE electrophoresed at 25 V/cm for 2.5 hr at 4 °C, and then stained with a 1:25,000 dilution of GelRed nucleic acid stain (Biotium Inc.) in 1X TBE and visualized on a UV transilluminator. PCR products were also assessed by Sanger sequencing (**Figure B.2-2**).

3.5.4 RNA preparation for EMSA and melting curve studies

RNA pseudoknot constructs were generated by *in vitro* transcription using N-terminally His-tagged T7 RNA polymerase prepared in-house using a method adapted from that described by He *et al.*[141] with the following modifications: T7 was expressed in a different strain of BL21 *E. coli*, 250 mM NaCl was added to all buffers to reduce co-purification of other proteins, and batch-binding to the nickel resin was used in place of column loading. Reaction conditions for *in vitro* transcription were modeled after those described previously[142,143]. PCR products were used as templates for the +6 series of pseudoknot RNA constructs. Briefly, 60 µL transcription reactions containing 150 nM PCR product, 120 mM HEPES-KOH (pH 7.6 at 22 °C), 0.01% (v/v) Triton X-100, 30 mM $MgCl_2$, 7.5 mM of each NTP, 40 mM DTT, 2 mM spermidine trihydrochloride, 0.2 mg/mL T7 RNA polymerase, and 0.01 U/µL Inorganic pyrophosphatase (MP Biomedicals) were incubated in a circulating water bath at 37 °C for 4.5 hr, and then mixed with an equal volume of 2X gel loading buffer (95% (v/v) formamide, 18 mM EDTA, 0.025% [w/v] each of SDS, bromophenol blue, and xylene cyanol) to stop the reaction. The reaction with loading buffer was heated in a 90 °C copper bead bath for 3 min and then snap cooled on ice. The transcript was gel purified as described above on a 20% Urea-PAGE gel, and identified by UV-shadowing with a 312 nm lamp. The bottom half of the band was cut from the gel and extruded through a needle-less 3 mL syringe. The RNA was eluted from the gel, precipitated and resuspended (without desalting) in milliQ water as described above. Extinction coefficients estimated using OligoCalc[140] were used to calculate the concentration of purified RNA from the A_{260} as described. *In vitro* transcription reactions to prepare the minimal aptamer series of constructs using partially double-stranded oligonucleotide templates were performed in a similar manner, except with a transcription reaction volume of 150 µL and with an incubation time of

16-18 hr. Complete sequences of all RNA constructs used in this study are presented in **Appendix Table B.2-3**.

3.5.5 3' fluorophore labeling of RNA constructs

RNA constructs prepared by transcription as described above were labeled with a Cy3 fluorophore at their 3′ end following a method described previously by Willkomm and Hartmann[144] with several modifications. Briefly, RNA constructs were first oxidized by incubating 5 μM RNA in 100 mM NaOAc (pH 5.2) with freshly prepared 2.5 mM sodium (meta)periodate (Fluka, 71859) on ice for 70 min, protected from light. Subsequently, the oxidized RNA was precipitated with the addition of 0.1 V of 3 M NaOAc (pH 5.2) and 2.5 V of cold absolute ethanol, followed by incubated on dry ice until frozen. The solution was inverted until just thawed and then centrifuged at 20,800 × *g* for 45 min at 4 °C to pellet the RNA. The supernatant was removed by pipetting, and the pellets were then washed with ~0.3 V of cold 70% (v/v) ethanol and centrifuged again for 20 min. The wash was removed by pipetting and the pellets were dried under vacuum.

The oxidized RNA was then coupled with a hydrazide derivative of the fluorophore Cy3 (GE Healthcare, PA13120). A typical 100 μL coupling reaction contained ~0.2 – 1.0 nmol of RNA, 50 nmol of Cy3 hydrazide (dye) dissolved in 10 μL of DMSO, and 100 mM NaOAc (pH 5.2). Solutions were degassed prior to the addition of dye, and the headspace above fully assembled reactions was flushed with nitrogen before capping the reaction tube. Reactions were protected from light and incubated at room temperature for 4 hr with agitation. In all subsequent steps, solutions were protected from light. After the end of the incubation, the Cy3-labeled RNA was precipitated with the addition of 0.1 V of 3 M NaOAc (pH 5.2) and 2.5 V of cold absolute ethanol, followed by incubated on dry ice until frozen. The solution was inverted until just

thawed and then centrifuged at 20,800 × *g* for 45 min at 4 °C to pellet the RNA. The supernatant was removed by pipetting, and the pellets were then washed with 2 V of cold 70% ethanol and centrifuged again for 20 min. The wash was removed by pipetting and the pellets washed again with 0.5 V of cold 70% ethanol and centrifuged again for 15 min. This final wash was removed by pipetting and the RNA pellets were dried under vacuum and resuspended in 30 μL cold milliQ water, and then desalted using Illustra MicroSpin G-50 columns that had been pre-equilibrated in milliQ water. The final concentration of RNA in the recovered solution was determined spectrophotometrically using a Nanodrop2000 spectrophotometer, using the respective extinction coefficient at 260 nm for the RNA (**Appendix Table B.2-3**) and ε_{550} = 150,000 M^{-1} cm^{-1} for Cy3. The contribution of dye to the absorbance at 260 nm was accounted for as follows: $A_{260,RNA} = A_{260} - 0.08 \times A_{550}$.

3.5.6 Expression and purification of E. coli *ribosomal protein S1*

The parent pCA24N plasmid containing the *rpsA* gene, encoding the *E. coli* ribosomal protein S1, with an additional N-terminal His-tag was prepared from the ASKA(-) clone JW0894 (National BioResource Project – *E. coli* at National Institute of Genetics)[145]. Because of the generalized cloning strategy used to create the ASKA library, the cloned ORF has a 7 amino-acid linker sequence, TDPALRA, between the 6×His-tag and the second native amino acid encoded by *rpsA*. To make the N-terminal His-tag cleavable, this 7 amino-acid sequence was mutated to the recognition sequence for TEV protease[146] using site-directed mutagenesis, after first verifying that a TEV recognition site, EXXYXQ^(G/S), was not already present in the S1 ORF.

When designing the mutagenesis primer, the number of base changes needed to achieve the desired sequence of amino acids ENLYFQ^G was minimized through the use of synonymous codons as follows. When the use of degenerate codons was possible, the degenerate codon that

required the fewest number of nucleotide changes was used (**Figure B.5-1**), but only when the codon prevalence is > 35% for instances of that amino acid in the *E. coli* ORFeome, as calculated in the Codon Usage Database[147]. Additionally, a second mutagenesis primer was designed to insert an additional stop codon at the C-terminus, removing additional amino acids not encoded in the native *rpsA* gene, resulting from the generalized cloning strategy. Candidate primer sequences were evaluated using similar criteria outlined above in **3.5.2**, ultimately yielding the mutagenesis primers ASKA-rpsA_inTEV_01R: 5′-gag caa aag att cag tGC CCT GAA AAT ACA GAT TCT Cat ggt gat ggt gat gg-3′ and ASKA-rpsA_inOchre_01F: 5′-aaa gca gct aaa ggc gag TAA cta tgc ggc cgc taa ggg-3′ (complementary flanking sequences are show in lower case). The mutagenesis reaction was performed essentially as described above in **3.5.2**, with slight modifications: both mutagenesis primers were present in the same reaction, *Pfu* Ultra DNA polymerase (Agilent) was used in place of *Pfu* Turbo, and the extension phase of the cycle was 12.5 min at 68 °C, and cells were plated onto LB-agar plates containing 170 µg/mL chloramphenicol. Candidate clones were screened by restriction endonuclease digestion with BamHI and those exhibiting only a single band were sequenced by to confirm successful mutagenesis at both sites, using the following sequencing primers: 5′-CAGGAAACAGCTATGACC-3′, 5′-ATTCGTGCGTTCCTGCCA-3′, 5′-GTCTGACATCTCCTGGAACG-3′, and 5′-CGAGCGTTCTGAACAAATCC-3′. The full sequence for the final plasmid (pCA24N_6xHis-TEV_rpsA) appears in **Figure B.5-2**.

pCA24N_6xHis-TEV_rpsA was expressed in the BLR(DE3) strain of *E. coli* using conditions loosely based on those described by Lancaster *et al.*[148]. An overview of the purification scheme is shown in **Figure B.5-3**. The standard chloramphenicol concentration (170 µg/mL) was found to be inhibitory, so the antibiotic concentration used in liquid media was

lowered to 68 μg/mL. 1 L of LB-Miller broth containing 68 μg/mL chloramphenicol was inoculated 1:500 from a saturated overnight culture and grown with shaking at 37 °C, and induced with 1 mM IPTG at an OD_{600} ~0.6. The culture was harvested 2 hr post-induction by centrifugation at 5,000 rpm for 15 min at 4 °C in a Beckman JLA 8.100 rotor. All subsequent steps were performed at 4 °C or on ice. The cell pellet was resuspended in 30 mL of buffer B (15 mM Tris-HCl [pH 7.05 at 25 °C], 30 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol, 0.1 mM PMSF), then pelleted by centrifugation at 6,800 × *g* for 10 min, and then stored at -80C °C for later use. The pellet was resuspended in 30 mL buffer B and lysed in 2 passes through an M-110L Microfluidizer processor (Microfluidics). The lysate was cleared by centrifugation at 10,400 × *g* for 45 min, and then combined with 5 mL of Ni-NTA Agarose resin (Qiagen, 30210) that has been pre-equilibrated in buffer B. The mixture was tumbled for ~2.5 hr, then transferred to a disposable Econo-Pac column (Bio-Rad Laboratories, 9704652) and drained. The resin was washed with 25 mL of buffer C (15 mM Tris-HCl [pH 7.05 at 25 °C], 30 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol, 10 mM imidazole [pH 8.0]) with 500 mM NaCl to reduce the amount of co-purifying RNA, and then washed again with 25 mL of buffer C to remove excess Na^+. Bound protein was eluted from the resin in 4 fractions using 15 mL total of buffer D (15 mM Tris-HCl [pH 7.05 at 25 °C], 30 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol, 250 mM imidazole [pH 8.0]). 5 μL aliquots from each fraction were run on a 9% Tris-glycine SDS-PAGE gel and stained by Coomassie Brilliant Blue R-250. Fractions containing significant amounts of 6×His-TEV-S1 were pooled and the concentration of 6×His-TEV-S1 was estimated from the A_{280} of the solution using a Nanodrop2000 spectrophotometer and an estimated ε_{280} = 48,930 M^{-1} cm^{-1} (ExPASy ProtParam, Swiss Institute of Bioinformatics). Approximately 41 mg of 6×His-TEV-S1 in 7.5 mL of buffer D was thoroughly mixed with ~0.5 mg of TEV protease

prepared in-house[149] and transferred to 10,000 MWCO dialysis tubing and dialyzed overnight into buffer E (15 mM Tris-HCl [pH 7.05 at 25 °C], 5 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol). The dialyzed solution was combined with 5 mL of Ni-NTA Agarose resin (Qiagen) that has been pre-equilibrated in buffer E, and tumbled for ~3 hr. The flow through from this second nickel affinity column was directly loaded onto a 5 mL Q Sepharose Fast Flow anion exchange column (GE Healthcare, 17-0510-01), pre-equilibrated with buffer E. The column was washed with 15 mL of buffer E, and then eluted with increasing amounts of buffer F (15 mM Tris-HCl [pH 7.05 at 25 °C], 600 mM NH_4Cl, 10 mM $MgCl_2$, 6 mM β-mercaptoethanol) in buffer E. A step-wise gradient was from 0 – 80% buffer F over 100 mL with a 10 mL step-size was used, and the S1 protein eluted between ~30 – 50% buffer F, as assessed by checking 20 μL aliquots of each fraction by 9% Tris-glycine SDS-PAGE. S1-containing fractions were pooled and concentrated using an Amicon Ultra-15 10,000 MWCO centrifugal filter unit (EMD Millipore) to a final volume of ~5mL and transferred to 10,000 MWCO dialysis tubing, and dialyzed into storage buffer A(10) (25 mM Tris-HCl [pH 7.05 at 22 °C], 100 mM NH_4Cl, 10 mM $MgCl_2$, 10% [v/v] glycerol, 6 mM β-mercaptoethanol). The S1 concentration was measured from the solution A_{280} after dialysis using a Nanodrop2000 spectrophotometer and an estimated ε_{280} = 47,440 M^{-1} cm^{-1} (ExPASy ProtParam, Swiss Institute of Bioinformatics), then aliquoted, snap frozen with liquid nitrogen, and stored at -80 °C. Protein aliquots were removed for use and thawed on ice; aliquots were kept for up to one week stored at -20 °C and then discarded.

3.5.7 preQ$_1$ and S1 electrophoretic mobility shift assays

Binding of S1 to the various pseudoknot variants (i.e., the +6 and minimal pseudoknot series, **Appendix Table B.2-3**) was assessed using electrophoretic mobility shift assays (EMSA), using

conditions adapted from McGinness *et al.*[128]. Assay conditions were converted to a minigel format (84 mm × 74 mm × 1.0 mm), using a Mini-PROTEAN® Tetra Cell apparatus (Bio-Rad). All solutions containing fluorophore-labeled RNA were protected from light. For titration experiments, $preQ_1$ and adenine were serially diluted in milliQ water at 10X the desired final concentration; serial dilutions of S1 were prepared similarly in buffer A(10). A typical binding reaction was prepared as follows: 1 µL of milliQ water, 10X $preQ_1$ dilution series, or 10X adenine dilution series was combined with 1 µL of 0.1 µM of 3′ Cy3-labeled RNA (0.1 pmol total) and 6 µL of 1.67X binding buffer (16.7 mM Tris-HCl [pH 7.5 at 22 °C], 167 mM NH_4Cl, 8.3% (v/v) glycerol, 1.67 mM DTT) in a 0.2 mL PCR tube. The mixture was refolded by heating in a 90 °C copper bead bath for 3 min and allowed to cool for at least 15 min at room temperature (refolded RNA solution). In $preQ_1$ titration experiments, 1 µL each of 1 mg/mL BSA and buffer A(10) were added to the refolded RNA, mixed thoroughly, and incubated at room temperature for 30 min. In S1 titration experiments, 1 µL each of 1 mg/mL BSA and 10X S1 dilution series were added to the refolded RNA, mixed thoroughly, and incubated at room temperature for 30 min. In S1 competition experiments, 1 µL each of 1 mg/mL BSA and 2.5 µM S1 in buffer A(10) were added to the refolded RNA, mixed thoroughly, and incubated at room temperature for 30 min. The buffer composition in the final reactions (excluding RNA, $preQ_1$, adenine, and S1) was 12.5 mM Tris-HCl, 125 mM NH_4Cl, 6% (v/v) glycerol, 100 µg/mL BSA, and 1 mM $MgCl_2$ in a final volume of 10 µL.

Following the 30 min incubation, reactions were mixed with 10 µL of cold loading buffer (10 mM Tris-HCl [pH 7.5 at 22 °C], 100 mM NH_4Cl, 100 µg/mL BSA, 60% [v/v] glycerol, 0.03% [w/v] bromophenol blue, 1 mM DTT) and placed on ice. Samples were electrophoresed at 4 °C on 9% (+6 series RNAs) or 12% (minimal series RNAs) native polyacrylamide gels in 1X

TGE (25 mM Tris base, 190 mM glycine, 1 mM EDTA) that had been prerun for ~1 hr at 50 V, and then run at 200 V until the bromophenol blue band was at the bottom edge of the gel, ~70 min for 9% gels, or ~90 min for 12 % gels. Gels were scanned in the glass plates on a Typhoon™ 9410 Variable Mode Imager (GE Healthcare) operating in Fluorescence mode, with 532 nm laser excitation, default emission filter set for Cy3 (580 BP 30), and 50 µm pixel size. The PMT voltage was adjusted to maximize signal without saturating the detector (typically 720 - 760 V). The fractions of bound and unbound RNA in resulting images were quantified in ImageQuant v5.2 (Molecular Dynamics) as described in **Appendix B.7**.

Data from the S1 titration binding experiments were analyzed using a two-site binding model as described previously[150]. The apparent dissociation constants $K_{d,slow}$ and $K_{d,fast}$ were determined in Prism (GraphPad; **Appendix B.10**) by simultaneously fitting the fraction of RNA bound in Complex 2 (θ_{slow}) and the fraction of RNA bound in Complex 1 (θ_{fast}) to **Eq. 6** and **Eq. 7**:

$$\boldsymbol{\theta}_{\mathbf{fast}} = \frac{[\mathbf{S1}]\boldsymbol{K}_{\mathbf{d,slow}}}{[\mathbf{S1}]^2 + [\mathbf{S1}]\boldsymbol{K}_{\mathbf{d,slow}} + \boldsymbol{K}_{\mathbf{d,fast}}\mathbf{K}_{\mathbf{d,slow}}} \quad \textbf{(6)}$$

$$\boldsymbol{\theta}_{\mathbf{slow}} = \frac{[\mathbf{S1}]^2}{[\mathbf{S1}]^2 + [\mathbf{S1}]\boldsymbol{K}_{\mathbf{d,slow}} + \boldsymbol{K}_{\mathbf{d,fast}}\boldsymbol{K}_{\mathbf{d,slow}}} \quad \textbf{(7)}$$

These equations use the assumption that binding of RNA by S1 does not deplete the available pool of free S1. Under the conditions of this assay, this is likely not a valid assumption and so in order to obtain more accurate values for $K_{d,slow}$ and $K_{d,fast}$ these equations should be modified to account for the change in the free S1 concentration as has been done previously for a 1-site binding model (**Appendix B.11**).

3.5.8 In-gel FRET electrophoretic mobility shift assay

Structural changes in the doubly fluorophore-labeled RNA construct (*Tte*^{smFRET}, Appendix Table B.2-3) upon S1 binding was assayed using a similar EMSA to that described above, with slight differences. The *Tte*^{smFRET} construct was gel purified on a 20% Urea-PAGE gel to rigorously remove RNA that was labeled with only Dy547 (*Tte*$^{Dy547\ only}$), as well as residual free Cy5 fluorophore that is not covalently attached to the RNA (**Figure B.6-2**). This additional step is required for this assay to simplify the interpretation of in-gel FRET. A solution of 1 μM RNA in milliQ water was refolded by heating for 2 min in a 70 °C copper bead bath then allowed to cool to room temperature over 20 min. The binding reactions were assembled with 1 μL of 1 μM RNA (1 pmol total), 1 μL of 10X S1 dilution series, and buffer in a final volume of 10 μL; the buffer composition in the final reactions (excluding RNA and S1) was 12.5 mM Tris-HCl, 125 mM NH_4Cl, 6% (v/v) glycerol, 100 μg/mL BSA, and 1 mM $MgCl_2$. Note that a higher concentration of RNA is used in these reactions compared to those described above. The assembled binding reactions were incubated at room temperature for 50 min, after which loading buffer was added as described above. The samples were electrophoresed at 4 °C on a 12% native polyacrylamide gel (18 cm × 14.5 cm × 1.5 mm) in 1X TGE that had been prerun for ~3 hr at 100 V, and then run at 18 mA for 3 hr. Reference lanes containing *Tte*$^{Dy547\ only}$ and a doubly Cy5 end-labeled DNA strand (5′-Cy5-CTTTACCACAAGGATGTG-Cy5-3′) were included for later use in correction of background and cross-talk between fluorescence channels. The gels was scanned in the glass plates on a Typhoon™ 9410 Variable Mode Imager (GE Healthcare) operating in Fluorescence mode. For in-gel FRET, the gel was imaged first with 532 nm laser excitation and the default emission filter set for Cy3 (580 BP 30) and PMT voltage of 580 (channel 1). The gel was then imaged with 532 nm laser excitation and the default emission filter set for Cy5 (670 BP 30) and PMT voltage of 650 V (channel 2). Cross-talk between channels and

background was corrected in FluorSep v2.2 (Molecular Dynamics) using automatic fluorochrome separation as described in the software's manual; the bands in the Tte^{Dy574} only and Cy5-DNA reference lanes were each boxed separately as references for channels 1 and 2, respectively. The approximate fractions of RNA in each band as a function of S1 concentration were calculated from a Cy5-only scan, in which Cy5 is directly excited (633 nm laser, 670 BP 30 filter set, PMT voltage of 550), in ImageQuant v5.2 (Molecular Dynamics).

3.5.9 Melting curve studies

Because Tris-based buffers are poorly suited for melting curve studies, melting experiments were performed using sodium phosphate as the buffering salt[135,151]. Melting experiments were performed on a Beckman DU® 640B spectrophotometer fitted with a High Performance Temperature Controller unit, Transport accessory, and T_m six-Cell Holder. A typical sample for melting curve analysis was prepared as follows: a solution of 0.3 µM RNA construct (no fluorophores) in 10 mM sodium phosphate (pH 7.0 at 22 °C), 100 mM NH_4Cl, and 1 mM $MgCl_2$ was refolded by heating for 3 min in a 90 °C copper bead bath, then transferred to a 70 °C copper bead bath for 3 min, then transferred to a 60 °C heating block for 3 min, and finally allowed to cool to room temperature over 20 min. 325 µL of refolded RNA solution was carefully transferred to each cuvette (Beckman, 523878) and tightly stoppered. The instrument was blanked with buffer without RNA, and absorbance at 260 nm was monitored with a 0.5 sec read averaging time with the absorbance 340 nm or 320 nm used for background correction. Cuvettes were allowed to equilibrate in the instrument at 10 °C for 15 min before the start of the discontinuous heating ramp. The cuvette holder was purged with a gentle of nitrogen during portions of the ramp < 20 °C to prevent condensation on the cuvette. The temperature was increased at a rate of 1 °C/min between 10 – 22 °C with a reading made every 1 °C, then at a rate

of 0.5 °C/min between 22 – 75 °C with a reading made every 0.5 °C, then at a rate of 1 °C/min between 75 – 95 °C with a reading made every 1 °C. For experiments where the cooling curve was also measured, the sample was held at 95 °C for 10 min at the end of the heating ramp, and then the same temperature ramp program was run in reverse. Otherwise, cuvettes were promptly removed at the end of the heating ramp and the RNA allowed to refold in the cuvettes while cooling from 95 °C to room temperature over 20 min. The refolded solutions were then removed from the cuvettes. For experiments done in the presence of saturating $preQ_1$, 300 µL of the refolded RNA cuvette solution was combined with 25 µL of 5.4 µM $preQ_1$ in the same buffer and returned to the cuvette. The resulting solution contained a 1.5X molar excess of $preQ_1$ over RNA. The same temperature ramp program was then run. For experiments done in the absence of Mg^{2+}, the buffer composition during refolding was 10 mM sodium phosphate (pH 7.0 at 22 °C), 100 mM NH_4Cl. Additional experiments to better characterize the RNA folding properties were also performed as describe above in 10 mM sodium phosphate [pH 7.0 at 22 °C] without any additional added mono or divalent ions.

3.5.10 Analysis of melting curves

Custom scripts were written in Matlab (The MathWorks) to partially automate the processing and analysis of data from melting curve experiments (**Appendix B.8** and **B.9**). As advocated by Owczarzy[152], an approximated second derivative of the absorbance *vs* temperature was plotted to aid the user in the appropriate selection of upper and lower baseline regions[135,151]. The fraction folded (α)[153] as well as the fraction unfolded (θ)[152] as a function of temperature (T) are calculated from the baseline-corrected absorbance values, according to the equations $\alpha = (\mathrm{A_U} - \mathrm{A})/(\mathrm{A_U} - \mathrm{A_L})$, and $\theta = (\mathrm{A} - \mathrm{A_L})/(\mathrm{A_U} - \mathrm{A_L})$, where A, A_U, and A_L are the absorbance, absorbance of the upper baseline, and

absorbance of the lower baseline, respectively. The resulting plot of α versus T was smoothed using a Savitzky-Golay FIR smoothing filter with a polynomial order of 1. The window for smoothing was varied from 5 – 17 points depending on the quality of the data. The local maxima in the first derivative plot of dα/d(1/T) corresponds to the melting temperature (T_m) for the special case of intramolecular unfolding present here[135]. This method of determining T_m is less sensitive to the choice of baselines than other methods[135], such as the maximum of the first derivative of absorbance versus temperature (dA/dT). Because many of the melting curves displayed clear 2-step melting behavior, the T_m for each apparent transition was determined by fitting the plot of dα/d(1/T) versus T with the sum of one of more Gaussians. Example output from the Matlab scripts appears in **Figure B.9-1**.

3.5.11 smFRET experiments

The doubly fluorophore-labeled RNA pseudoknot construct (*Tte*^{smFRET}, **Appendix Table B.2-3**) used for single molecule experiments in this study is identical to that used in previous studies by Suddala and Rinaldi *et al.*[83]. To reduce non-specific binding of protein that results in high background fluorescence, quartz slide surfaces for single molecule experiments were passivated with a mixture of PEG/biotin-PEG as described previously[122] using established protocols[118]. Microfluidic channels (~80 – 100 µL capacity) were assembled using the passivated slides and coverslips[118,154], and the biotin-PEG reacted with 0.2 mg/mL streptavidin in T50 buffer (10 mM Tris-HCl [pH 8.0 at 2 °C], 50 mM NaCl) for 10 min, then washed again with T50. A typical sample was prepared as follows: 6.25 nM *Tte*^{smFRET} RNA was refolded by heating in a 90 °C copper bead bath for 2 min with 2 µL of 100 mM Tris-HCl (pH 7.5 at 22 °C), 0.5 µL of 4 M NH_4Cl, and 2 µL of 50% (v/v) glycerol in a total volume of 16 µL. The RNA in solution was refolded by allowing it to cool for at least 15 min at room temperature, after which 2 µL of 1

mg/mL BSA and 2 μL of storage buffer A(10). The composition of the resulting 20 μL solution was 5 nM RNA, 12.5 mM Tris-HCl, 125 mM NH_4Cl, 6% (v/v) glycerol, 100 μg/mL BSA, and 1 mM $MgCl_2$. For samples prepared in the absence of Mg^{2+}, buffer A(0) (25 mM Tris-HCl [pH 7.05 at 22 °C], 100 mM NH_4Cl, 10% (v/v) glycerol, 6 mM β-mercaptoethanol) was used in place of buffer A(10). For some samples, 625 nM $preQ_1$ was present during heating and refolding of the RNA; this resulted in a concentration of 500 nM $preQ_1$ after the addition of BSA and buffer A(10) or A(0). The refolded sample was then diluted to a final RNA concentration of 25 pM in buffer I (12.5 mM Tris-HCl, 125 mM NH_4Cl, 6% (v/v) glycerol, 100 μg/mL BSA). Depending on the particular condition being tested, buffer I was also variously supplemented with 1 mM $MgCl_2$ and 500 nM $preQ_1$. This diluted RNA solution was applied to the slide and allowed to incubate for 10 min. Excess, unbound RNA was washed by flowing at least 100 μL of buffer I, supplemented with 500 nM $preQ_1$ and/or 1 mM $MgCl_2$ depending on the particular condition tested. Finally, at least 100 μL of buffer I supplemented $preQ_1$ or 1 mM $MgCl_2$ as well as an oxygen scavenging system, consisting of 5 mM protocatechuic acid, 50 nM protocatechuate-3,4-dioxygenase (to slow photobleaching)[155] and 4 mM Trolox, to reduce photoblinking[120], was applied to the slide and allowed to equilibrate for 5 min before imaging on a prism-TIRF microscope, as described previously[83]. Fluorescence emission from single molecules excited with a 532 nm diode laser was recorded at 10 frames per second (integration time of 100 ms) in mj2 format using an intensified CCD camera (I-Pentamax, Princeton Instruments) using a custom acquisition script written in Matlab. In some instances, the Cy5 fluorophores were directly excited using a 638 nm diode laser for the last ~50 frames of the movie to confirm the presence of both fluorophores (Dy547 and Cy5). Movie files were converted from mj2 to pma format and fluorescence-time traces extracted using IDL (Exelis Visual Information Solutions).

Alternatively, fluorescence time traces were also extracted from the mj2 files directly using custom Matlab scripts (The MathWorks). Genuine fluorescence time traces were selected manually and analyzed using custom Matlab (The MathWorks) scripts as described previously[83]. FRET distribution histograms were built by combining the apparent FRET efficiencies observed in 100 frames from each trace in a given condition. The resulting histograms were fit using OriginLab 9 (OriginLab Corporation) as follows: data from each condition in a particular dataset (**Figure 3-6e, f** or **g**) were concatenated and then fit to the sum of three Gaussians to determine the peak centers shared across conditions within the dataset. The histogram for each condition was then fit separately to the sum of three Gaussians, the positions of which were fixed to the previously determined peak center values. This method assumes that states with the same characteristic mean FRET values exist in each condition and that the addition or removal of S1 or $preQ_1$ changes only the degree to which a given state is populated. Interfluorophore distances were calculated from the apparent FRET efficiencies for each dataset as described in the text and appear in **Table 3-3.**

CHAPTER 4: Riboswitch mRNA interactions with the ribosome at the earliest steps of initiation: Towards the development of single molecule fluorescence assays of translation initiation[3]

4.1 Overview

Even as information is added to our mechanistic description of translational attenuation by the class-I preQ$_1$ riboswitch, new questions continue to arise regarding the nature of the interaction between the riboswitch-containing mRNA and the ribosome. Much has been learned from *in vitro* translation studies using the *Tte* mRNA, however questions remain about the applicability of mesophilic ribosomes to the study of mRNA from a thermophilic species. To address this, a new candidate mRNA containing a class-I preQ$_1$ riboswitch from *Bacillus anthracis* (a mesophile) with features similar to the *Tte* mRNA was identified and shown to be amenable to study using our current set of established assays. In light of the ability of ribosomal protein S1 from *E. coli* to partially unfold the *Tte* riboswitch pseudoknot, and the surprisingly modest translational repression of *Tte* mRNA attributable to ligand, we looked at the influence of S1 on translation of preQ$_1$ riboswitch-containing mRNAs. The inhibitory effects of ligand on mRNA translation are not greatly potentiated by the loss of S1; however, a dramatic effect was observed

[3] Single molecule FRET experiments were performed in conjuction with May Daher. Cloning of the gene containing the preQ$_1$ riboswitch and prediction of the transcript boundaries for the *Bas* mRNA from *B. anthracis* were performed with the excellent assistance of David J. Smith, who received support from the Universtiy of Michigan Undergraduate Reseach Opportunity Program. 30S ribosomal subunits lacking S6 were prepared by Matthew S. Marek.

on the translational coupling of co-transcribed genes, which invokes a role of S1 in the translation of polycistronic mRNA. These results further underscore the need for better knowledge of how the 30S subunit binds to structured mRNA, and to mRNA in general. To this end, we describe our initial steps towards the development of a highly adaptable single molecule assay to characterize early 30S-mRNA interactions.

4.2 Introduction

In the still relatively young riboswitch field[56], the vast majority of studies have focused on areas such as the identification of novel classes of riboswitches and their ligands[59,67,75,156,157] and structural characterization of the riboswitch aptamer domains[64,158-160]. While there have been an increasing number of advances in the study of riboswitches with their expression platforms[108,161-163], studies of this kind are still only beginning to elucidate the true behavior of riboswitches in a physiological context.

Of particular relevance here are the riboswitch studies that deal explicitly with regulation at the level of translation[108,164-168]. One illustrative example is that of the S_{MK} box motif found in the 5′ UTR of *metK*, a gene encoding S-adenosylmethionine (SAM) in a number of species of Gram-positive bacteria, such as *Lactobacillus acidophilus* and *Enterococcus faecalis*[164]. In this translationally acting riboswitch, the Shine-Dalgarno (SD) sequence is sequestered by binding of, and directly participates in recognition of, the ligand SAM[169]. In work from the lab of Prof. Tina Henkin (Ohio State University) in which the S_{MK} box was first described, Fuchs *et al.*[164] showed that this riboswitch exerts an ~5-fold repression of a lacZ reporter under conditions when intracellular SAM is abundant. This (albeit modest) repression clearly demonstrated that this riboswitch acts on the translational level, and subsequent studies showed that binding of SAM by the riboswitch attenuates, but does not abolish, binding of the riboswitch-containing mRNA by

the 30S subunit[165]. These studies highlight the more general question of how the ribosome interacts with structured mRNA and more specific questions about how the 30S interacts with translationally acting riboswitches (such as the SAM-III riboswitch and the preQ$_1$ riboswitch found in *Tte* mRNA). Thinking about the nature of 30S interactions with mRNA also prompts questions regarding the complicating influences of additional factors such as ribosomal protein S1, which has a known role in mediating the initial interaction between the 30S subunit and mRNA[24,30,34,35].

Single molecule techniques are uniquely positioned to aid in unraveling the mechanistic details of complex, multistep processes such as the initiation of translation. Numerous single molecule studies have provided key insights into the structural dynamics of riboswitches[83,159,170], as well as the more complex dynamics of the ribosome during translation[171-174], such as the fluctuations of bound tRNAs as the ribosome samples between classical and hybrid states[137], and the movement of tRNAs through the ribosome during multiple rounds of decoding and elongation[175]. Such single molecule studies allow for a single step to be isolated in the context of a larger multistep process, where different outcomes can be distinguished for individual molecules, such as the rejection of a non-cognate aminoacyl-tRNA in one elongating ribosome, or the successful accommodation and subsequent translocation in another.

While the elongation phase of translation has received the most attention, single molecule assays to look at various stages of the initiation phase of translation have begun to be developed[176,177], including studies examining the influence of the IF2:initiator tRNA conformation on the 30S intersubunit interface and how this correlates with 50S subunit joining[178], as well as on the interactions between IF2 and IF3 and the resulting effects on subunit joining[179,180]. Both of these studies look at the late stages of initiation in which the 50S subunit

joins and the ribosome prepares to enter elongation. Initial binding of the mRNA by the ribosome[37-39,132] and initiation complex assembly have long been studied through the use of elegant ensemble techniques[6,181,182]; however, fewer single molecule studies[176] have focused on these earliest stages of initiation.

Building on our experience developing the SiM-KARTS technique described in **Chapter 2**, this chapter describes initial steps towards developing a single molecule assay with the ultimate goal of examining the early stages of initiation in which binding between the 30S subunit and riboswitch-containing mRNA occurs, as well as additional experiments to assess how other factors that can play a role in this initial binding step, namely the presence of S1, influence the translation of *Tte* mRNA, as well as a related preQ$_1$ riboswitch-containing mRNA from *Bacillus anthracis*.

4.3 Results

4.3.1 A highly similar preQ$_1$ riboswitch-containing mRNA is found in B. anthracis

Thermoanaerobacter tengcongensis is a thermophilic, Gram-negative bacterium that shares high genetic similarity with the mesophilic, Gram-positive bacterium *Bacillus halodurans*[183]. The preQ$_1$ riboswitch in the 5′ UTR of the *Tte* mRNA from *T. tengcongensis* has undergone extensive structural characterization[63,64,83,85], which makes it an attractive subject for functional studies. By nature of originating in a thermophile, however, the interpretation of *in vitro* translation studies performed on *Tte* mRNA can be potentially convoluted by effects arising solely from translating a thermophilic mRNA using heterologous, mesophilic translation factors and ribosomes. To address this concern, a second riboswitch-containing candidate from a mesophilic bacterial species was sought.

A putative class-I preQ$_1$ riboswitch has been identified[184] in *B. halodurans* upstream of a gene cluster related to the biosynthesis of Queuosine (NCBI accession NC_002570.2; 2373342 – 23773299). This riboswitch, however, is located too far from the first ORF in the operon (~165 nt upstream) to act as a translational riboswitch and is therefore more likely to act transcriptionally. Fortunately, numerous other examples of translationally acting preQ$_1$ riboswitches are known, particularly from other species of *Bacillus*, including *B. anthracis*. The class-I preQ$_1$ riboswitch motifs in *T. tengcongensis* and *B. anthracis* were originally identified bioinformatically by Roth and coworkers[59], (**Figure 4-1a**). In *B. anthracis* the riboswitch is found upstream of a single gene, BAS1509, also called *queT*, and is found in many species of Firmicutes encoding a hypothetical membrane transporter protein suggested to play a role in the uptake of Queuosine biosynthetic intermediates. Interestingly, the key difference between the *Bas* mRNA and *Tte* mRNA transcripts is that in *T. tengcongensis*, the riboswitch is found in the 5′ UTR of a bicistronic operon encoding two genes: TTE1564, a *queT* homolog, and the distal gene TTE1563. TTE1563 in the +2 reading frame, and overlaps part of the last nine codons of TTE1564. TTE1563 encodes the enzyme 7-cyano-7-deazaguanine reductase (QueF), a nitrile reductase responsible for converting preQ$_0$ to preQ$_1$. The QueT proteins from *B. anthracis* and *T. tengcongensis* are similar, sharing 30% and 51% amino acid identity and similarity, respectively. The riboswitches in the transcripts from these two organisms are located upstream of highly similar genes and so for this and other reasons discussed below, the transcript from *B. anthracis* (*Bas* mRNA) was selected as a good mesophilic candidate for further study.

The transcript boundaries for the putative *Bas* mRNA transcript were predicted as described in **Materials and Methods 4.5.1** and the coding sequences and associated 5′ and 3 ′ UTRs were cloned into a plasmid for *in vitro* transcription. While there is no crystal structure

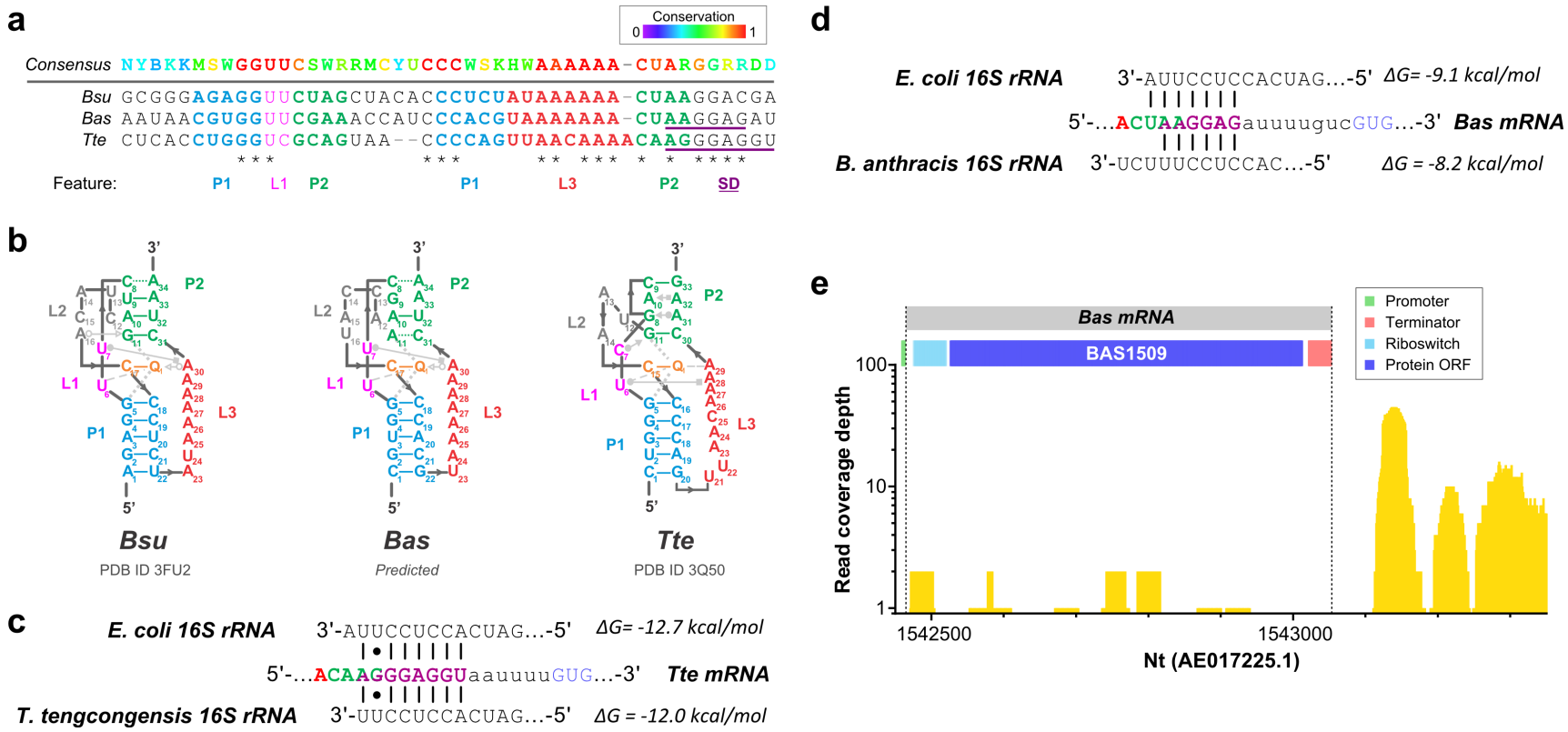


Figure 4-1 Comparisons between translational preQ$_1$ riboswitches from *T. tengcongensis* and *B. anthracis*.

(**a**) Sequence comparison between preQ$_1$ riboswitch aptamer domains from different species. The top line shows the consensus sequence for the preQ$_1$ (class-I) riboswitch, color coded according to sequence conservation as reported in the Rfam database[185] (http://rfam.xfam.org/family/RF00522). Subsequent lines show the aptamer sequences for the preQ$_1$ riboswitches found in *B. subtilis* (*Bsu*, transcriptional), *B. anthracis* (*Bas*) and *T. tengcongensis* (*Tte*) with secondary structural features color coded as in **Figure 2-1a**. (**b**) Tertiary structures of different preQ$_1$ riboswitches with key interactions shown with Leontis-Westhof[86] notation. The structure for the *Bas* riboswitch is predicted based on sequence similarity with *Bsu*. (**c**) Base-pairing interactions between the *Tte* mRNA transcript and *E. coli* or *T. tengcongensis* ribosomes (reproduced from **Figure 2-2a**. (**d**) Same as in **c**, but for the *Bas* mRNA. (**e**)Transcript schematic and read coverage map for the putative *Bas* mRNA transcript from *B. anthracis* transcriptome sequencing[186]. Nucleotide numbering is with respect to the NCBI accession AE017225.1.

available of this preQ$_1$ riboswitch variant, key interactions between the riboswitch and preQ$_1$ ligand, as well as tertiary interactions can be predicted with high confidence by comparison with other preQ$_1$ riboswitch variants that share very high sequence similarity, in particular the transcriptionally acting riboswitch from *B. subtilis* (*Bsu*), for which such structural information does exist (**Figure 4-1a, b**). To build further confidence in the accuracy of the putative transcript prediction, evidence at the transcriptional level for this transcript was mined from next-generation RNA-Seq transcriptome profiling data previously reported in the literature for *B. anthracis*[186]. Sequence-level support for a large fraction of this transcript (**Figure 4-1e**), was found by mapping sequencing reads as described in **Materials and Methods 4.5.2** and **Appendix C.3**. Although the number of sequencing reads (read coverage) for this transcript are very low, especially in comparison to read coverage for adjacent open reading frames in the genome, this is perhaps to be expected from a regulated transcript.

4.3.2 Considerations when using heterologous ribosomes for translation studies

The use of heterologous ribosomes is often an appropriate[165] and convenient choice. Fortuitously, the 16S rRNA sequence in *E. coli* ribosomes is capable of forming very similar interactions with the SD sequences present in both of the riboswitch-containing mRNAs used here (**Figure 4-1c**, **d**), and has been shown to successfully translate preQ$_1$-riboswitch containing mRNA (**Figure 2-1**), despite being from a distant phylogenetic species. The use of heterologous *E. coli* ribosomes, however, still requires careful consideration.

One notable difference between *E. coli* and ribosomes from many commonly studied Gram-positive bacteria is the well-known role played by S1 in translation of some mRNAs by *E. coli* ribosomes[24,28,30,187]. Early work examining translation by ribosomes from Gram-positive species, found that in some[26] but not all[188] cases, the addition of *E. coli* S1 failed to improve the

translation of mRNA, despite its ability to bind to the respective Gram-positive ribosome. These effects were also largely dependent on features in the mRNA translation initiation region, most notably the strength of the SD sequence[26]. Additionally, it was previously thought that S1 homologs were not present in *B. subtilis* and some related *Bacillus* species[188-190]. However, this was later found not to be true[126,191], with the most salient difference being that the S1 from *B. subtilis* and homologs in related species contain only four[126] out of six total OB-fold repeats (a single-stranded RNA binding domain) found in *E. coli* S1. This is also the case for *T. tengcongensis* S1, which bears genetic similarity to *B. halodurans*, despite being Gram-negative like *E. coli* (**Figure C.4-1**). S1 from *E. coli* is known to bind to, and in some cases enhance translation of, ribosomes from *Bacillus* species as discussed above, and indeed it has been shown that only the three N-terminal repeats are involved in ribosome binding, with a strong requirement[30,127] for repeats 1 and 2. Domain repeats 4-6 are dispensable, but cell viability is reduced when these repeats are deleted[30,43]. S1 from *B. subtilis* was not found associated with the ribosomes isolated during late log-phase growth[192], however this may be because the expression of S1 in *B. subtilis* occurs in the sporulation phase of growth[191,193]. Taken together, this prompts one to ask what the functional role of S1 is in *B. subtilis* and related species. For example, it reopens the question of whether these S1 homologs are ever found present on the ribosome, given their unique pattern of expression that is growth-phase specific. The possibility of ribosomes with growth-phase specific composition also brings one to ask what implications that may have for translation of other genes that are not constitutively expressed, as is the case presumably for preQ$_1$ riboswitch containing mRNAs.

In the most general of terms, *E. coli* ribosomes can arguably be converted into a closer approximation of *B. subtilis* ribosomes as they exist in log-phase growth through the removal of

S1. Indeed, even for *E. coli* ribosomes, S1 is loosely bound and is present only in sub-stoichiometric amounts[29] and the interactions of ribosomes with some mRNAs is dramatically altered when S1 is absent[24-26,190]. In light of the ablity of isolated S1 to unfold the preQ$_1$ riboswitch and the modest translational repression observed by *in vitro* translation in the presence of ligand, one hypothesis is that the magnitude of the preQ$_1$-induced down regulation observed *in vitro* would be more dramatic if not for the activity of S1 associated with the heterologous *E. coli* ribosomes used in these experiments. To this end, *in vitro* initiation and translation studies were performed (described below) using S1-depleted ribosomes to examine what, if any, effects S1 has on the translation of preQ$_1$ riboswitch-containing mRNAs.

4.3.3 Heterologous ribosomes can initiate on Bas *mRNA*

As a first step, the ability of heterologous *E. coli* ribosomes to initiate translation on *Bas* mRNA was assessed. The efficiency of 30S initiation complex formation on *Bas* mRNA, *Tte* mRNA, and CAT mRNA by wild-type *E. coli* ribosomes was measured using a filter-binding assay in which radiolabeled initiator tRNA is differentially retained in neutral or positively charged membranes depending on the extent of incorporation into 30S ICs (**Figure 4-2a**). Low levels of fMet-tRNA^{fMet} were bound by 30S subunits in the absence of added mRNA and retained at significantly higher levels when mRNA was included, as expected given that the presence of mRNA stabilizes tRNA binding[194]. The efficiency of 30S IC formation was slightly higher for *Bas* mRNA compared to *Tte* mRNA, which is reasonable given that the *Bas* mRNA originates from a mesophilic organism. The riboswitch in the *Bas* mRNA shares high sequence similarity with the riboswitch from *Bsu* (**Figure 4-1a**) particularly in the P2 stem (directly adjacent to the SD sequence). Thus, the *Bas* mRNA is expected to have a less rigidly structured and more dynamic riboswitch fold than *Tte*, as was observed previously for the *Bsu* riboswitch[83]. For

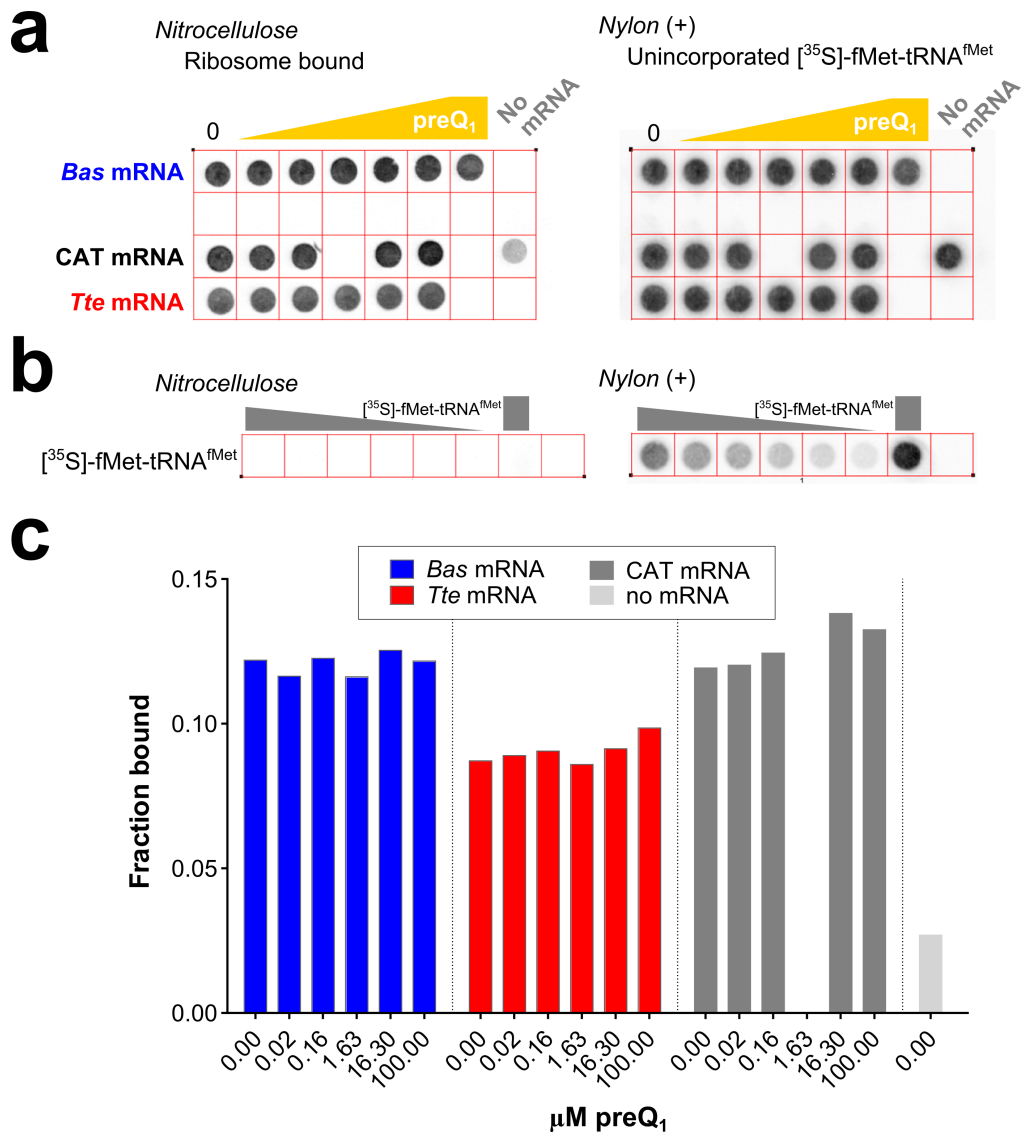


Figure 4-2 Assessment of 30S IC formation efficiency as a function of preQ$_1$ concentration.

(**a**) Autoradiograph of filter binding membranes. Efficiency of 30S IC formation as a function of mRNA species and preQ$_1$ concentration is assessed by taking the ratio of initiator tRNA counts in 30S ICs, which are trapped in the nitrocellulose (top) membrane, over the sum of counts in both membranes. In the absence of added mRNA (3rd row, last column), only low levels of radiolabeled tRNA are bound by ribosome and thus retained in the top membrane. Note that a shorter exposure time is required for the Nylon membrane. The concentrations of preQ$_1$ used are shown in the horizontal axis in **c**. (**b**) A serial dilution of tRNA is included to determine the σ factor, a measure of non-specific retention of tRNA in the top membrane in the absence of ribosomes and mRNA. The σ factor was calculated using 4, 2, 1.4, 0.70, 0.34, 0.17 and 9 pmol of radiolabeled tRNA as described in **Materials and Methods 4.5.10**. (**c**) Quantification of 30S IC-bound tRNA in the presence of various concentrations of preQ$_1$. 30S IC formation was similar for CAT and *Bas* mRNAs, and slightly greater than *Tte* mRNA. No significant dependence of 30S IC formation on preQ$_1$ concentration is observed.

reactions containing preQ$_1$, the mRNA was heated and refolded in the presence of ligand before the addition of 30S subunits, tRNA, additional preQ$_1$, and the other remaining components required for complex formation.

Somewhat surprisingly, pre-incubation with, and complex assembly in the presence of, preQ$_1$ did not have a dramatic effect on 30S IC formation (**Figure 4-2b**) for either *Bas* mRNA or *Tte* mRNA. This suggests that preQ$_1$ is not able to prevent the 30S from initiating on the mRNA when in the presence of tRNA, in contrast to what was observed for the S_{MK} box and its ligand SAM[165]. A variation of this assay was attempted using non-radiolabeled initiator tRNA and fluorophore-labeled mRNA (i.e., *Tte*$^{+30}$-Cy3, discussed below) to examine the amount of mRNA bound by the 30S as a function of preQ$_1$ concentration. The results of this assay were inconclusive, however, due to the high levels of non-specific retention of Cy3-labeled mRNA in the nitrocellulose filter, even in buffer alone (data not shown), likely as a result of increased hydrophobicity of the mRNA from the fluorophore.

4.3.4 E. coli *S1 has unexpected effects on translation of riboswitch-containing mRNAs*

E. coli 30S subunits depleted of S1 (30S$^{\Delta S1}$) were prepared as described in **Materials and Methods 4.5.4** (**Figure 4-3a**, **b**) and used for *in vitro* translation experiments, similar to those described in **Chapter 2**. Using a preparation of wild-type *E. coli* ribosomes in which ~30% of 30S subunits contained S1 (1X salt-washed 30S, **Figure 4-3a-c**), translation of the *Tte* mRNA in the presence of CAT mRNA was reduced in the presence of saturating (100 μM) preQ$_1$ (**Figure 4-3d**) as seen previously (**Figure 2-1c, d**). It should be noted that a lower molar ratio of CAT protein to the sum of *Tte* proteins was observed in the experiment in **Figure 4-3d** than that in **Figure 2-1c, d**. This discrepancy could potentially result from differences in the ribosome preparations used (separated subunits versus salt-washed ribosomes, respectively), or possibly

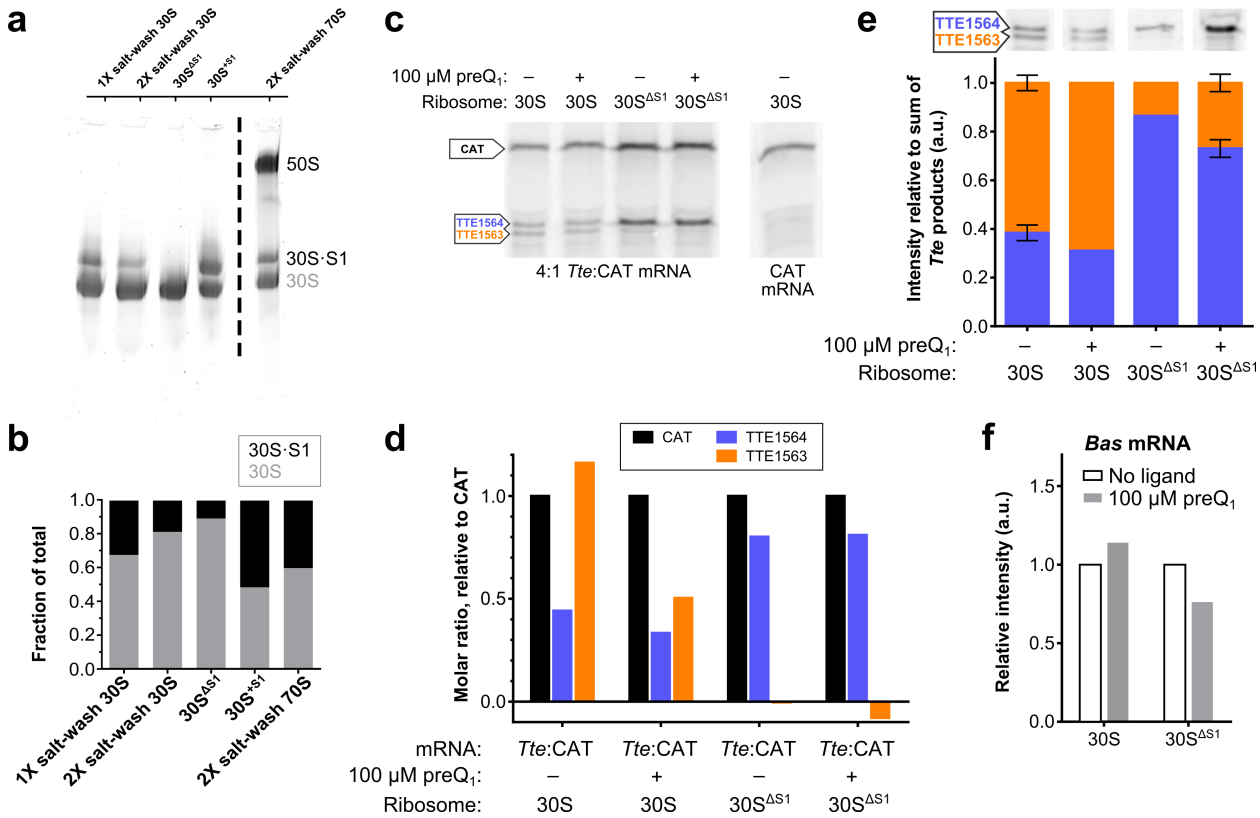


Figure 4-3 Effects of S1-depletion on *in vitro* translation of riboswitch containing mRNAs.

(**a**) Assessment of S1 content in various preparations of 30S subunits and ribosomes. 30S can be efficiently depleted of S1 ($30^{\Delta S1}$), while reconstitution of depleted subunits with a 2-fold stoichiometric excess of recombinant S1 ($30S^{+S1}$) restores S1 to approximately the levels present in salt washed 70S ribosomes. (**b**) Quantification of 30S and 30S•S1 bands in **a**. Quantification of the 30S•S1 band is likely slightly overestimated due to smearing from the 30S band. (**c**) Autoradiograph of *in vitro* translation reactions using 1X salt-washed 50S and 30S subunits (30S) or S1-depleted 30S subunits ($30S^{\Delta S1}$) in the presence or absence of saturating $preQ_1$. (**d**) Quantification of experiment shown in **c**. Intensity is normalized according to the number of methionine residues in each protein (9, 5, 2 for CAT, TTE1564, and TTE1563, respectively), and reported relative to the CAT product in the same lane. Negative values result from a slight over-correction for the contribution of CAT degradation products (light smear), which is more prevalent in translation assays using L-[^{35}S]-methionine, that co-migrate with the TTE1564 and TTE1563 bands. (**e**) Effect of S1-depletion on translation of *Tte* mRNA. Autoradiograph and quantification of translation of *Tte* mRNA using L-[^{35}S]-methionine or L-[^{35}S]-cysteine under the same conditions as in **c**. Intensity is normalized for the number of cysteine or methionine residues in each protein (1, 1 for cysteine; 5, 2 for methionine), and normalized to the sum of the TTE1564 and TTE1563 bands in the same lane. Error bars represent the standard error of the mean of at least 2 samples. (**f**) Effect of S1-depletion on the *in vitro* translation of BAS1509 from *Bas* mRNA in the presence or absence of ligand. Intensity is normalized to the corresponding "No ligand" condition.

from the more complex normalization required when performing the assay using radiolabeled L-methionine instead of L-cysteine (see **Materials and Methods 4.5.8**).

As in the *in vitro* translation experiments presented in **Chapter 2**, the total amount of mRNA present in these experiments is in 2-fold excess over ribosomes. Thus, the translation initiation sites of the various mRNAs present in the reaction must compete for ribosome binding. Translation of TTE1563 from *Tte* mRNA likely results from independent initiation at the *queF* start codon some fraction of the time (the observed ratio of TTE1564:TTE1563 is 40:60, whereas a 50:50 ratio or lower is expected if translation of TTE1563 is completely dependent on translation of TTE1564) in the *in vitro* assays described here, whether in isolation (**Figure 4-3e**) or in the presence of CAT mRNA (**Figure 2-3c**). However, the fact that there is a decrease in translation in the presence of ligand for both TTE1564 and TTE1563, largely preserving the observed ratio of the two reading frames, demonstrates that there is indeed translation coupling between these two ORFs, as one would expect for a bicistronic operon[195].

In translation reactions performed using $30S^{\Delta S1}$, little change was seen with ligand for translation of the TTE1564 ORF, which is proximal to the riboswitch; however, translation of the downstream reading frame TTE1563 was largely abolished upon depletion of S1, irrespective of the presence of ligand. The ratio of TTE1564 relative to CAT was also increased, potentially due to the increased availability of ribosomes in the absence of active translation of TTE1563. This striking result suggests that S1 is required for re-initiation and translational coupling of these two genes in this bicistronic operon when translated by *E. coli* ribosomes.

The fact that no significant change was observed in the translation of the proximal reading frame (i.e. TE1564) by $30S^{\Delta S1}$ ribosomes in the absence or presence of ligand (**Figure 4-3d**) is in

itself somewhat surprising given the ability of S1 to bind the preQ$_1$ riboswitch (demonstrated in **Chapter 3**), and its requirement for the successful translation of other mRNAs[23], particularly those with structured TIR such as the *rpsO* mRNA, which also contains a pseudoknot[30]. It was expected that the removal of S1 would accentuate the effects of ligand on translation efficiency. By contrast, a modest decrease was observed in the translation of BAS1509 in the presence of ligand, but only when using ribosomes depleted of S1 (**Figure 4-3f**). It should be noted, however, that this experiment was done using only *Bas* mRNA without the presence of CAT mRNA to use as an internal reference. Thus, these measurements compare band intensities in different lanes, without a more robust normalization, and as such the magnitude of the changes in this experiment should be treated with caution. This experiment therefore bears repeating in the presence of CAT mRNA, as is routinely done for translation of *Tte* mRNA.

The 30S IC filter-binding findings as well as these intriguing results from *in vitro* translation in the absence of S1 highlight the need for additional techniques that provide a more detailed view of what exactly is occurring during the initial steps of initiation when binding of mRNA by the 30S occurs, and when translational riboswitches are most likely to exert their regulatory control over expression. With this goal in mind, the initial steps taken towards developing a single molecule fluorescence assay for initiation and described in the remainder of this chapter.

4.3.5 Truncation and fluorophore-labeling of the Tte *mRNA does not impede 70S IC formation*

While many methods for the preparation of long, fluorophore labeled mRNAs exist, it is often convenient to prepare and work with shorter labeled RNAs, provided they recapitulate the behavior of the larger RNA. To this end, a number of truncations of the *Tte* mRNA with a variety

of lengths were designed and fluorophore labeled as described in **Materials and Methods 4.5.3**. The truncation lengths were chosen so as to create a set of mRNAs that occupied the mRNA channel on the 30S subunit to different extents (**Figure 4-4a**). The longest of these, *Tte*$^{+30}$-Cy3 was truncated 30 nucleotides into the TTE1564 ORF, and is expected to completely fill the mRNA channel and extend into solvent at both the 5′ and 3′ ends. The shortest, *Tte*$^{-11}$-Cy3, is truncated 11 nucleotides upstream of the ORF at the end of the riboswitch aptamer, and thus contains only a partial SD sequence. To confirm that 3′-fluorophore labeling of the truncated mRNA does not impede the ability of the ribosome to proceed through all steps of initiation and form mature initiation complexes, the efficiency of 70S IC formation on *Tte*$^{+30}$-Cy3 was compared to full-length *Tte* mRNA. 70S IC formation efficiency was measured using a ultracentrifugation assay based on the extent of incorporation of radiolabeled initiator tRNA, depicted in **Figure 4-4b**. *Tte*$^{-11}$-Cy3 was also included for comparison. The results of this assay showed that 70S IC formation is equivalent for *Tte*$^{+30}$-Cy3 and *Tte* mRNA demonstrating that the fluorophore labeling at that position does not interfere with initiation, and is in good agreement with the overall extent of tRNA binding seen in filter-binding assays (**Figure 4-4c**). As expected, initiator tRNA incorporation was not detectable in reactions containing *Tte*$^{-11}$-Cy3, which lacks a start codon and thus cannot support initiation.

4.3.6 Successful preparation of fluorescently labeled ribosome for single molecule studies

Multiple strategies for site-specific fluorescent labeling of the ribosome have been developed for use in single molecule studies. One example targeting RNA components of the ribosome was first described by Dorywalska *et al.*[196], in which insertions are made at specific positions in the rRNA, providing exposed “handles” to which small, labeled nucleic acids can be hybridized. Another strategy developed at around the same time focuses instead on the protein

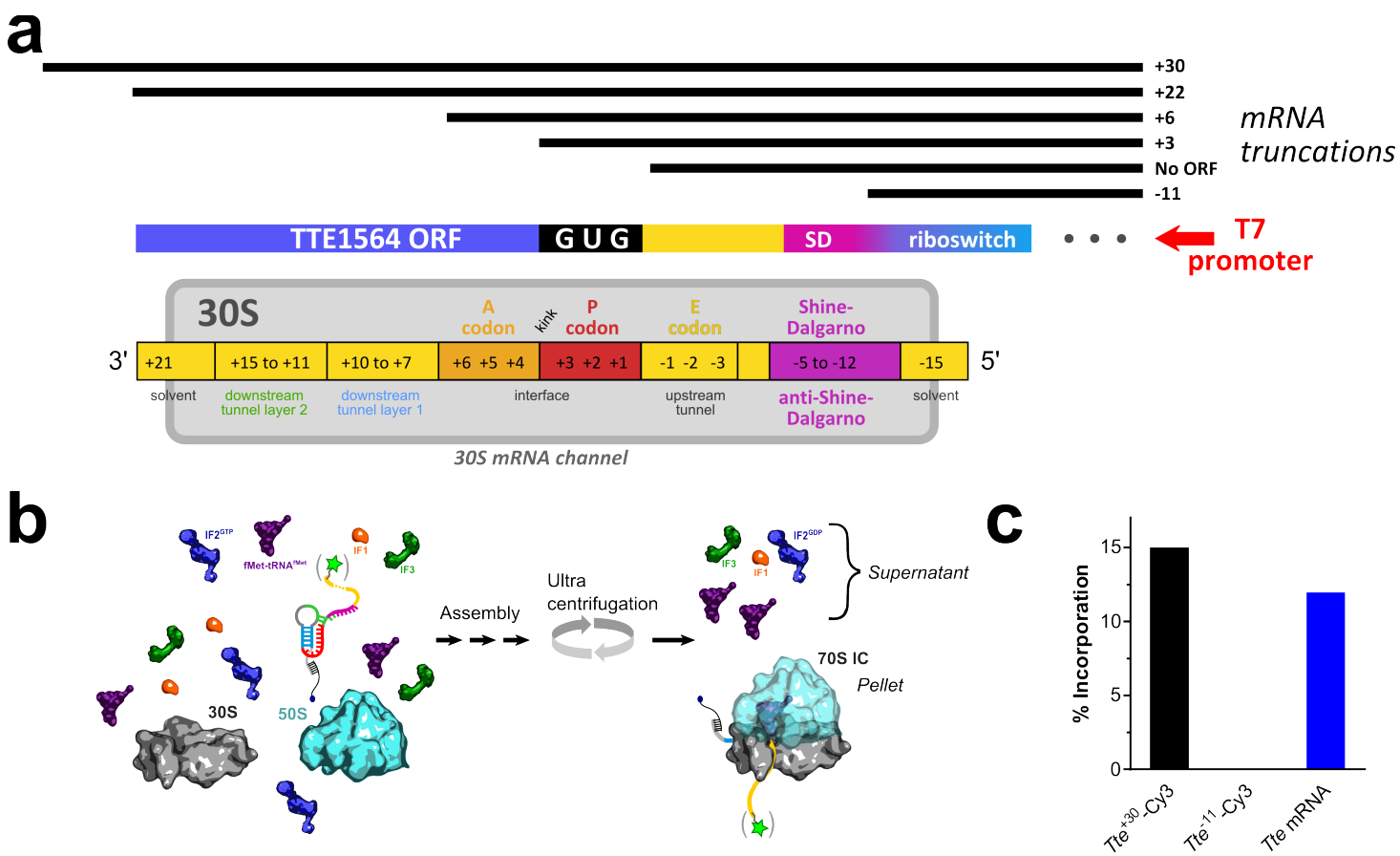


Figure 4-4 *Tte* mRNA truncations to determine minimal RNA required for 30S IC formation.

(**a**) Schematic diagram of *Tte* mRNA truncation constructs and their expected occupancy in the mRNA channel of the 30S during initiation. Positions of key features on the 30S subunit are indicated along with the (approximate) nucleotides of the mRNA that occupy that position, where +1 is the first nucleotide of the open reading frame. mRNA channel representation is adapted from Ref. 197. (**b**) Schematic of the 70S IC formation assay, where complex formation is assessed by the amount of [^{35}S]-fMet-tRNA^{fMet} found in pelleted 70S ICs relative to that present at the start of the reaction. 70S IC formation on *Tte* RNA constructs of different lengths and with or without fluorophore (green star) were compared. (**c**) Comparison of 70S IC formation efficiency on full length *Tte* mRNA and fluorophore-labeled *Tte* mRNA truncations. As expected, 70S ICs do not form on mRNA that lacks a start codon (*Tte*$^{-11}$-Cy3), whereas efficiency of initiation is similar for full length *Tte* mRNA and *Tte*$^{+30}$-Cy3.

components of the ribosome; single cysteine mutants of selected ribosomal proteins are expressed recombinant and subsequently reconstituted into 30S[198-200] and 50S subunits[201]. For the development of this assay, we employed the latter strategy, labeling ribosomal protein S6 in the 30S subunit.

The wild-type S6 has no natural cysteine residues, and so a single cysteine mutant, S6 D41C, was developed by Ermolenko *et al.*[200], and expressed and purified using established protocols[198]. For fluorophore labeling, initial attempts were unsuccessful and so notable changes to the established literature protocol were made. These are discussed briefly below and in **Materials and Methods 4.5.14**.

In the protein labeling strategy detailed by Hickerson *et al.*[198], labeling of the single cysteine residue with a fluorophore maleimide derivative takes place under native conditions in the presence of high salt, after which the protein is denatured with urea. Fluorophores often have a large non-polar surface area and so their removal can be difficult due to many hydrophobic interactions with the protein. Labeled S6 is then purified from unincorporated dye by ion exchange chromatography on an SP Sepharose column at pH 6.3. In our experience, it was necessary to change the type of ion exchange resin used, as well as the buffering salt. The original buffer used for ion exchange is prepared using Tris-HCl, pH 6.3 at 21 °C. This is an inherently poor choice of buffering salt, as the final pH is outside of the useful range for Tris (pKa = 8.07). This, however, was easily remedied by substituting Bis-tris propane HCl (pKa = 6.5), for Tris-HCl.

More importantly, S6 is only of the very few small ribosomal proteins that has a pI < 7 (pI = 4.9 as determined experimentally by Kaltschmidt[202]), and thus is expected to be *negatively*

charged at neutral, or even slightly acidic, pH. Other proteins that are also in this cohort are ribosomal proteins S1 (pI = 4.7)[28] and S2 (pI = 6.7)[202]. The theoretical pI for the S6 D41C mutant is predicted to be ~5.36 (ExPASy ProtParam, Swiss Institute of Bioinformatics), and so is also expected to be negatively charged at neutral pH, even after covalent modification with Cy3 or Cy5 maleimide (GE Healthcare PA23031, PA25031) as these fluorophores carry an additional net negative charge. As such, the original SP Sepharose cation exchange step described for work up after fluorophore labeling is inappropriate, and unsurprisingly S6 did not bind to the column. Switching to an anion exchange resin easily solved this issue, and fluorophore labeled protein of high purity was successfully obtained (**Figure 4-5a**, **b**), and subsequently incorporated into reconstituted fluorescent 30S subunits (**Figure 4-5c-e**).

4.3.7 Fluorescence co-localization at the single molecule level indicates formation of 30S IC

As a proof of concept and to establish suitable imaging conditions, initial experiments were performed in which 30S ICs were assembled *in vitro* as depicted in **Figure 4-6a**, immobilized on the slide surface, and then imaged. Reactions were made using the *Tte*$^{+30}$-Cy3 truncation RNA, pre-annealed with the biotinylated DNA capture strand, and 30S subunits reconstituted with Cy5-labeled S6 (30S-S6-Cy5). Hundreds of distinct, immobilized mRNA molecules were observed per field of view (**Figure 4-6a**), either with or without a bound ribosome, as indicated by the presence of a co-localized Cy5 signal (**Figure 4-6b**). Approximately 8% of all Cy3-labeled mRNAs were found to be co-localized with a Cy5-labeled 30S subunit in the presence of fMet-tRNA^{fMet} (**Figure 4-6c**), which is in good general agreement with the levels of 30S IC formation seen in the ensemble assays of initiation discussed above. In the presence of initiation

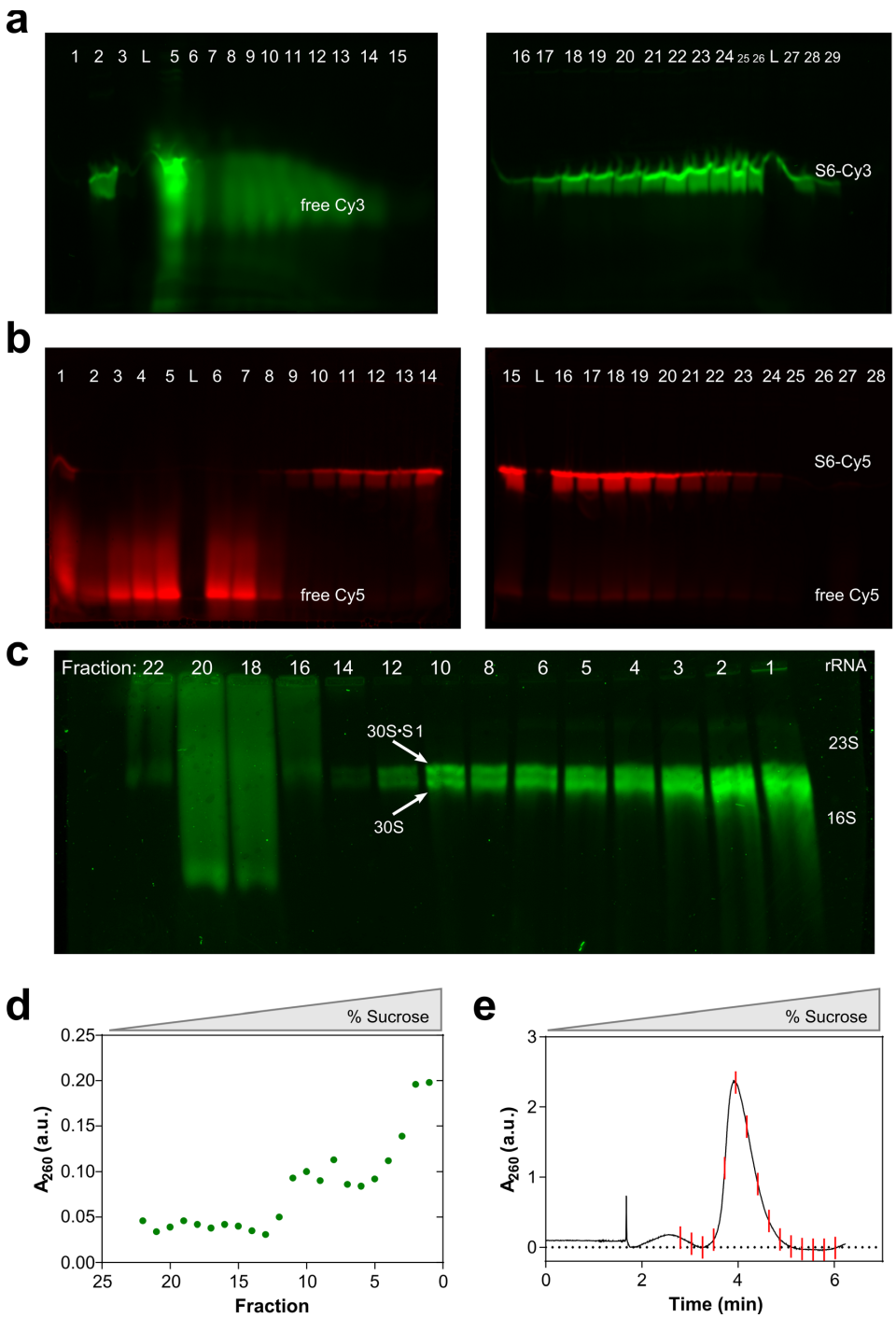


Figure 4-5 Preparation of fluorescently-labeled S6 and reconstitution into 30S subunits.

(**a**) Anion exchange under denaturing conditions effectively removes the majority of free dye present after the Cy3-maleimide after labeling. Samples were electrophoresed on 13% SDS-PAGE gels and scanned on a Typhoon 9410 variable mode imager using the appropriate excitation laser and emission filter sets. Lanes: 2) reference sample of previously purified S6-Cy3; 5) crude labeling reaction; 6-29) anion exchange

Figure 4-5 Preparation of fluorescently-labeled S6 and reconstitution into 30S subunits (continued).

gradient fractions; L, ladder. (**b**) Removal of free Cy5-maleimide is similar, but slightly less efficient than for Cy3-maleimide. Lanes: 1) crude labeling reaction; 2-5) column flow through; 6-25) anion exchange gradient fractions; 27) high-salt wash; L, ladder. (**c**) Successful reconstitution of 30S subunits with S6-Cy3 is confirmed by composite agarose-acrylamide gel electrophoresis of the sucrose gradient fractions shown in (**d**). Positions of the 23S and 16S bands in a purified rRNA reference sample were determined by post-staining with SYBR-Gold. Light fractions contain unincorporated S6-Cy3 while heavy fractions contain fluorescent 30S subunit. (**d**) Absorbance profile, measured manually by Nanodrop, for sucrose gradient fractions shown in **c**. (**e**) Sucrose gradient absorbance profile for 30S subunits reconstituted with S6-Cy5. Red marks indicate fraction boundaries.

factors, but in the absence of fMet-tRNA^{fMet}, ~2% of mRNAs were co-localized with a 30S subunit, as would be expected given that binding of mRNA by the 30S is greatly stabilized by initiator tRNA[132,194]. These results demonstrate that we are able to successfully look at the early stages of initiation at the single molecule level.

4.4 Discussion

Answering fundamental questions about how control over the 30S subunit is realized by the preQ$_1$ riboswitch requires a multi-pronged approach because the process of initiation of translation is multifactorial, involving many "moving parts" and variables including the intrinsic structural characteristics of the riboswitch itself and the structural topology of the 30S, which is altered in different ways by the binding of initiation factors and tRNA[203-205], among others.

One approach to these questions is to look at multiple variations in order to get a better picture of the underlying theme. Applied in this context, we have developed the *Bas* mRNA as another riboswitch mRNA candidate for study in addition to the already established *Tte* mRNA, with the goal that in studying both, we may better discriminate between findings that arise from species-specific features of the riboswitch (thermophilic *vs* mesophilic) from those that are

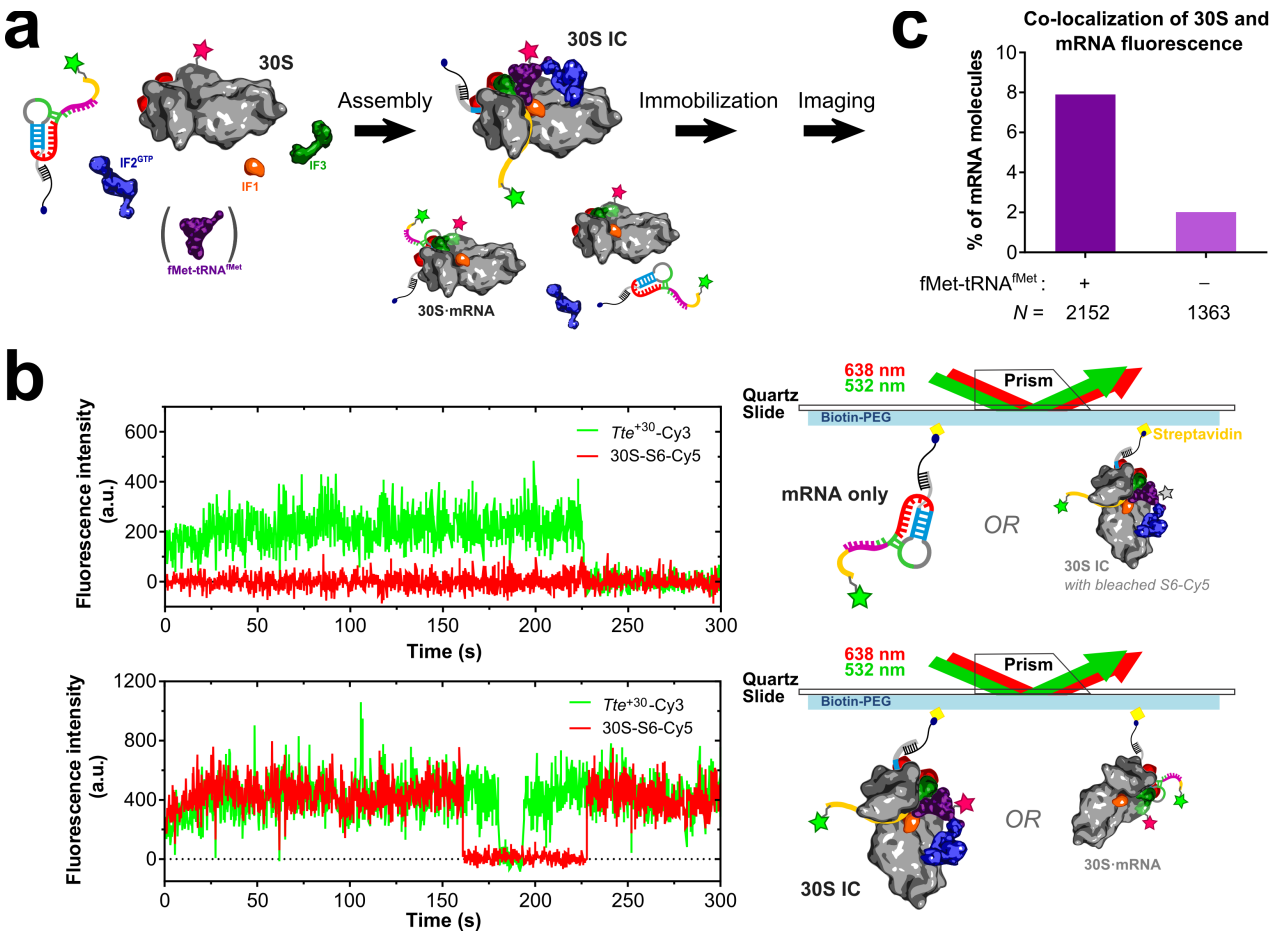


Figure 4-6 Single molecule 30S initiation complex formation assay.

(**a**) Assay overview schematic. Fluorescently labeled mRNA (*Tte*$^{+30}$-Cy3) and 30S subunits (30S-S6-Cy5) are incubated with initiation factors (IF) in the presence of absence of initiator tRNA (fMet-tRNA^{fMet}, purple), diluted to the proper concentration for immobilization on the slide surface, and then imaged. (**b**) Example single molecule fluorescence-time traces and the likely types of immobilized species each represents. The prism-TIRF microscopy setup used is depicted as in **Figure 2-4b**, with the evanescent field omitted for clarity. (**c**) Quantification of co-localization of Cy3 and Cy5 fluorescence, indicating a ribosome-bound mRNA, with and without initiator tRNA. *N* is the total number of immobilized mRNA molecules in the dataset for each condition.

common to the mechanism of regulation. The 30S IC filter binding assay in **Figure 4-2** is designed to provide information on the ability of the ribosome to initiate on various mRNAs, and does in fact demonstrate the ability of the 30S to initiate on *Bas* mRNA. As shown by SiM-KARTS in **Chapter 2**, the SD sequence of the preQ$_1$ riboswitch, at minimum in the case of *Tte* but likely also for *Bas*, remains accessible even in the presence of saturating concentrations of ligand and so it is possible that the formation of the stable 30S IC, as reported by retention of radiolabeled tRNA, is not as sensitive a readout to the effect of ligand as an earlier step of initiation would be, namely the initial binding of the mRNA by the 30S subunit. This assay can easily be modified to examine this earlier step by using radiolabeled mRNA and omitting initiator tRNA and/or initiation factors, as has been done previously in studies of other translational riboswitches[165]. Overall, our data show that the *Bas* mRNA is amenable to further study with the initiation assays and translation components we have available.

In vitro translation assays using S1-depleted subunits showed that, in the case of the *Tte* mRNA, the nature of the ribosome being used can greatly affect the outcome. It is in some ways amazing that the mesophilic, heterologous *E. coli* ribosome is capable of translating a thermophilic mRNA at 37 °C, particularly when the riboswitch fold can reasonably be estimated to have a melting temperature of greater than 60 °C in the presence of ligand. It was hypothesized that *E. coli* S1 may have played role in the translation of *Tte* mRNA, perhaps by facilitating binding and unfolding of the riboswitch, given its known requirement in the translation of other highly structured mRNA[30]. Instead, *in vitro* translation results strongly suggest that S1-depletion does not accentuate the effects of ligand on translation efficiency. Conversely, translation efficiency appears to decrease for *Bas* mRNA when translated in the presence of ligand. But, because this mRNA is translated in isolation, caution must be used when

interpreting this finding since a more robust normalization by the amount of concurrently-translated CAT is not available in the current set of experiments. The seemingly different effects that S1 has on even these two highly related mRNAs also makes it interesting to speculate about a possible role for S1 homologs in Gram-positive such as *B. subtilis*, whose expression of S1 temporally regulated, as a mechanism for altering of translational specificity of the ribosome under specific growth conditions.

What is abundantly clear is that *E. coli* S1 appears to be required for translational coupling of TTE1564 and TTE1563 when translated by the heterologous *E. coli* ribosome. Translational coupling in the expression of the two ORFs in the *Tte* mRNA is indicated by the fact that expression of both proteins decreases in the presence of $preQ_1$ when using wild-type ribosomes. By contrast, a decrease in the expression of only TTE1564 (the gene immediately downstream of the riboswitch) would be expected if translation of each ORF were completely independent of the other. Moreover, when S1-depleted ribosomes are used, expression of the distal gene TTE1563 is abolished. This finding suggests a potential, more generalized role for S1 in reinitiation on polycistronic mRNAs in *E. coli*, especially when subsequent ORFs contain less optimal combinations of start codon and SD sequence. Translation of TTE1563 begins with the non-canonical start codon UUG, which and is preceded by a weaker SD sequence than is TTE1564, and this combination is known to be much less efficiently translated for *E. coli* ribosomes, particularly when depleted of S1[26]. One possible scenario for this translational coupling is that a ribosome translating the TTE1564 ORF reaches the stop codon, terminates, and releases the newly synthesized QueT protein. That same ribosome then backtracks a short distance to initiate at the start of the TTE1563 ORF. This reinitiation event, however, is mediated by ribosome-bound S1. In the absence of ribosome-bound S1, the ribosome is unable to locate

the start of the second ORF due to its weaker SD sequence and non-canonical start codon and instead diffuses away. This could explain the observation that, while some degree of independent translation of TTE1563 is possible with *E. coli* ribosomes that contain S1, the majority arises as a result of translational coupling with TTE1564 and so in the absence of S1, that coupling is lost.

Finding such as these underscore the continuing need for sophisticated approaches to studying processes such as the binding of riboswitch-containing mRNA by the 30S, given the number of different variables (e.g. ribosome composition, presence of ligand) that could potentially come together to influence the final regulatory decision. To this end, we have established conditions for a single molecule assay to examine interactions between the mRNA and 30S with great inherent flexibility. Factors such as the presence of various IFs and tRNA, the order of addition or time of removal of ligand, the choice of fluorophore labeled species (mRNA and 30S, tRNA and mRNA, mRNA and S1 protein), etc. can be easily manipulated and thus the assay can be easily tailored to answer a variety of questions.

4.5 Materials and Methods

4.5.1 Selection and cloning of sequences encoding the Bas *mRNA*

The preQ$_1$ riboswitch from *Bacillus anthracis* was selected for study from the list of known riboswitch sequences in Rfam[185] and those previously identified by Roth *et al*.[59]. Criteria similar to those described by Eichhorn[206], namely the distance between the 3′ end of the riboswitch and the start codon of the adjacent downstream reading frame, and the presence or absence of predicted strong stem loop structures located in that region, were used to discern whether a riboswitch was likely to be translationally acting (as opposed to transcriptionally acting). Reference genomic sequences for *B. anthracis* were downloaded from the National Center for

Biotechnology Information (AE017225.1, http://www.ncbi.nlm.nih.gov). The boundaries of the putative RNA transcript were predicted using BPROM and FindTerm (SoftBerry, http://linux1.softberry.com/), which found a promoter and transcription start site upstream of the riboswitch and a Rho-independent transcription terminator immediately downstream of the BAS1509 ORF. The expected sequence of the 3′ end of the 16S rRNA from *B. anthracis*[207,208] was compared to the nucleotides 3′ of the riboswitch aptamer and 5′ of the start codon to identify the putative Shine-Dalgarno sequence. Base-pairing interactions between the 16S rRNA from *B. anthracis* and the SD sequence of the *Bas* mRNA were predicted using the RNA Fold Bimolecular function in RNAstructure v5.6. From these folding results, connectivity table files (`*.ct`) were manually constructed and the free energies of hybridization for the isolated SD•16S rRNA sequences was calculated using the Efn2 module in RNAstructure at 37 °C (310.15 K).

The DNA region encoding the complete mRNA transcript was amplified by PCR from *B. anthracis* str. Sterne (34F2) genomic DNA, generously provided as a colony prep in PCR-Lyse solution (Epicenter) by the lab of Prof. Philip C. Hanna at the University of Michigan. Sequences of the forward and reverse primers were: 5′-TGT TGC TTA AAA AAC GAA TAA CGT GG-3′, and 5′-ggt ggt <u>AAG CTT</u> TAA AAA AGA GCT ATC CCT ATA GGA ATA GC-3′, respectively. The HindIII restriction site in the reverse primer is underlined and non-complementary extensions are shown in lowercase. Only the forward primer was 5′-phosphorylated as described above in **Material and Methods 3.5.2**, after which the PCR reaction was performed, containing 2 µL of colony prep solution as the genomic template, 0.5 µM of each DNA primer, 1 U Phusion High-Fidelity DNA polymerase (New England Biolabs, M0530), 1X Phusion HF Buffer, 200 µM of each dNTP in a 50 µL final volume. The PCR

reaction conditions were: initial denaturation at 98 °C for 30 s, followed by 30 cycles of denaturation (98 °C for 10 s), annealing (70 °C for 30 s), and extension (72 °C for 18 s), followed by a final extension at 72 °C for 5 min. When the reaction was complete, 10 µL of 3M NaOAc (pH 5.2) was added to each reaction and the PCR product was purified using the QIAquick PCR purification kit (Qiagen), and the concentration measured using a Nanodrop2000 spectrophotometer. An annealed DNA duplex with a single 5′-phosphate (5′-GGT GGT TCT AGA TAA TAC GAC TCA CTA TAG GG-3′ and 5′-P-CCC TAT AGT GAG TCG TAT TAT CTA GAA CCA CC-3′) encoding the T7 promoter and an XbaI restriction site was ligated to the PCR product using T4 DNA ligase. The resulting ligation product was digested with XbaI and HindIII, ligated between the XbaI and HindIII sites in the pUC19 plasmid, and transformed into the JM109 strain of *E. coli* (Promega). Clones carrying the correct plasmid were identified by Sanger sequencing. The complete sequence and map of the resulting plasmid (pUC19c_BAS1509) is shown in **Figure C.1-1**.

4.5.2 Analysis of RNA-Seq data in support of the Bas *mRNA transcript*

RNA transcriptome profiling data from the work of Passalacqua *et al.*[186] were used to look for evidence of the *Bas mRNA* at the transcriptional level. SOLiD sequencing read data (GSE13543) were downloaded from the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/). Bowtie (http://bowtie-bio.sourceforge.net/index.shtml) was used to align sequence reads to a reference genome index was constructed from the NCBI accession numbers AE017225.1 and AE017336.2. Aligned reads files were converted from SAM to BAM format using SAMtools (http://samtools.sourceforge.net/). Finally, the coverageBed function in bedtools (http://bedtools.readthedocs.org/en/latest/) was used to determine the read coverage depth for each nucleotide in the genomic region encoding the putative *Bas* mRNA

transcript; read coverage was then summed over all growth conditions in Microsoft Excel. Example unix commands used to perform this analysis are provided in **Appendix C.3**.

4.5.3 RNA preparation

Full-length CAT, *Tte*, and *Bas* mRNAs were prepared by *in vitro* transcription with the MEGAscript T7 transcription kit (Ambion), using 4 pmol of the respective linearized plasmid DNA as the template (FspI for CAT; HindIII for *Tte* and *Bas* mRNA), in a reaction volume scaled according to the *mass* of the added DNA (20 µL per 1 µg pDNA). The transcription reactions were gel purified on denaturing 5% Urea-PAGE gels using buffer and extraction conditions similar to those described in **Materials and Methods 3.5.4**. The full sequence for the *Bas* mRNA is shown in **Appendix C.2**.

DNA templates for *in vitro* transcription of the truncated mRNA constructs were prepared by PCR using gel-purified DNA primers using conditions as described in **Materials and Methods 3.5.3**. The 3′-biotinylated DNA capture strand was similarly purified. Primer sequences and primer-specific annealing temperatures appear in **Appendix Table C.4-1**. Truncated RNAs were transcribed at 37 °C for 4 hr using the MEGAShortscript T7 transcription kit (Ambion) in a reaction volume of 60 µL, containing 150 – 200 nM PCR product as template, and on denaturing 8 or 12% Urea-PAGE gels using buffer and extraction conditions similar to those described in **Materials and Methods 3.5.4**. Complete sequences of the RNA truncations and DNA capture strand appear in **Appendix Table C.4-2**. RNA constructs prepared by transcription as described above were labeled with a Cy3 fluorophore at their 3′ end using the method described in **Materials and Methods 3.5.5**.

4.5.4 Ribosome preparation

Wild-type salt-washed ribosomes and separated ribosomal subunits were prepared from the MRE600 strain of *E. coli* as described above in **Materials and Methods 2.5.6** with only minor modifications. Preparation of 2X salt-washed separated subunits and 2X salt-washed ribosomes is exactly as described above. To prepare 1X salt-washed 30S and 50S subunits, the pelleted material after the first sucrose cushion centrifugation step was washed twice with 1 mL of storage buffer, resuspended in 2 mL of storage buffer by gentle stirring, and the dialyzed against 3 changes of buffer C to dissociate the subunits. Zonal centrifugation in a Beckman SW-28 Ti to isolate the separated subunits was performed as described.

4.5.5 Preparation of 30S subunits lacking S6

For single molecule studies, 30S subunits lacking S6 ($30S^{\Delta S6}$) were isolated from the JW4158-3 strain of *E. coli*, part of the Keio collection[209]; in this strain, the gene encoding ribosomal protein S6, *rpsF,* is replaced by a kanamycin resistance cassette. This strain was kindly provided by Prof. Dmitri Ermolenko at the University of Rochester Medical Center. Cells was grown in LB media containing 50 µg/mL of kanamycin sulfate. 2X salt-washed 30S subunits were prepared exactly as described for wild-type ribosomes.

4.5.6 Preparation of S1-depleted 30S subunits

30S subunits depleted of S1 ($30S^{\Delta S1}$) were prepared following a method adapted from that described by Lauber *et al.*[31] with buffer conditions similar to those used by Duval *et al.*[30]. Because S1 has a high affinity for polyU RNA, it can be efficiently removed from the 30S subunit by incubation with polyU resin. 125 mg of polyuridylic acid–agarose resin (Sigma, P8563) was swelled in 15 mL of polyU wash buffer (20 mM Tris-HCl [pH 7.5 at 22 °C] with 100 mM NaCl) overnight at 4 °C and then added to a clean Poly-Prep chromatography column (Bio-Rad, 7311550) that had been trimmed to 2 mL. All subsequent steps were performed at 4

°C or on ice to prevent degradation of the polyuridylic acid. The column was placed inside of a 15 mL BD Falcon tube and centrifuged in a swinging bucket rotor at 100 × *g* for 1 min. The column was then washed by adding 1 mL of polyU wash buffer and centrifuging for an additional minute. This wash step was repeated 12 times in total to extensively remove loosely bound or degraded polyU RNA from the column. The column was then equilibrated with six 1 mL washes with S1-depletion buffer (20 mM Tris-HCl [pH 7.5 at 22°C], 1 M NH_4Cl, 10 mM $MgCl_2$, 60 mM KCl, and 1 mM DTT). Approximately 1.6 nmol of wild-type 1X salt-washed 30S subunits were then applied to the column and mixed with the resin, allowed to incubate for 1 min, and then centrifuged for 1 min at 100 × *g*. The eluate was collected and passed through the column an additional 4 times. The column was then washed with four times with 500 µL of fresh S1-depletion buffer. The eluate and column washes were then pooled and diluted with an additional 1 mL of ribosome storage buffer (recovered volume ~3.3 mL), transferred to an ultracentrifuge tube (Beckman Coulter, 349622), and centrifuged at 55,000 rpm in a Beckman TLA-100.3 rotor for 14 hr to pellet the 30S subunits. The supernatant was removed by pipetting and the pellet was washed once with 300 µL of ribosome storage buffer. The pellet was then resuspended with gentle pipetting in 100 µL of ribosome storage buffer. The concentrations of the final pellet wash and the resuspended subunit solution were measured on a Nanodrop2000 spectrophotometer, using a molar extinction coefficient of 1.45×10^7 M^{-1} cm^{-1} (or alternatively, 69 pmol of 30S per A260 unit[210]). The presence of more intense protein bands (irrespective of the mRNA used) in reactions performed using $30S^{\Delta S1}$ subunits potentially speaks to the inherent difficulty in precise quantification of ribosome concentrations. Furthermore, a different extinction coefficient of 1.44×10^7 M^{-1} cm^{-1} has also been used for $30S^{\Delta S1}$ subunits (67 pmol per A260 unit[35]). In total, ~25%, i.e. 400 pmol, of the starting material was recovered. The solution

was then aliquoted, flash frozen with liquid nitrogen, and stored at -80 °C. The efficiency of depletion was in excess of 90%. 30S subunits reconstituted with S1 ($30S^{+S1}$) were prepared by incubating $30S^{\Delta S1}$ with a 2-fold excess of recombinant S1 in ribosome storage buffer on ice for 5 min[211].

4.5.7 Characterization of S1 content in 30S ribosomal subunits

The efficiency of S1-depletion and the S1-content of various 30S preparations was assessed on a composite non-denaturing 3% polyacrylamide gel containing 0.5% agarose in 1X TBE according to the method described by Dahlberg *et al.*[29,211]. The gel was pre-run at 200 V for 5 hr at 4 °C, after which the running buffer was discarded and residual buffer removed from the wells. Samples containing ~0.7 pmol of 30S subunits, of 0.4 pmol of 70S ribosomes, were mixed with an equal volume of warm (~40 °C) 0.5% agarose in 1X TBE with 0.001% bromophenol blue, immediately pipetted into the wells, and allowed to solidify in-place. The gel was run at 200 V for ~6.5 hr at 4 °C with recirculating buffer, followed by staining with SYBR-Gold nucleic acid stain, and imaged using a Typhoon 9410 variable mode imager (GE Healthcare, 488 nm laser excitation and 555 BP 20 emission filter). Gel images were quantified as described in **Appendix B.7**.

4.5.8 In vitro *translation assays using wild-type and S1-depleted subunits*

In vitro translation studies were carried out in a similar manner to that described in **Materials and Methods 2.5.7**. *In vitro* transcribed mRNAs were translated using the PURExpress® Δ Ribosome Kit (New England Biolabs). For each reaction, 3 µL of a 3.2 µM solution mRNA (3.2 µM CAT, 3.2 µM *Tte* mRNA, 3.2 µM *Bas* mRNA, or a mixture of 0.64 µM CAT and 2.56 µM *Tte* mRNA) was re-folded in the presence of 0 or 100 µM $preQ_1$ by heating to 70 °C for 2 min, followed by slow cooling to room temperature for 20 min, and then placed on ice. The remaining

components required for translation were master-mixed and aliquoted to each reaction (1.9 μCi L-[^{35}S]-cysteine or 4.7 μCi L-[^{35}S]-methionine, 4 μL PURExpress Solution A, 1.2 μL Factor Mix, and 4.8 pmol each of 2X salt-washed 50S subunits and 1X salt-washed 30S or 30S$^{\Delta S1}$ subunits), along with additional preQ$_1$ required to maintain a final concentration of 0 or 100 μM preQ$_1$ in a final reaction volume of 12 μL. Reactions were incubated at 37 °C for 2 h, frozen on dry ice, and stored at -20 °C. All following steps for sample worked up and SDS-PAGE are exactly as described in **Materials and Methods 2.5.7**. When calculating the molar of ratio of each *Tte* protein band relative to CAT, an additional correction was employed for the contribution of co-migrating CAT degradation products (i.e., subtraction of a fixed fraction of the CAT band, 0.11 × CAT for TTE1564 and 0.09 × CAT for TTE1563), which are more prominent in *in vitro* translation assays performed using radiolabeled L-methionine.

4.5.9 Preparation of aminoacylated initiator tRNA

Initiator tRNA^{fMet} from *E. coli* (Sigma, MP Biomedicals) was aminoacylated using a protocol adapted from one provided by the lab of Prof. Rachel Green at Johns Hopkins University[212]. The formyl-donor, 10-formyltetrahydrofolate, was prepared from calcium folinate pentahydrate as described previously[213]. S100 extracts were prepared from the MRE600 strain of *E. coli* using a protocol provided by the Green lab[212], which is based on established protocols[214]. A typical reaction contained 10 μM tRNA^{fMet}, 2 mM ATP, 20 μM cold L-methionine, either 2 μM each of methionyl tRNA transferase (MetRS) methionyl-tRNA^{fMet} formyltransferase (MTF) or 0.25 μg/μL S100 extracts, in buffer 207 (25 mM Tris-acetate [pH 7.5 at 22 °C], 100 mM NH_4Cl, 30 mM KCl, 10 mM $Mg(OAc)_2$, 1 mM DTT) in a final volume of 100 μL. For preparation of radiolabeled tRNA, 2.2 μM L-[^{35}S]-methionine (Perkin Elmer, NEG709A) was also included. Reaction mixtures were incubated at 37 °C for 30 min, after which the radioactive counts in 1 μL

of a 1:20 dilution of the reaction were measured by scintillation counting. 10 µL of 3 M NaOAc (pH 5.2) and 2 µL of 0.5 M EDTA (pH 8.0) were added to stop the reaction. The reaction was then extracted twice with phenol saturated w/ TE (pH 6.6) and once with chloroform saturated with TE (pH 6.5). 2.5 – 3V of cold ethanol was then added to the aqueous phase and stored overnight at -20 °C. The precipitated tRNA was collected by centrifugation at 20,800 × *g* for 45 min at 4 °C, and the fMet-tRNAf^{Met} pellet was resuspended in 25 µL of 30 mM KOAc (pH 5.2) with 1 mM DTT, aliquoted and stored at -20 °C. Radioactive counts in 1 µL of a 1:20 dilution of the resuspended tRNA solution were measured by scintillation counting. The approximate aminoacylation efficiency was calculated from the ratio of total L-methionine to tRNA^{fMet} and the measured counts before and after precipitation, with the assumption that 100% of the input tRNA is recovered after the reaction. Formylation efficiency was assessed by TLC following the method described by Walker and Frederick[215].

Plasmids for the expression of recombinant MetRS (pET21a_MetRS) and MTF (pQE16_MTF) were kindly provided by Prof. Simpson Joseph at the University of California, San Diego. Recombinant His-tagged MetRS and MTF were expressed in the BLR(DE3) strain of *E. coli* and purified using a method modified from those described previously[216,217]. Cells were grown at 37 °C with shaking in LB media with 100 µg/mL ampicillin to an OD_{600} ~0.6, induced with 1 mM IPTG for 3.5 hr, then collected by centrifugation at 8,000 × *g* for 20 min at 4 °C. All subsequent steps were performed at 4 °C or on ice. For preparation of MTF, cell pellets from 1 L of culture were resuspended in 25 mL of lysis buffer (50 mM HEPES-KOH [pH 7.6 at 22 °C], 1 M NH_4Cl, 0.1% Triton X-100, 10 mM $MgCl_2$, 0.5 mM PMSF and 7 mM β-mercaptoethanol) and lysed in a single pass through a M-110L Microfluidizer processor (Microfluidics),or by sonication (Fisher Sonic disembrator 550; level 5, microtip, 10 s process/ 30 s rest for 3 min total

process time). Lysate was then clarified by centrifugation in a Beckman Ti-70 rotor for 1 hr at 31,500 pm. The clarified lysate was incubated with 5 mL of TALON affinity resin (Clontech), which had been equilibrated with 15 mL of equilibration buffer (lysis buffer without Triton X-100 or PMSF), and tumbled for ~2.5 hr. The slurry was poured into a disposable gravity column, drained, and washed with 3 CV of wash buffer (equilibration buffer containing 10 mM imidazole). The bound protein was eluted with 3 CV of elution buffer (equilibration buffer containing 250 mM imidazole) in 4 fractions. Elution fractions containing the protein of interest were pooled and concentrated using a 10,000 MWCO Amicon Ultra-15 centrifugal filter unit (EMD Millipore) to approximately 3.7 mL. Concentrated protein was then transferred to 10,000 MWCO dialysis tubing and dialyzed against 3 changes of 1 L of storage buffer (50 mM HEPES-KOH [pH 7.6 at 22 °C], 50% [v/v] glycerol, 100 mM KCl, 10 mM $MgCl_2$, and 7 mM β-mercaptoethanol). After dialysis, the protein concentration was adjusted to 200 – 300 μM with storage buffer, aliquoted, and flash frozen with liquid nitrogen and stored at -80 °C. Purification for MetRS was identical, except that the wash buffer did not contain imidazole.

4.5.10 Filter-binding assay for 30S initiation complex formation

The principle for this assay is based on the double-filter method described with Wong and Lohman[218], with several modifications as suggested in Bevilacqua and Cech[150] and others[219], in which the binding of radiolabeled nucleic acids by proteins or other macromolecules is assessed by filtering binding reactions through a pair of stacked membranes and measuring the amount of radioactivity retained in each. 2X salt-washed 30S subunits used for all assays were activated by incubation at 37 °C in ribosome storage buffer for 50 min immediately prior to use. For each reaction, 3 μL of 3 mM mRNA was re-folded in the presence of various concentrations of $preQ_1$ by heating to 70 °C for 2 min, followed by slow cooling to room temperature for 15 min. The

remaining components required for 30S IC formation were master-mixed and aliquoted to each reaction tube containing refolded mRNA, along with additional $preQ_1$ required to maintain the desired concentration of $preQ_1$ in the final reaction volume of 15 μL. The final reaction contained 0.5 mM GTP, 0.9 μM each of IF1, IF2 and IF3, 0.6 μM [^{35}S]-fMet-tRNA^{fMet}, 0.9 μL of 10 μM 30S subunits in ribosome storage buffer (0.6 μM final), 0.6 μM mRNA, 0 – 100 μM $preQ_1$, and 1X binding buffer (20 mM HEPES-KOH [pH 7.7 at 22 °C], 70 mM NH_4Cl, 50 mm KCl, 7 mM $Mg(OAc)_2$, and 0.1 mM DTT). 30S IC reactions were incubated in a water bath at 37 °C for 12 min and then transferred to ice while the dot-blot apparatus was prepared. A serial dilution of [^{35}S]-fMet-tRNA^{fMet} in 1X binding buffer was also prepared in order to calculate the σ factor, which is used to correct for the level of non-specific nucleic acid binding[218].

Membranes were pre-wet in binding buffer without DTT for at least 30 min. Filter papers were wet briefly immediately before use. A membrane stack was constructed by stacking (from top to bottom): a reinforced nitrocellulose membrane (Optitran BA-S 85, Whatman #10-439-191), a Whatman 3MM filter paper, a positively-charged nylon membrane (BrightStar-Plus, Ambion), and a second Whatman 3MM filter paper. The membrane stack was clamped inside of 96-well dot-blot manifold (Mini-fold, Schleicher & Schuell) and vacuum was briefly applied to dry the membranes, which were then washed with cold binding buffer containing DTT (100 μL per well), and vacuum applied again. The tRNA serial dilution and binding reactions were pipetted into the wells of the manifold, drawn through the membranes under vacuum, and then washed 100 μL cold binding buffer. Vacuum was applied for 2 – 3 minutes until the membranes appeared dry and then membranes were wrapped in saran wrap and imaged using a storage phosphor screen and Typhoon 9410 Variable Mode Imager (GE Healthcare Life Sciences) and quantified using ImageQuant v 5.2 (Molecular Dynamics). Radiolabeled tRNA that is

successfully incorporated into 30S ICs is preferentially retained in the nitrocellulose filter, while unbound tRNA is trapped in the positively-charged nylon filter. The fraction of bound and unbound tRNA were calculated as described previously[218].

4.5.11 70S initiation complex formation assay

70S initiation complexes were prepared essentially as described in Zaher *et al.*[220]. Briefly, 8 µM *Tte* mRNA or fluorophore-labeled truncations (*Tte*$^{+30}$-Cy3 and *Tte*$^{-11}$-Cy3) were annealed with 16 µM DNA capture strand in a volume of 5 µL by heating to 95 °C for 45 seconds, followed by slow cooling to room temperature over 15 min, then placed on ice. The remaining components required for 70S IC formation were master-mixed and aliquoted to each reaction tube containing the annealed mRNA and capture strand in a final reaction volume of 40 µL. The final reaction contained 2 mM GTP, 3 µM each of IF1, IF2 and IF3, 3 µM [^{35}S]-fMet-tRNA^{fMet}, 1 mM $MgCl_2$, 1.5 µM 2X salt-washed 70S ribosomes, 1 µM mRNA, 2 µM capture strand, and 1X polymix buffer[221,222] (5 mM potassium phosphate [pH 7.5 at 22 °C], 5 mM $Mg(OAc)_2$, 0.5 mM $CaCl_2$, 8 mM putrescine, 1 mM spermidine trihydrochloride, 95 mM KCl, 5 mM NH_4Cl, 1 mM DTT). The total Mg^{2+} concentration in the reaction from all of the added components was ~7.5 mM. 70S IC reactions were incubated in a water bath at 37 °C for 50 min and then placed on ice and the radioactive counts in 1 µL of the reaction were measured by scintillation counting.

Successfully formed 70S ICs were purified away from unincorporated initiator tRNA and initiation factors by centrifugation through a sucrose cushion as follows: reactions were diluted with 200 µL of cold 1X polymix buffer supplemented with an additional 15 mM $MgCl_2$ to prevent complex dissociation, carefully layered onto a cold 1.3 mL sucrose cushion (1.1 M sucrose, 1X polymix buffer, 15 mM MgCl2, 0.5 mM EDTA) in an ultracentrifuge tube (Beckman Coulter, 349622), and centrifuged in a Beckman TLA-100.3 rotor at 69,000 rpm for

2.5 hr at 4 °C. The supernatant was carefully removed by pipetting and the pelleted material was resuspended by gentle pipetting in 40 μL of 1X polymix buffer. The radioactive counts in 1 μL of the resuspended material were measured by scintillation counting and the efficiency of 70S IC formation was calculated by taking the ratio of counts after and before centrifugation, accounting for the fact that initiator tRNA is present in excess over mRNA. An example protocol is included in **Appendix C.5**.

4.5.12 Expression and purification of and initiation factors

pProEx-HTb plasmids for the expression of initiation factors (IF) 1, 2, and 3 with a TEV-cleavable, N-terminal His-tag were kindly provided by Prof. Ruben L. Gonzalez, Jr.[149]. IF1, IF2 (γ isoform) and IF3 were expressed in the BLR(DE3) strain of *E. coli* and purified essentially as described by Fei *et al.*[223] with slight modification. A His-tagged version of TEV protease was also expressed and purified according to a protocol provided by the Gonzalez lab at Columbia University, New York[149].

Preparation of IF3 was as follows: cells were grown at 37 °C in 1 L of Terrific Broth with 100 μg/mL carbenicillin to an OD600 of ~0.6 – 0.8, then induced with 1 mM IPTG for 3 hr, and pelleted by centrifugation at 4,000 × *g* for 20 min at 4 °C. All subsequent steps were performed at 4 °C or on ice. Pelleted cells were washed by resuspending in 30 mL cell wash buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 60 mM NH4Cl, 10 mM $MgCl_2$, 0.1 mM PMSF, 5 mM β-mercaptoethanol), re-pelleting at 10,100 × *g* for 15 min, and then stored at -80 °C. The cell pellet was resuspended 25 mL of IF3 lysis buffer (cell wash buffer with 10 mM imidazole) and lysed by sonication (Fisher Sonic disembrator 550; level 5, microtip, 10 s process/ 30 s rest for 4 min total process time). Cell debris was pelleted by two rounds of centrifugation for 30 min at 10,400 × *g*. The clarified lysate was incubated with 5 mL of TALON affinity resin, which had been

equilibrated with 5 CV (= 25 mL) of IF3 wash buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 30 mM imidazole, 10 mM $MgCl_2$, 0.1 mM PMSF, 5 mM β-mercaptoethanol), and tumbled for ~2.5 hr. The slurry was poured into a disposable gravity column, drained, and washed with 10 CV of IF3 wash buffer. Bound protein was eluted in 10 fractions with a total of 5 CV IF3 elution buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 30 mM imidazole, 10 mM $MgCl_2$, 0.1 mM PMSF, 5 mM β-mercaptoethanol). Fractions containing IF3 were identified by SDS-PAGE on 13% Tris-glycine gels with Coomassie staining, pooled, mixed with His-tagged TEV protease at a ratio of 0.05 mg of TEV protease per mg of IF3, and placed in 10,000 MWCO dialysis tubing. TEV cleavage of the N-terminal His tag proceeds during overnight dialysis against three 1 L changes of TEV cleavage buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 200 mM NaCl, 0.1% Triton X-100, 2 mM β-mercaptoethanol).

Once the cleavage reaction had gone to completion, as judged by SDS-PAGE (~18 hr), the dialyzed solution was incubated with 5 mL of TALON affinity resin, which had been equilibrated with 5 CV of TEV protease buffer supplemented with 30 mM imidazole, and tumbled for ~1.5 hr. At this step, the His-tagged TEV protease and cleaved fragments bind the resin while cleaved IF remains in solution. The slurry was poured into a disposable gravity column and the flow through collected. The column was washed with 2 CV of TEV protease buffer supplemented with 30 mM imidazole and combined with the previous flow through. The column flow through was applied to a 5 mL SP-Sepharose Fast-Flow cation exchange column that has been equilibrated with 5 CV of IEX buffer 1 (40 mM Tris-HCl [pH 7.5 at 4 °C], 30 mM NaCl, 40 mm NH_4Cl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol). The column was then washed with 3 CV of IEX buffer 1. Bound IF3 was eluted over 20 CV using a step-wise linear gradient from 100 % IEX buffer 1 to 100% IEX buffer 2 (40 mM Tris-HCl [pH 7.5 at 4 °C], 750 mM

NaCl, 40 mm NH_4Cl, 5 mM $MgCl_2$, 2 mM β-mercaptoethanol), 5% per step. IF3 eluted between 55-65% IEX buffer 2, as determined by SDS-PAGE. Fractions containing purified IF3 were pooled and dialyzed overnight against three 1 L changes of 2X IF storage buffer (20 mM Tris-acetate [pH 7.5 at 4 °C], 100 mM KCl, 10 mM β-mercaptoethanol). After dialysis, the solution was concentrated to ~200 μM using 10,000 MWCO Amicon Ultra-15 centrifugal filter unit (EMD Millipore), and then an equal volume of autoclaved 100 % glycerol was added to the solution. The final protein concentration was determined using the Bradford assay and stored at -20 °C.

Preparation of IF1 was similar to IF3 up through the collection of flow through from the second TALON affinity column. At that stage, purity was assessed by SDS-PAGE on a 15% Tris-glycine gel. IF1 was concentrated using a 3,000 MWCO Amicon Ultra-15 centrifugal filter unit and dialyzed in 3,500 MWCO dialysis tubing against three 0.5 L changes of 2X IF storage buffer. After dialysis, the solution was concentrated to ~400 μM using 3,000 MWCO Amicon Ultra-15 centrifugal filter unit, and then an equal volume of autoclaved 100 % glycerol was added to the solution. The final protein concentration was determined using the Bradford assay and stored in aliquots at -20 °C.

Preparation of IF2-γ was as follows: cells were grown at 37 °C in 1 L of Terrific Broth with 100 μg/mL ampicillin to an OD600 of ~0.8, then induced with 1 mM IPTG for 1.5 hr, and pelleted by centrifugation at 4,000 × *g* for 20 min at 4 °C. All subsequent steps were performed at 4 °C or on ice. Pelleted cells were washed by resuspending in 30 mL IF2 cell wash buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 100 mM NaCl, 0.1 mM PMSF, 2 mM β-mercaptoethanol), re-pelleting at 4,000 × *g* for 15 min, and then stored at -80 °C. The cell pellet was resuspended 25 mL of IF2-γ lysis buffer (20 mM Tris-HCl [pH 7.5 at 4 °C], 10 mM imidazole, 300 mM NaCl,

0.2 mM PMSF, 0.88 U/mL DNase I [Takara, 2215A], 2 mM β-mercaptoethanol) and lysed by sonication (Fisher Sonic disembrator 550; level 5, microtip, 2 rounds of 10 s process/ 30 s rest for 2 min process time per round, with 15 min rest on ice between rounds). The lysate was cleared by a single round of centrifugation at 10,400 × *g*. Because IF2-γ binds poorly to the affinity resin, the cleared lysate was dialyzed against three 1 L changes of IF2-γ wash buffer without imidazole (20 mM Tris-HCl [pH 7.5 at 4 °C], 300 mM NaCl, 2 mM β-mercaptoethanol). The dialyzed solution was incubated with 7 mL of TALON affinity resin, which had been equilibrated with 5 CV of IF2-γ wash, and tumbled for ~2.5 hr. The slurry was poured into a disposable gravity column, drained, and washed with 10 CV of IF2 wash buffer. Bound protein was eluted in 6 fractions with a total of 6 CV IF2-γ elution buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 250 mM imidazole, 500 mM NaCl, 2 mM β-mercaptoethanol). The first 3 elution fractions were identified as containing IF2-γ by SDS-PAGE on 9% Tris-glycine gels with Coomassie staining. These fractions were combined and the protein concentration measured with the Bradford assay. The pooled fractions were diluted to ~65 mL (~2.2 mg/mL IF2-γ) and mixed with His-tagged TEV protease at a ratio of ~4 µg of TEV protease per mg of IF2-γ. The solution was placed in 10,000 MWCO dialysis tubing. TEV cleavage of the N-terminal His tag proceeds during overnight dialysis against three 1 L changes of TEV cleavage buffer (10 mM Tris-HCl [pH 7.5 at 4 °C], 200 mM NaCl, 0.1% Triton X-100, 2 mM β-mercaptoethanol).

After cleavage, the dialyzed solution was concentrated to < 30 mL using a 10,000 MWCO Amicon Ultra-15 centrifugal filter unit, and then incubated with 5 mL of TALON affinity resin, which had been equilibrated with 3 CV of TEV protease buffer, and tumbled for ~4 hrs. The slurry was poured into a disposable gravity column and the flow through collected. The column was washed with 3 CV of TEV protease buffer and combined with the previous flow through.

The column flow through was applied to a 5 mL SP-Sepharose Fast-Flow (GE Healthcare) cation exchange column that has been equilibrated with 10 CV of IEX buffer 1. The column was then washed with 3 CV of IEX buffer 1. Bound IF2-γ was eluted over 30 CV using a step-wise linear gradient from 100% IEX buffer 1 to 75% IEX buffer 2, 3.75% per step. IF2-γ eluted between ~20 – 38% IEX buffer 2, as determined by SDS-PAGE. Fractions containing purified IF2 were pooled and dialyzed overnight against three 1 L changes of 2X IF2-γ storage buffer (20 mM Tris-acetate [pH 7.5 at 4 °C], 100 mM KCl, 20 mM $Mg(OAc)_2$, 10 mM β-mercaptoethanol). After dialysis, the solution was concentrated to ~200 μM using 10,000 MWCO Amicon Ultra-15 centrifugal filter unit (EMD Millipore), and then an equal volume of autoclaved 100% glycerol was added to the solution. The final protein concentration was determined on a Nanodrop2000 spectrophotometer using a molar extinction coefficient 27,390 M^{-1} cm^{-1}. The final solution was aliquoted and stored at -20 °C.

4.5.13 Expression and purification of ribosomal proteins

The pET24b plasmid for the over-expression of a single-cysteine mutant (D41C) of small ribosomal protein S6 was kindly provided by Prof. Dmitri Ermolenko at the University of Rochester Medical Center. S6 D41C was expressed and purified according to the method described by Hickerson *et al.*[198]. Ribosomal protein S1 was prepared exactly as described in **Materials and Methods 3.5.6**.

4.5.14 Fluorescent labeling of ribosomal protein S6 D41C

S6 D41C was labeled according to the method described by Hickerson *et al.*[198] with several crucial modifications. Notably, the method for removal of free dye after labeling was modified to use a different type of ion exchange resin and buffer.

Briefly, S6 D41C was expressed and purified as described above, and then dialyzed against three 1 L changes of labeling buffer (80 mM HEPES-KOH [pH 7.5 at 22 °C], 1 M KCl, and 1 mM tris(2-carboxyethyl)phosphine hydrochloride). Additionally, the headspace above the dialysis container was flushed with nitrogen. The labeling reaction was scaled to largely fill a 1.7 mL microcentrifuge tube so as to minimize the amount of headspace above the solution during the incubation. The labeling reaction was performed with a ~5-fold excess of dye to S6 D41C, at a final protein concentration of 40 μM. Thus, 64.7 nmol of dialyzed S6 D41C was placed in a 1.7 mL microcentrifuge tube, diluted to a final volume of 1.62 mL with labeling buffer, and incubated at room temperature for 20 min. A single Cy3-maleimide (GE Healthcare, PA23031) dye packet containing ~324 nmol of dye was dissolved in 30 μL anhydrous DMSO at the end of the incubation period and immediately added to the protein solution. The headspace in the microcentrifuge tube was flushed with nitrogen, and the reaction was allowed to proceed with gentle agitation for ~4 hr at room temperature, protected from light. The reaction was then quenched by the addition of β-mercaptoethanol to a final concentration of 0.5% (v/v), diluted with dilution buffer (20 mM Bis-tris propane-HCl [pH 6.3 at 22 °C], 6 M urea, 6 mM β-mercaptoethanol) to a final volume of 42.5 mL to reduce the concentration of KCl to below 40 mM, and placed on ice. All subsequent steps were performed at 4 °C or on ice. An anion exchange column was prepared using 5 mL of Q-Sepharose Fast Flow resin, and equilibrated with 15 mL of IEX buffer A (dilution buffer with 20 mM KCl). The diluted labeling reaction was applied to the column and drained. Free dye and labeled protein were separated using a linear step-wise gradient from 100% IEX buffer A to 22% IEX buffer B (dilution buffer with 1 M KCl), over 10 CV, 2.2% per step. When a visibly pink band of labeled protein began to move off the column (~14% IEX buffer B), 1 mL fraction were collected. 5 μL aliquots from each fraction

were assessed for protein content by SDS-PAGE on 13% Tris-glycine gels and imaged on a Typhoon 9410 variable mode imager, using the default excitation and emission filter sets for Cy3 (532 nm excitation, 580 BP 30 emission filter). Fractions containing pure labeled protein were pooled and concentrated to a final volume of ~4.5 mL using a 10,000 MWCO Amicon Ultra-15 centrifugal filter unit that had been pre-wet with 1 mL of 5% Tween-20 for 5 min, and then rinsed with 15 mL of milliQ water. The concentrated protein solution was then dialyzed into buffer 5[198] (80 mM HEPES-KOH [pH 7.5 at 22 °C], 1 M KCl, 20 mM $MgCl_2$ 6 mM β-mercaptoethanol). After dialysis, the dialyzed solution was diluted with 10 mL of buffer 5, and concentrated to ~5 mL using a 10,000 MWCO Amicon Ultra-15 centrifugal filter unit that had been pre-wet with 1 mL of 5% Tween-20 for 5 min, and then rinsed twice with 5 mL of buffer 5. An the protein solution was again diluted with an additional 5 mL of buffer 5, concentrated to a final volume of 625 µL, aliquoted and flash frozen in liquid nitrogen and stored at -80 °C. The concentration of recovered, fluorescently labeled protein (S6-Cy3) was determined using a Nanodrop2000 spectrophotometer, using the molar extinction coefficients for the protein (ε_{280} = 11,460 M^{-1} cm^{-1}) and for Cy3 (ε_{552} = 150,000 M^{-1} cm^{-1}). Note that the protein concentration is calculated from the A_{280} after subtracting 0.08 × A_{552} to account for the contribution of Cy3 at 280 nm as described in the product manual for the dye. The labeling efficiency calculated by dividing the Cy3 concentration by the protein concentration. Approximately 36% of the input protein was recovered, with a labeling efficiency of ~80%. The same procedure was used to label S6 D41C with Cy5-maleimide (S6-Cy5), with similar results (22% recovery, 80% labeling efficiency). The molar extinction coefficient for Cy5 at 650 nm is 250,000 $M^{-1}cm^{-1}$, and 0.05 × A650 is used to correct for the contribution of Cy5 at 280 nm.

4.5.15 Fluorescent 30S subunit preparation by reconstitution with fluorophore-labeled S6

Fluorescently-labeled 30S subunits were prepared by reconstituting $30S^{\Delta S6}$ subunits purified from the JW4158-3 strain of *E. coli* (which lacks S6) with S6-Cy3 or S6-Cy5, using conditions loosely based on established protocols[198,224]. Reconstitution reactions contained 500 – 560 pmol of $30S^{\Delta S6}$ subunits and a 1.5 molar excess of S6-Cy3 or S6-Cy5 in a final volume that was scaled for the approximate pmol of 30S subunits in the reaction (2.5 µL per pmol). The reactions were supplemented with additional buffer components such that the final solution contained 80 mM HEPES-KOH (pH 7.6 at 22 °C), 330 mM KCl, 20 mM $MgCl_2$, and 6 mM β-mercaptoethanol. Reactions were assembled on ice and then incubated in a water bath at 37 °C for ~20 min protected from light, and mixed by inverting after the first 10 min of incubation. All subsequent steps were performed on ice or at 4 °C and protected from light. Two reconstitution reactions were pooled and concentrated to ~600 µL using a 50,000 MWCO Amicon Ultra-4 Centrifugal filter unit that pre-wet with 1 mL of 5% Tween-20 for 5 min, and then rinsed with milliQ water. The 300 µL of the concentrated solution was layered onto each of two 4.6 mL ~10-20% sucrose gradients, prepared by freezing and slow thawing a 15% (464 mM) sucrose solution in buffer 7 without detergent[198] (50 mM HEPES-KOH [pH 7.5 at 22 °C], 100 mM KCl, 20 mM $MgCl_2$, 6 mM β-mercaptoethanol). Reconstituted subunits were separated from fluorophore-labeled S6 by zonal centrifugation in a Beckman SW-50.1 rotor for 2 hr 20 min at 45,000 rpm. Alternatively, 750 µL of the completed reconstitution reaction was directly layered onto 2.5 mL sucrose gradient as described above, centrifuged in a Beckman TLA-100.3 rotor at 70,000 rpm for 1 hr, and fractionated manually by piercing the bottom of the centrifuge tube with a heated needle and collecting fractions by gravity. Separation achieved with this method is acceptable, though not nearly as precise due to the use of a fixed angle rotor and manual fractionation.

Gradient fractions containing 30S-S6-Cy5 or 30S-S6-Cy3 subunits were pooled and concentrated in a 100,000 MWCO Amicon Ultra-0.5 mL Centrifugal filter unit to ~140 µL. Residual sucrose was removed by buffer exchange, achieved by repeated dilution with ribosome storage buffer and concentration by centrifugation until less than 0.1% of the original buffer remained. The concentration was measured after reduction to a final volume of ~90 µL. 30-40 pmol aliquots of reconstituted 30S subunits were made in microcentrifuge tubes, flash frozen in liquid nitrogen and stored at -80 °C. Successfully incorporation of fluorophore-labeled S6 can be confirmed by running small aliquots of gradient fractions on a composite non-denaturing 3% polyacrylamide gel containing 0.5% agarose in 25 mM Tris-HCl (pH 8 at 22 °C) with 1 mM $MgCl_2$ as previously described[211], and imaged afterwards on a Typhoon 9410 variable mode imager.

4.5.16 Single molecule 30S initiation complex formation assay

Fluorescent 30S ICs were assembled in a manner similar to that described above for 70S ICs, but with modifications to adapt the assay for single molecule experiments. Fluorescently labeled 30S-S6-Cy5 subunits were activated by incubation at 37 °C in ribosome storage buffer for 10 min prior to use. *Tte*$^{+30}$-Cy3 mRNA was annealed with 2-fold excess of DNA capture strand by heating to 95 °C for 45 seconds, followed by slow cooling to room temperature over 15 min. 1 pmol of the mRNA:capture strand solution was then combined with either 30 pmol of fMet-tRNA^{fMet}, or with an equivalent volume of 30 mM KOAc (pH 5.2) with 1 mM DTT for reactions done in the absence of tRNA. The remaining components required for 30S IC formation were master-mixed and aliquoted to each reaction tube containing annealed mRNA and capture strand, in a final reaction volume of 10 µL. The final reaction contained 2 mM GTP, 3 µM each of IF1, IF2 and IF3, 3 µM fMet-tRNA^{fMet} (when present), 4.69 µL of 3.2 µM 30S-S6-Cy5 subunits in

ribosome storage buffer (1.5 μM final), 1 μM mRNA, 2 μM capture strand, and 1X polymix buffer. The total Mg^{2+} concentration in the reaction from all of the added components was ~8.5 mM. 30S IC reactions were incubated in a water bath at 37 °C for ~30 min and then placed on ice.

Quartz slides for single molecule experiments were passivated with a mixture of PEG/biotin-PEG and assembled with microfluidic channels as described in **Materials and Methods 3.5.11**. After coating with streptavidin, slides were washed with ~300 μL of 1X polymix buffer, and then further passivated with 0.5 mg/mL BSA and 10 μM of an annealed 16-mer duplex DNA (5′-CTG CGT TGT AGG CTC G-3′; 5′-CGA GCC TAC AAC GCA G-3′) in 1X polymix buffer. The 30S IC reaction was then diluted to a final mRNA concentration of 50 pM with 1X polymix buffer supplemented with an additional 100 pM capture strand, 3.5 mM $Mg(OAc)_2$, and 4 mM Trolox, and 100 μL of this diluted reaction was applied to the slide and incubated at room temperature for 10 min. The slide was then washed with 100 μL of 1X polymix buffer supplemented with 3.5 mM $Mg(OAc)_2$, followed by 200 μL of 1X polymix buffer supplemented with 3.5 mM $Mg(OAc)_2$ and an oxygen scavenging system (10 mM protocatechuic acid, 100 nM protocatechuate-3,4-dioxygenase[155], 4 mM Trolox).

Both Cy5 and Cy3 dyes were directly and simultaneously excited using 638 nm red and 532 nm green diode lasers, respectively, on a prism-type TIRF microscopy setup as described above. Emission from both fluorophores was simultaneously recorded using an intensified CCD camera (I-Pentamax, Princeton Instruments) at 200 ms time resolution using the Micro-Manager software. Fluorescence time traces were extracted from the raw movie files using IDL (Research Systems) and analyzed using Matlab (The MathWorks) scripts. Genuine single molecule traces with a combined Cy3 and Cy5 intensity of >200 intensity units and a signal to noise ratio >3

were sorted into categories according to whether Cy3 and Cy5 were both present in the trace and photo-bleached (or blinked) in a single step (indicating co-localization of ribosome and mRNA), one-step Cy3 photobleaching (single mRNA present), multi-step Cy3 photobleaching (multiple mRNAs), one-step Cy5 bleaching (single 30S-S6-Cy5), or multi-step Cy5 photobleaching (multiple 30S-S6-Cy5 subunits).

CHAPTER 5: Conclusions and Outlook

5.1 Overview

The work presented in this dissertation has focused on the complex and often unexpected ways in which tertiary RNA structure brings about a regulatory response, with a particular focus on the realm of translation, as well as deciphering the how nature has evolved to interact with such RNA structures. More specifically, a variety of more classical biochemical and biophysical techniques have been applied in combination with new and existing single molecule techniques in order to study how the translational class-I preQ$_1$ riboswitches operate in a closer facsimile of their native contexts. In addition, already established knowledge about the structural dynamics and characteristics of these riboswitches has been employed as tool to tackle the difficult problem of studying how ribosomal protein S1, a familiar but in many ways still mechanistically enigmatic RNA binding protein, interacts with structured RNA.

5.2 SiM-KARTS reveals greater nuance in the activity of the preQ$_1$ riboswitch in *Tte* mRNA

SiM-KARTS is a powerful tool with the ability to report on site specific structural dynamics of RNAs in their native sequence context. Using SiM-KARTS to investigate changes in local structure around the expression platform of the riboswitch in *Tte* mRNA, we found that opportunities for hybridization with the SD sequence occurs in bursts, and that these “bursts” of accessibility occur both in the absence and presence of ligand. The influence of ligand is

primarily to alter the lifetime of the RNA in the bursting state, as opposed to complete sequestration of the SD sequence upon ligand binding, which would presumably be accompanied by complete silencing of expression. The end result is a more temporally heterogeneous accessibility of the SD sequence as a function of ligand.

In vitro translation of the *Tte* mRNA showed an ~40% reduction in expression in the presence of saturating ligand concentrations, which is less than would be expected if one assumes the more binary model of riboswitch regulation (i.e., a true all or nothing switching response), which runs counter to our observations. While the dynamic range of translational regulation in the mRNA is more modest than one might have anticipated, again assuming the binary model of riboswitching, it is in fact not without precedent. Similar modest repression has been seen previously for other translational riboswitches, notably the SAM-III/S_{MK} box riboswitch[164] where only a ~5-fold change was observed for an *in vivo* reporter gene placed under the control of this translational riboswitch. In another example from the adenosylcobalamin (Ado-Cbl) responsive riboswitch in *btuB* mRNA, again a modest 2-3 fold decrease in ribosome binding at the mRNA start site was observed *in vitro* in the presence of saturating concentrations of Ado-Cbl (Figure 3a from Ref. 167, open circles). While this is certainly not the case for all translational riboswitches, a more nuanced response to ligand is not unique to the class-I $preQ_1$ riboswitch in *Tte* mRNA.

As alluded to in **Discussion 2.4**, the modest modulation of translation activity brought about directly through ligand binding in some riboswitches suggests that this regulation may work in concert with other forces in the cell. For example, moderate changes in the ribosome occupancy of the mRNA may result in increased susceptibility to regulation through other mechanisms such as through regulation by small RNAs (sRNAs) in conjunction with Hfq[225], or

through direct exposure of sequence features that promote mRNA decay in addition to reduced translation, as is the case for the *lysC* riboswitch[108].

Outlook

For any technique, it is important to be equally cognizant not only of its particular strengths, but also of its limitations. In the case of SiM-KARTS, careful consideration must be given to the sequence of the probe oligonucleotide (probe strand). Depending on what particular portion of a large RNA one is interested in structurally characterizing, there is a distinct possibility that the possible sequence space for the SiM-KARTS probe strand can be quite limited. As an example, one may wish to characterize the transient structure of a particular microRNA (miRNA) binding site present in the 3′ UTR of a gene that contains multiple binding sites for the same miRNA. In this instance, one might blindly choose to use a probe strand with exact complementarity to the miRNA binding site, but this is unlikely to afford the necessary specificity to inform on the local structural dynamics of the particular miRNA site of interest given the availability of other sites with identical sequence complementarity. The solution would be to choose as the target binding sequence not the 7 nucleotide seed sequence of the miRNA site itself, but rather an adjacent sequence with a sufficiently unique sequence so as to the provide sufficient specificity while still reporting on the local structure.

SiM-KARTS also lends itself to multiplexing of probes strands that are labeled with different fluorophores when interrogating a single large RNA and such a combinatorial approach may better provide the desired information. For example, if one already has good reason to expect a large scale change in the conformation of the RNA under study, probes of different colors that target sites that are expected to be differentially accessible depending on the

conformational state of the RNA may be used concurrently. Rather than a change in the duration or distribution of binding events as was employed in **Chapter 2**, one would instead observe and derive information from changes in binding of each probe (**Figure 5-1**).

Another general consideration is the timescale of structural changes over which SiM-KARTS can effectively be applied. Instrumentation constraints aside, the rates of diffusion and the kinetics of probe hybridization impose a natural upper limit on the fastest dynamics that are measurable with SiM-KARTS. For structural changes that occur on these very fast timescales that are on the order of the time required for probe hybridization, whether the act of observation itself (i.e., probe binding and dissociation) perturbs the structural dynamics of the RNA needs to also be considered. In general, the SiM-KARTS technique is better suited to larger scale changes in structure that occur on the order of milliseconds or longer, such as the rearrangement of individual domains in a tertiary structure[226]. For example, the class-II preQ$_1$ translational riboswitch from *Lactobacillus rhamnosus* also utilizes a pseudoknot structure, in this case a HL$_{out}$-type pseudoknot, to bind ligand and control accessibility of the SD sequence[60,61]. In contrast to the class-I preQ$_1$ riboswitch, in the class-II riboswitch the entirety of the SD sequence participates in formation of one of the stems of the pseudoknot (P3 stem) and there is an additional intervening stem-loop (P4 stem) that further modulates ligand binding and dissociation[227,228].

Previous single molecule studies found that large scale structural changes occur in the RNA and that these changes are dramatically slowed in a Mg^{2+} and preQ$_1$ dependent manner, with lifetimes in the various conformations on the order of hundreds of milliseconds. In the case of this riboswitch, one would also expect the SD sequence to exhibit bursts of accessibility in the

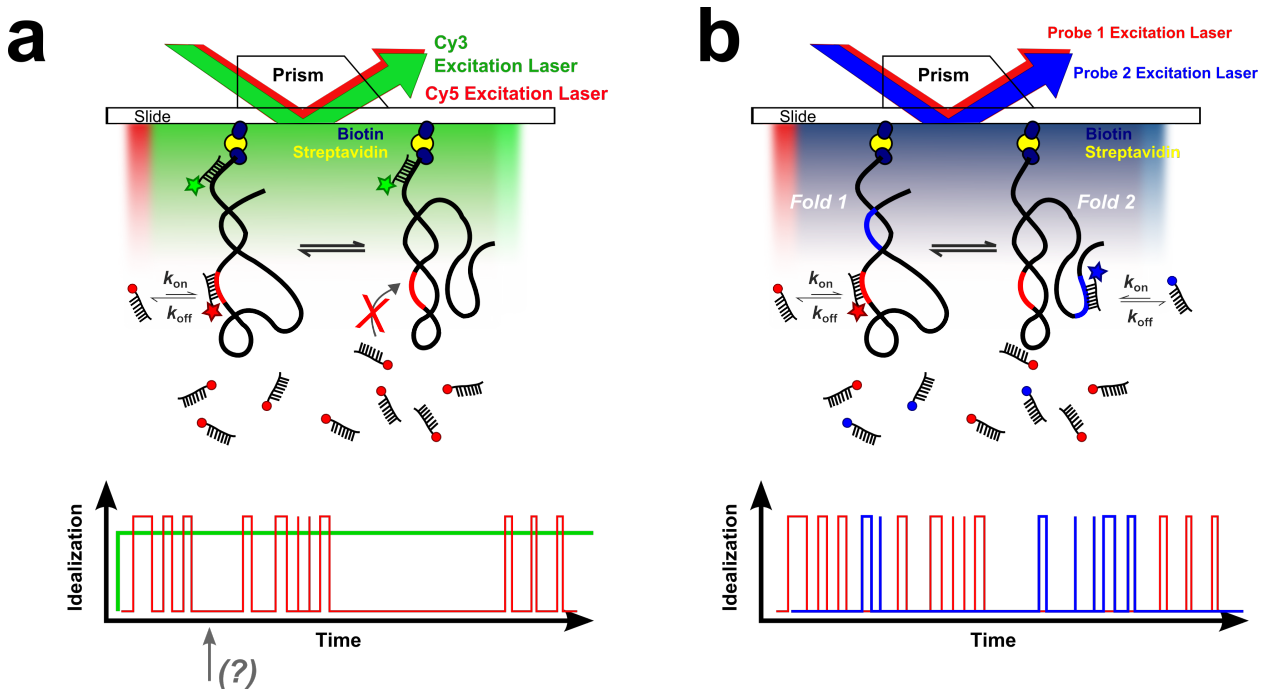


Figure 5-1 Comparison of single-probe and multiplexed SiM-KARTS.

(**a**) Schematic depiction of SiM-KARTS as it is implemented in **Chapter 2**. In some instances, it may be difficult to discern what represents a change in RNA fold (arrow and question mark), or what represents non-specific binding of the probe. Idealization of the binding events shows that the Cy3-labeled probe stays bound throughout the observation window and is used solely for the purpose of identifying the immobilized mRNA of interest. (**b**) Example schematic of multiplexed SiM-KARTS in which multiple probe strands are used simultaneously to monitor target sequences that have different, or in this case mutually exclusive, accessibility depending on the fold of the RNA. This method can potentially be used to gather more detailed data about concurrent changes in the RNA, or simply be used to address concerns over probe specificity. Note that the choice of colors for each fluorophore and excitation source are arbitrary.

absence of ligand as seen in **Chapter 2**, given that molecules exhibit transitions to a low FRET state when labeled with donor and acceptor fluorophores that report on the state of the P3 stem (WT/11-57, Ref. 228). The exact kinetic signature for the class-II riboswitch would likely be somewhat different from that of the class-I, however, as probe binding depends on the dynamics of the structure and the inherent accessibility of the intended binding site[92]. These characteristics, i.e. conformational rearrangements on a longer timescale and a more complete sequestration of the SD sequence, make the class-II preQ$_1$ riboswitch and other riboswitches that undergo similar types of dramatic rearrangements, such as the SAM-II[82,229] and SAM-III[169] riboswitches, good examples of dynamic RNA structures that are amenable to study using SiM-KARTS.

Recently, a third class of preQ$_1$ riboswitch (class-III) found in *Faecalibacterium prausnitzii* has also been described[62]. Again, this riboswitch is comprised of a complex HL$_{out}$-type pseudoknot fold, however only a portion of the SD sequence is involved in basepairing with an anti-SD sequence (as part of the P5 stem), leading to partial sequestration of the SD sequence. Interestingly, the addition of ligand increases the number of riboswitch molecules exhibiting measurable structural dynamics, and in doing so promotes the tendency for P5 to dock and thus sequester the SD sequence. This type of complex behavior suggests that SiM-KARTS may be well suited to further characterize the interplay between ligand binding and structural changes in this somewhat unusual riboswitch in the full context of its host mRNA.

One can also imagine how SiM-KARTS may be applied to the study of RNA-protein interactions, for example monitoring stepwise changes in RNA conformation and accessibility during the packaging of an RNA viral genome by viral capsid proteins. Overall, the flexibility of the SiM-KARTS technique to study virtually any large RNA, makes this an important new tool for the real-time study of structural dynamics.

5.3 *E. coli* S1 is capable of binding and unfolding pseudoknot RNA

In **Chapter 3**, the design and characterization of a series of RNA pseudoknots is described. These pseudoknots were used to begin the characterizing the interactions of S1 with a model RNA possessing a well-defined tertiary structure, using a variety of techniques. Because these pseudoknots variants are based on the one found in the class-I preQ$_1$ riboswitch, it is possible to change their stability by the simple addition of a small molecule (i.e., the riboswitch ligand, preQ$_1$). We found that S1 is able to bind to the pseudoknots with affinities that were inversely proportional to the structural stability of the pseudoknot, and that the additional stabilization that preQ$_1$ affords when bound is sufficient to prevent or displace S1. Taken together with the results of previous characterization of this riboswitch, in particular that structural dynamics of the P2 stem, it suggests a general mechanism for S1 binding of tertiary structure where a sequence for which S1 has high affinity must either be exposed or present in a reasonably dynamic structure such that S1 can bind and begin to passively unwind the RNA.

Outlook

The mechanistic study begin in **Chapter 3** has already yielded important information about the binding capacity of S1. However, this line of experimentation can likely provide additional mechanistic detail. For example, the single molecule trace data indicate that the conformational states occupied in the presence of S1 are in some instances rapidly changing, as evidence by frequent changes between FRET states for individual molecules. Analysis of these transition probabilities though HMM modeling[121] will likely provide insight. For example, such analyses would indicate whether binding and unfolding of the pseudoknot by S1 (indicated by transitions to the low FRET state) occur preferentially from the mid-FRET state, which has a less compact

fold and is more prevalent in the absence of ligand, as seems to be indicated by the S1 titration assays.

Another avenue that has not yet been fully leveraged is the fact that these pseudoknots are capable of binding a variety of ligands[59] with different affinities (**Figure 5-2**), and thus are expected to confer different degrees of stability to the WT and other pseudoknot variants. Early estimates of the ability of S1 to unwind polynucleotide secondary structure reported that its unwinding capacity was equivalent to that of heating to 50-55 °C, depending on sequence[131]. Through the use of these ligands, we can likely be more quantitative and perhaps more narrowly determine the upper limit of structural stability that S1 is able to interrogate and unfold.

Similarly, it should also be possible to more precisely identify the sequence(s) within the pseudoknot RNAs with which S1 is interacting (**Figure 5-3**) through more direct means, e.g. through the use of chemical footprinting reagents such as terbium[230] or with cleavage protection assays using S1 nuclease. Lastly, because S1 binding appears to be reversible, it may also be useful to directly determine the binding and dissociation rates of *E. coli* S1 with surface-immobilized pseudoknot RNAs by co-localization of RNA (Cy3) and non-specifically fluorophore labeled S1 (Cy5). These experiments would provide invaluable mechanistic insight into the interactions of S1 with RNA possessing tertiary structure.

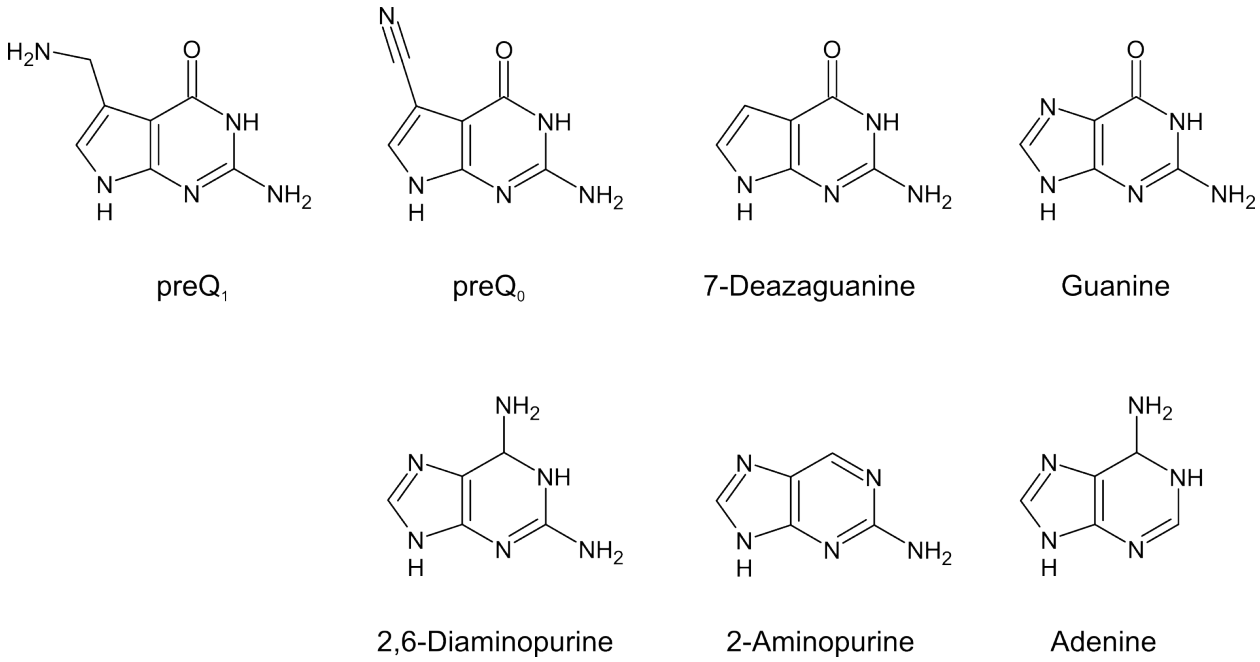


Figure 5-2 Nucleobase ligands that afford various degrees of stabilization in the preQ$_1$ riboswitch.

These nucleobase ligands, arranged in order of decreasing binding affinity from left to right, are expected to confer different degrees of stability to the pseudoknot and thus will likely be useful in more quantitatively determining the strength of S1 binding interactions.

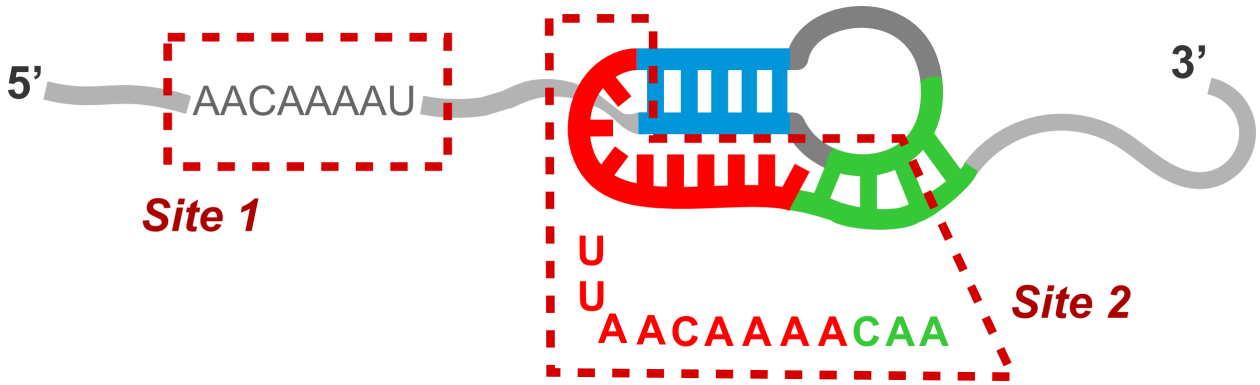


Figure 5-3 Proposed S1 binding sites in pseudoknot RNA constructs.

The formation of two apparent RNA-protein complexes in S1 titration assays suggest that two distinct binding sites exist (**Figure 3-4**). Only Complex 2 is destabilized by ligand binding indicating that this second site is located within the pseudoknot fold (Site 2). Based on the higher affinity of S1 for A/U-rich sequences, binding sites corresponding to Complexes 1 and 2, respectively, are proposed.

5.4 *In vitro* translation of preQ$_1$ riboswitch-containing mRNAs by heterologous ribosomes highlight important differences in translation machinery

It is common for *in vitro* assays to be performed using what are often considered "standard" conditions, and this often includes the use of materials from a well-studied model organism. For the majority of cases, the underlying biological principle is often well conserved and thus the results that are obtained are broadly applicable. In rarer instances, an unexpected result can bring to light fundamental differences that make a particular phenomenon unique and thus reinforce the original motivation for intellectual pursuit of a given topic. Such is arguably the case here with respect to the study of riboswitch-containing mRNAs with heterologous ribosomes.

The motivation for the *in vitro* translation studies presented in **Chapter 4** examining S1 dependence was the not-unreasonable hypothesis that S1 may be important for the translation of the thermophilic *Tte* mRNA by heterologous *E. coli* ribosomes. In work by Nou and Kadner[167], a >20-fold reduction was observed in the ability of high-salt washed ribosomes, prepared under almost identical buffer conditions as those described in **Materials and Methods 2.5.6**, to bind at the start codon of *btuB* RNA, which contains Ado-Cbl responsive riboswitch. One notable effect of high-salt washing is the displacement of S1 from the 30S subunit (compare 1X *vs* 2X salt-washed 30S in **Figure 4-3a**, **b**). One possible explanation for the observed decrease in ribosome binding is that in the process of salt-washing, a significant fraction of S1 was removed from the ribosomes, and thus the ability of these ribosomes to properly initiate on *btuB* mRNA was severely impaired due to a strict requirement for S1. In another better known example, the TIR of *rpsO* mRNA is known to possess a pseudoknotted structure in the vicinity of its SD sequence and there is a dependence on S1 for proper initiation on this mRNA[30]. In the case of translation of the *Tte* mRNA, S1 does not seem to mediate initiation at the 5′ end of the mRNA. Depletion

of S1 from the *E. coli* ribosomes did not seem to potentiate the down-regulation of translation on the *Tte* mRNA, as would be expected if S1 were instrumental in mediating translation of this thermophilic mRNA that can, in broad terms, be described as having a stable pseudoknot directly adjacent to the SD sequence.

Nevertheless, one very clear outcome of translation with S1-depleted *E. coli* ribosomes is that there is a dramatic loss of translational coupling between the upstream and downstream cistrons (**Figure 4-3d**). This implicates S1 as potentially a required cofactor in the translational coupling of adjacent genes in polycistronic mRNA, particularly when the downstream gene possess a less efficient combination of start codon and SD sequence strength. This is potentially a very interesting area of study as organization of genes into operons is one of the very basic strategies that bacteria employ to regulate the expression of related genes.

Outlook

Proteomic characterization of *T. tengcongensis* has been reported in the literature[231,232], and while evidence of TTE1563 (QueF) was found at the protein level, the same was not true for TTE1564 (QueT). This could indicate that QueT is likely expressed at very low levels or only under very specific growth conditions, as is suggested by RNA transcriptome profiling data[186] for the QueT homolog (BAS1509) in *B. anthracis*, which is also regulated by a preQ$_1$ riboswitch (**Figure 4-1**). Since RNA transcriptome profiling data has been recently made available for *T. tengcongensis* culture at a variety of temperatures[233], it may be informative to mine this sequencing data to determine whether there was any evidence in their data of *queT* gene (TTE1564) expression, and if so, build better picture of the expression profile of this gene. Translation in Gram-negative bacteria, of which *T. tengcongensis* is one, was traditionally

considered to have much stronger S1-dependence in contrast to Gram-positive bacteria[26]. However, *T. tengcongensis* shares more genetic similar to the Gram-positive species *B. halodurans*, thus raising the question of whether the putative S1 homolog is expressed in this thermophilic species, and whether it is associated with the ribosome. For this reason, it may also be informative to determine whether *rpsA* was expressed in these RNA and proteomic studies[231-233], and perhaps provide insight into what an appropriate ribosome might look like for translation of the *Tte* mRNA.

5.5 Initial conditions have been established for single molecule assays of initiation

The latter half of **Chapter 4** describes the development and motivation for single molecule assays that examine the earliest stages of initiation in which an mRNA is bound by the 30S ribosomal subunit. While initial conditions have been established providing proof of concept, there is the potential for optimization to better leverage the strengths of single molecule microscopy.

Outlook

Proof of concept experiments demonstrating that it is possible to visualize individual 30S ICs via the co-localization of fluorescence from 30S subunits and mRNA were successful, but ultimately suffer from the overall low efficiency of complex formation (maximally 8% in the presence of initiator tRNA). Additionally, it is not trivial to determine the reconstitution efficiency of 30-S6-Cy5 subunits and thus determine whether a significant percentage of "dark" ribosomes are present on the slide. A set of alternative immobilization strategies is presented in **Figure 5-4b, c**. In these strategies, an alternative labeling strategy for the ribosome is used[196] in which an extended loop engineered into h33a of the 16S rRNA is used as a handle through which to either

label the ribosome with a fluorophore labeling strand (**Figure 5-4b**), or used to directly immobilize the subunit on the slide surface[234] (**Figure 5-4c**). The scheme depicted in **Figure 5-4b** has been used by a previous member of the Walter lab, Matthew S. Marek, with some initial success. Each method has its advantages, however direct immobilization of the ribosomal subunit is slightly more appealing for several reasons related to its increased flexibility and practicality. In this method, the labeling strand is biotinylated and this moiety is used to anchor the labeling strand and annealed ribosome to the surface through a biotin-streptavidin interaction. This labeling strand can be additionally labeled with a single fluorophore at the opposite end for purposes of localization. The major advantage to this method is the relative ease with which subunits can be labeled and immobilized, as well as the option for immobilizing "dark" ribosomes. When dark ribosomes are immobilized, mRNA in solution that localizes at the surface is assumed to be interacting with immobilized 30S subunit, which can be easily verified through simple control experiments conceptually similar to those presented in **Figure A.2-2**. This format shares some similarity with SiM-KARTS in that a labeled RNA in solution (in this case the mRNA rather than the probe strand) is only detected when specifically bound at the surface through interactions mediated by the immobilized ribosome. From a practical standpoint, this immobilization scheme can both address the challenge in preparing sufficient amounts of material (e.g., ribosome), as well as the low efficiency of complex formation. In these experiments, high picomolar solutions of ribosomes are applied to the slide, incubated for a short period, and then washed, leaving only immobilized ribosomes. If complex formation efficiency is low, the surface density of ribosomes can be easily increased by increasing the ratio of biotin-PEG:PEG used to passivate the slide surface. In this way, many sites (i.e. ribosomes) for mRNA binding exist, and so even fractional interaction should provide a sufficient number of single

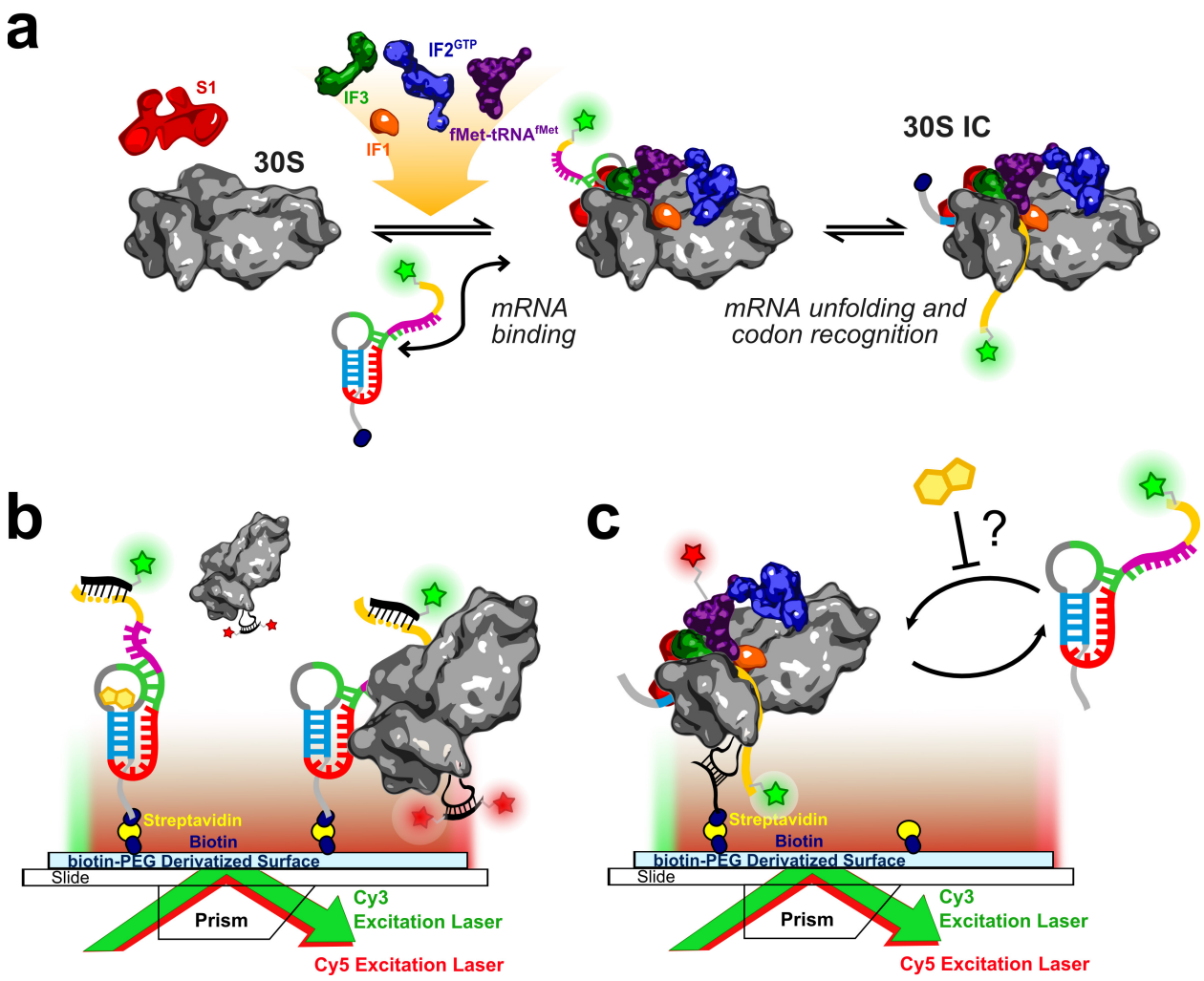


Figure 5-4 Single molecule assays of early steps of initiation.

(**a**) A highly simplified overview of the early steps of initiation whose characterization is the aim of these single molecule assays. mRNA, along with initiation factors and initiator tRNA assemble on the 30S subunit. In this process, the ribosome-bound mRNA must unfold and be correctly positioned in the mRNA channel, which is dictated in part through interactions between the SD sequence and the 16S rRNA. Correct selection of the start codon and basepairing with the bound initiator tRNA yields the 30S IC, which is then ready for joining by the 50S subunit. (**b**) Initiation assay in which mRNA is immobilized. (**c**) Initiation assay in which a dark 30S subunit is immobilized, allow for labeling of another initiation component, such as fMet-tRNA^{fMet} (red star).

molecule traces. In the reversed immobilization scheme shown in **Figure 5-4b**, solutions of labeled ribosomes with relatively high concentration likely need to be used, the preparation of which can be burdensome. The format depicted in **Figure 5-4c** has the additional advantage that this setup is easily adapted to monitor FRET between the mRNA and various other components required for initiation, such as initiator tRNA.

APPENDICES

Appendix A: Supplementary material for Rinaldi and Lund *et al.*

The figures and tables that appear in this appendix are adapted from the Supplementary Results included as part of the manuscript currently under consideration at *Nature Chemical Biology*.

A.1 Uncropped gel images from Tte *mRNA and CAT mRNA* in vitro *translation assays*

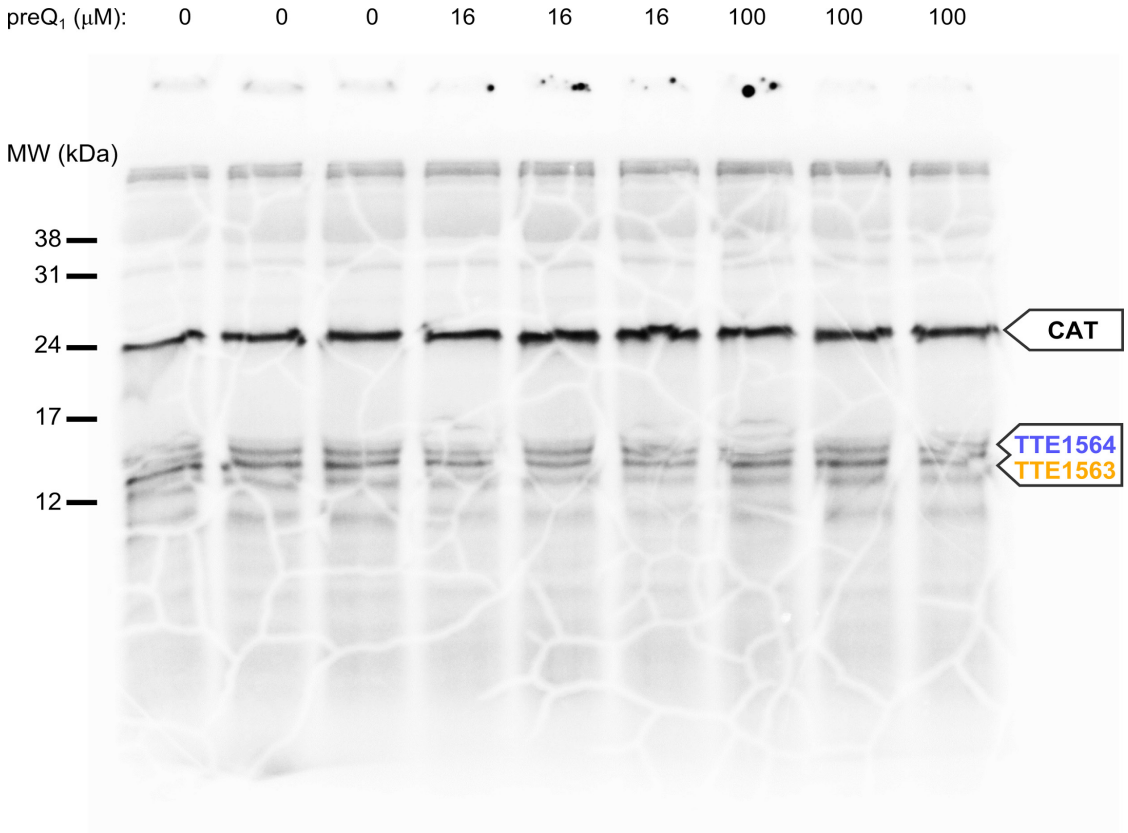


Figure A.1-1 Full *in vitro* translation gel image of *Tte*:CAT mRNA in the absence or presence of preQ$_1$

Full gel representing lanes quantified in **Figure 2-1d** and **Figure 2-3**

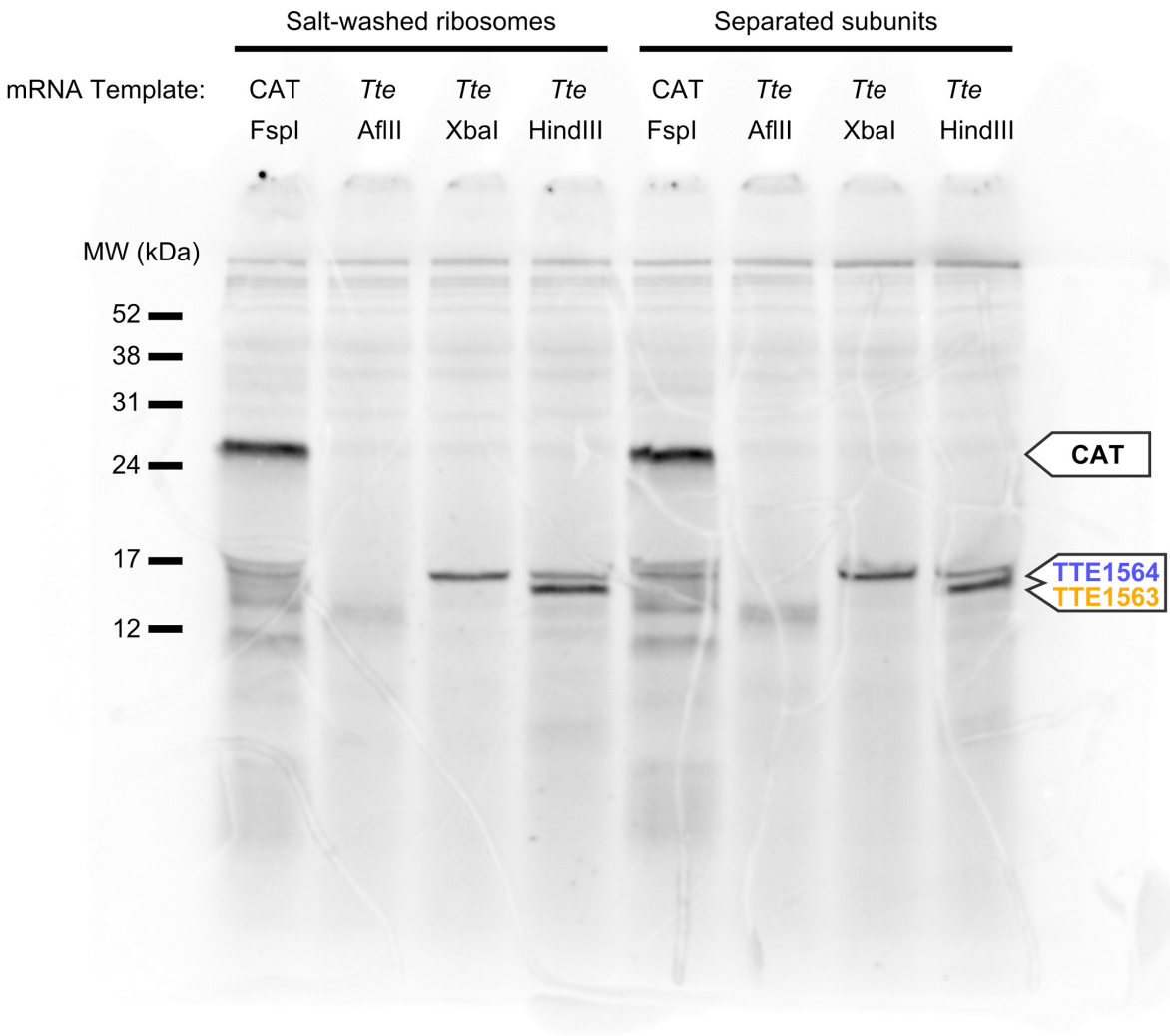


Figure A.1-2 Comparison of the performance of salt-washed ribosomes and separated subunits in *in vitro* translation experiments

Full gel for representative lanes shown in **Figure 2-2c**.

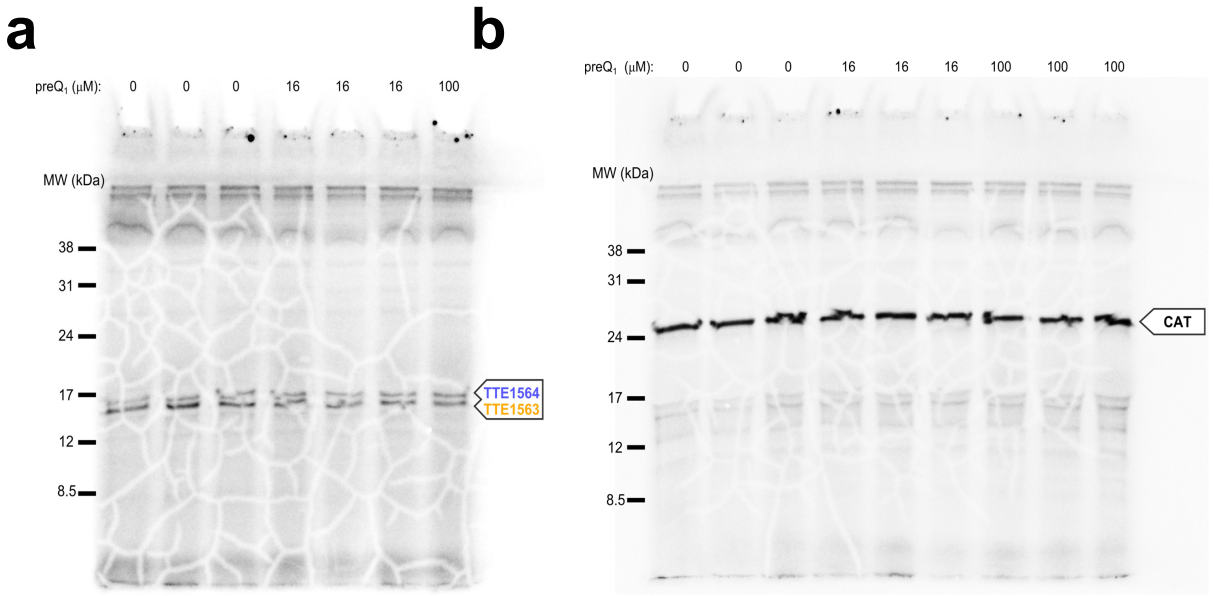


Figure A.1-3 Full *in vitro* translation gel image for unmixed *Tte* or CAT mRNA as a function of preQ$_1$ concentration

Full gel representing lanes quantified in **Figure 2-3a.**

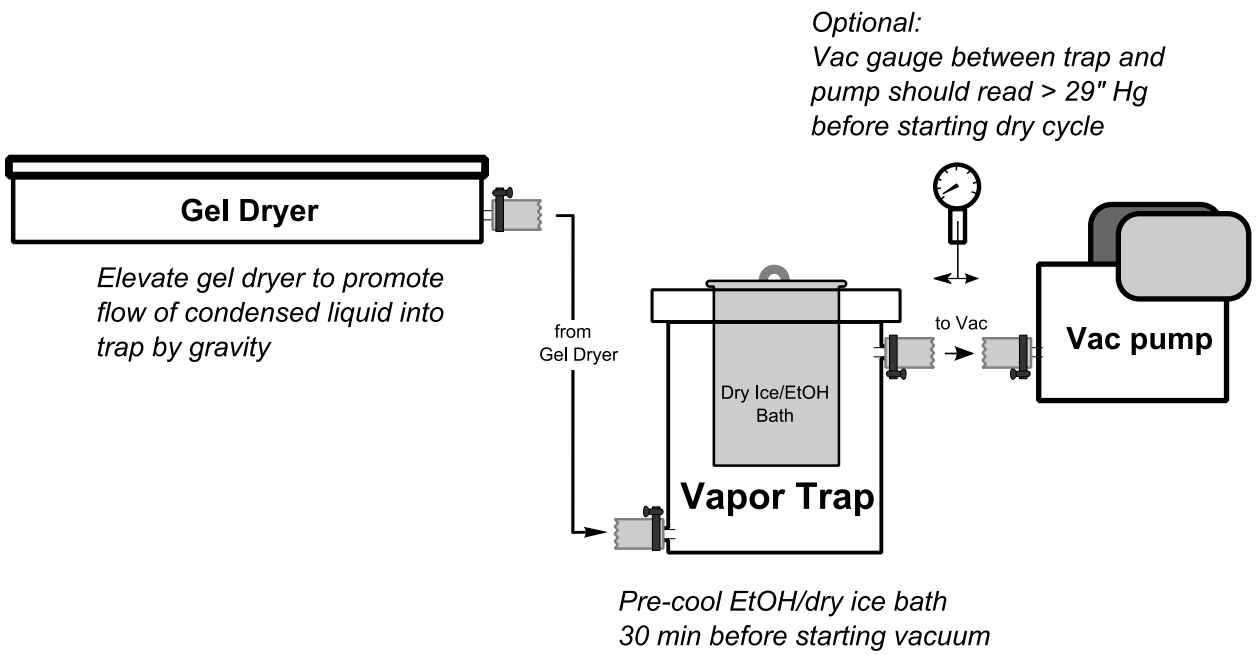


Figure A.1-4 Gel drying set up for drying ^{35}S *in vitro* translation gels

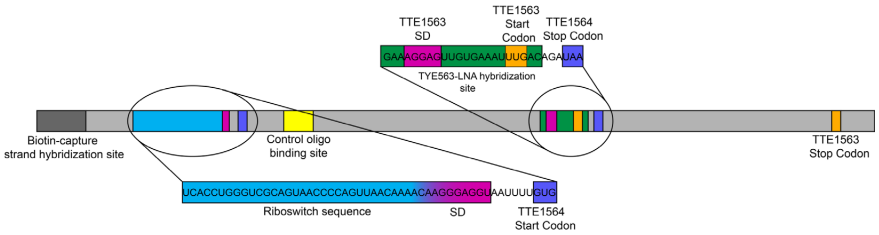


Figure A.2-1 Map of points of interest in the *Tte* mRNA used for SiM-KARTS

mRNA map describing the relative locations of important gene features and hybridization sites used in SiM-KARTS experiments. The cartoon is not drawn to scale.

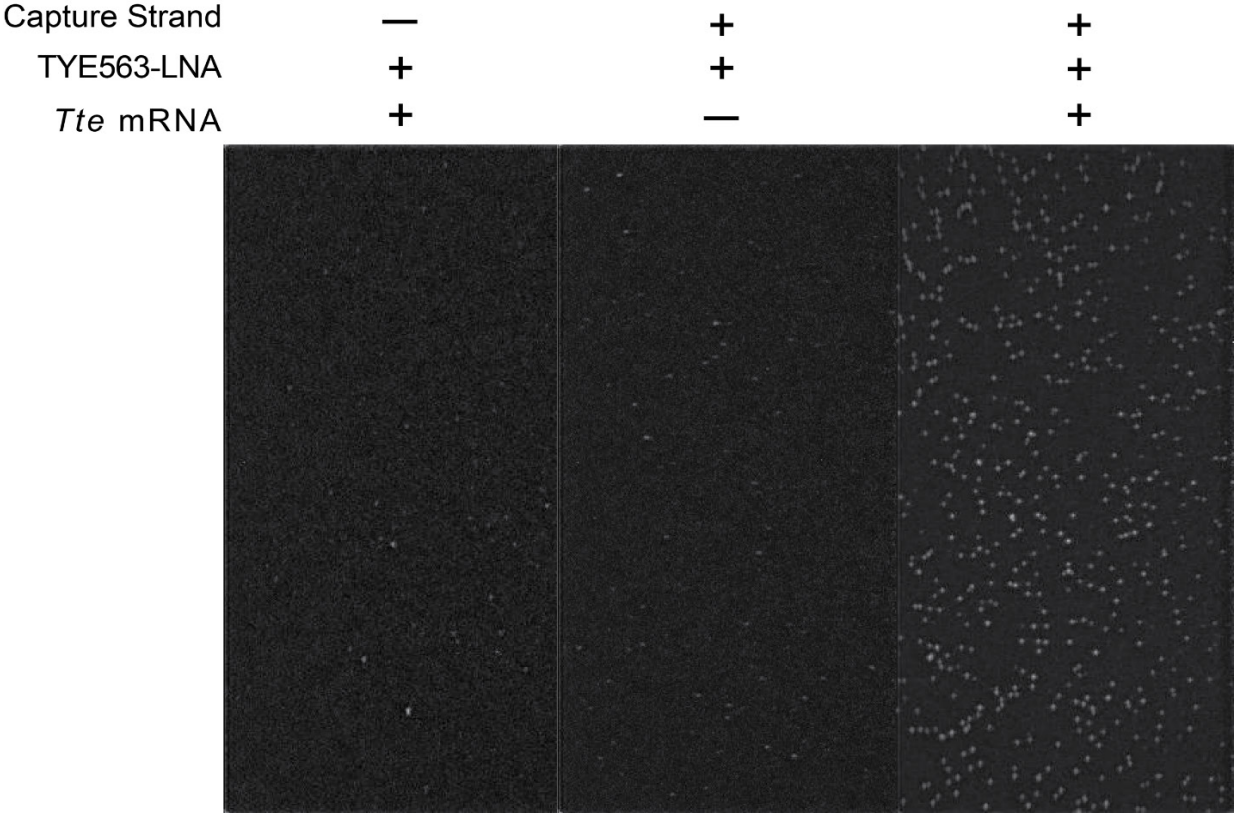


Figure A.2-2 TYE563 emission confirming specific mRNA surface immobilization during SiM-KARTS experiments

Three fields of view displaying the TYE563 emission channel using the prism-TIRFM illumination conditions. *Tte* mRNA, TYE563-LNA and biotin-capture strand were selectively annealed together in SiM-KARTS buffer. Background fluorescence is observed in the left and middle panel, where the biotin-capture strand and mRNA, respectively, were omitted from the complex. TYE563 emission was only observed when all three components were annealed together, shown in the right panel.

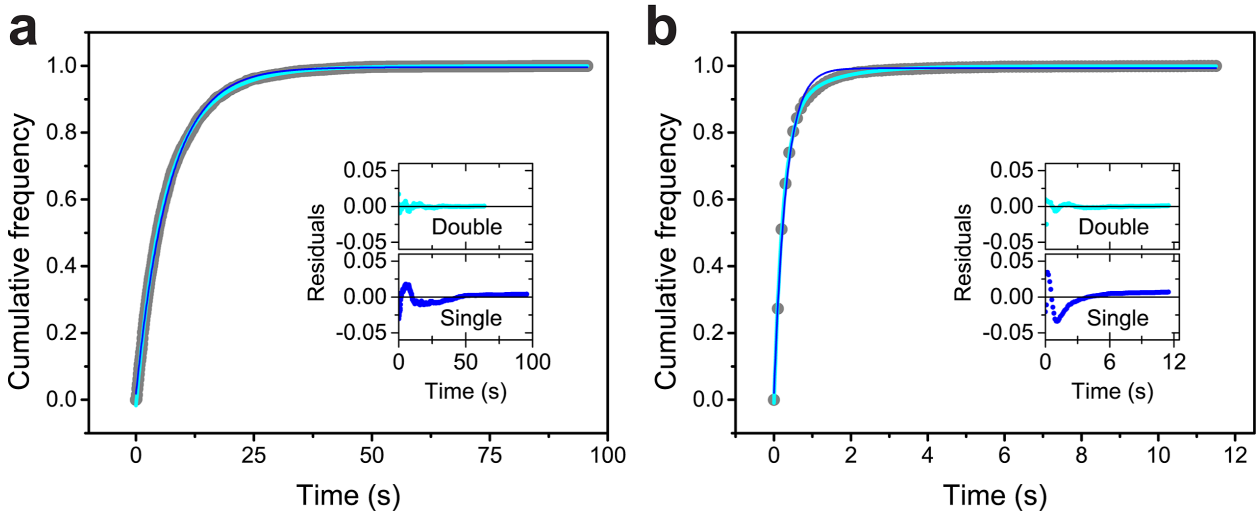


Figure A.3-1 Example kinetic data of a SiM-KARTS experiment in the absence of ligand

SiM-KARTS plot displaying the cumulative ISIs (**a**) and bound dwell times (**b**) of the anti-SD probe at zero ligand concentration. The dark blue line represents a single-exponential fit, whereas the cyan line represents a double-exponential fit for both plots along with their corresponding residuals.

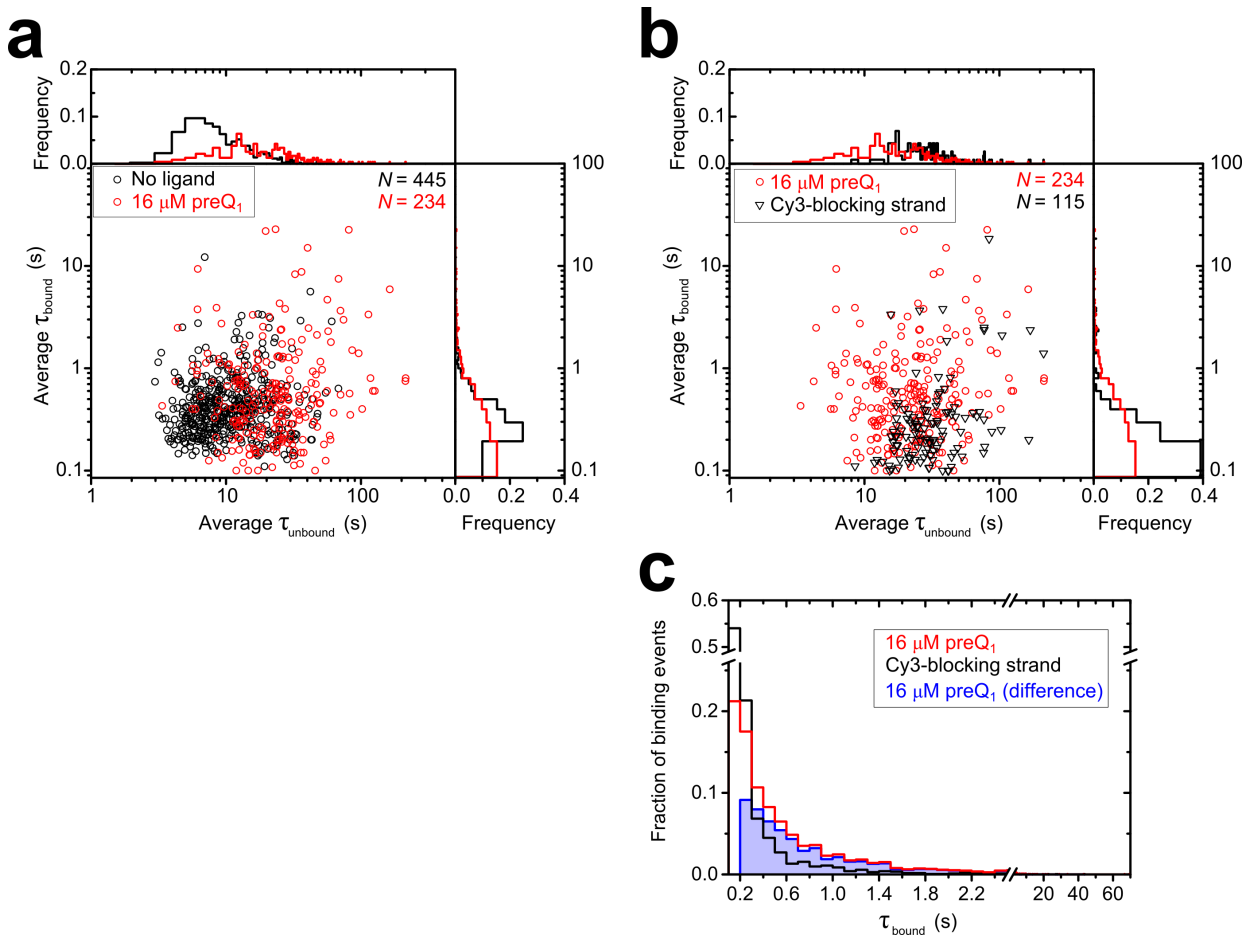


Figure A.3-2 Average bound and unbound times of the anti-SD probe for mRNA molecules in equilibrium SiM-KARTS experiments as a function of ligand or blocking strand.

(**a**) The average time each *Tte* mRNA molecule spent with anti-SD probe bound (τ_{bound}) and without anti-SD probe bound ($\tau_{unbound}$) was calculated in the absence of ligand and at 16 μM (saturating) ligand conditions. A slight shift is observed towards longer unbound times in the presence of ligand, but subpopulations within a single ligand condition were not observed. For six molecules in the 16 μM ligand condition, fewer than two binding events were observed in the fluorescence time trace and thus these molecules do not contribute data points to the plot, despite being part of the data set. (**b**) Average τ_{bound} and $\tau_{unbound}$ times for each molecule were calculated for each *Tte* mRNA in the presence of 16 μM preQ$_1$ (reproduced from **a**) and compared to the τ_{bound} and $\tau_{unbound}$ times for *Tte* mRNA whose expression platform has been blocked (reproduced from **Figure 2-10c**). A clear shift is observed towards shorter bound times. (**c**) Histogram of the bound dwell times for binding events observed in the 16 μM dataset presented in **a** (red), the blocking strand dataset from **Figure 2-10c** (black), and the 16 μM preQ$_1$ dataset after removing binding events that last only for a single frame and subtracting a proportional number of binding events according to the bound dwell time distribution observed in the blocking strand dataset, as described in **Results 2.3.11** (blue). The majority (60%) of the original binding events remain after taking this difference.

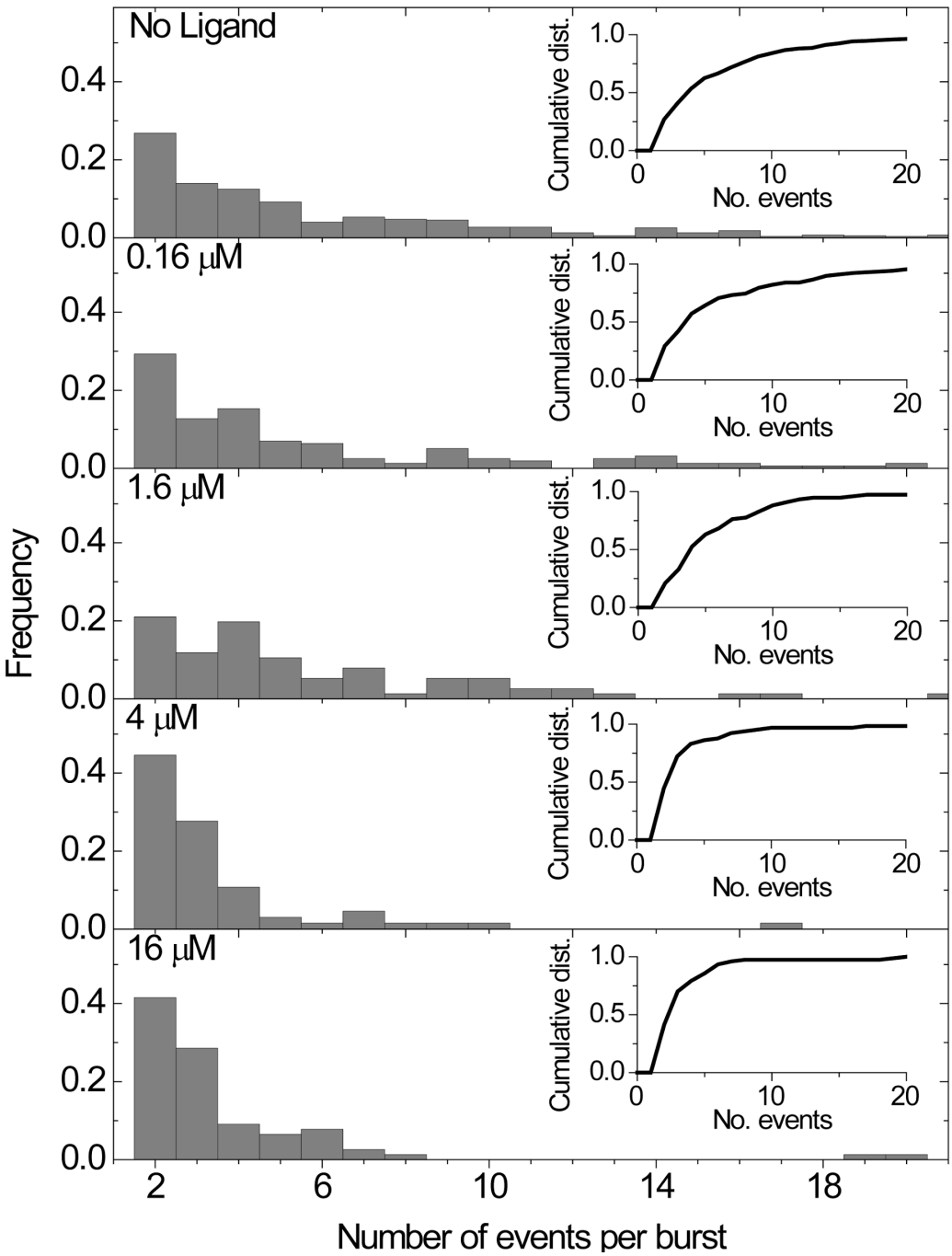


Figure A.3-3 The number of anti-SD probe binding events per burst decreases with increasing ligand concentration.

Histograms and cumulative distribution plots (inset) depicting the number of anti-SD probe binding events per burst at different ligand concentrations.

A.4 Sequences of nucleic acids used for in vitro translation and SiM-KARTS experiments with Tte mRNA and CAT mRNA

Sequences for all primers, oligonucleotides, and mRNA transcripts used in this study are presented below, written in 5′ to 3′ direction.

Primers:

FspI mutagenesis primer: GAGGGGTTTTTTGCGCAAAGGAGGAACTATATCC

Tte mRNA cloning primers (for SiM-KARTS):

Forward: GATCATGGATCCTAATACGACTCACTATAGGGGAACTCCTACTACAAGTTGCTAAGAGGC

Reverse: GATCATAAGCTTGCTTCCTCATCGTTCTCTGTAAACTC

Oligonucleotides:

DNA Capture Strand: GCCTCTTAGCAACTTGTAGTAGGAGTTCCAAAAAAAAAA-biotin

Cy3-Blocking Strand DNA: Cy3-GGCACAAAATTACCTCCCTTGTTTTGTTAACTGGG

LNA probe: TYE563-+GT+CAAATTT+CA+CAA+CT+C+CTTT+C, where a preceding "+" indicates a locked nucleic acid (LNA) base

RNA Anti-SD probe: Cy5-(aminohexyl-linker)-GAUCACCUCCUU

RNA Control probe: Cy5-(aminohexyl-linker)-GCAACAAGAGC

***Tte* mRNA for SiM-KARTS:**

GGGGAACUCCUACUACAAGUUGCUAAGAGGCUAUUUUUUAGUUCAAAUUACUCAUACAAUCAUGUU
AAAAUUAAUCGCAGUGAGCAACAAAAUGCUCACCUGGGUCGCAGUAACCCCAGUUAACAAAACAAG
GGAGGUAAUUUUGUGCCCAAAAAAAGAAUAAAAGAUUUAGCUGAAAUUGCUCUUGUUGCAGCAAUU
UAUUUCGCACUCACAAUUAUAUUUUCGUCCAUUUCGUUUUUACCCGUUCAAUUUCGAAUCGGGGAA
AUUACGAAAUCCAUUGUAGUAUUCAAUAAAAAAUAUGCUAUUUCCAUGAUGAUAGGAAAUUUUUUUG
CAAAUUUGUUUAGCCCAUUUGCUGGUGCAAUGGAAUUAAUUUUUAUGCCUCUUUCGAACUUAAUAG
GCUGUACAAUUGGAUACUACAUUGGAAGACUUACUCACAAAGCGAUAGGAGCUAUAUUCAUAGCCC
UUUGGAUUGCAGCAUCAGUUGCAAUUACUUUAAAGGUUUCUGCAGGCAUACCAUUUAUUCCGACUU
UCUUAAGCGUGGGAGUAGCGGAAACUGUACUUUUGGUAACUGGAUAUUUUUUGCUUUUCACAAUU
GAAAAGAAAGGAGUUGUGAAAUUUGACAGAUAAAUAUAAAGAGAGAAGAUUUGACAUUUACGGUUA
CGAAAAAAUUGACAAAGAAGUUCUAGAAUCUAUUGAAUAUGAGUAUCCUGAAAAAAAUACUAUCGUG

GAGUAUAUUACCGAUGAAUUUUCUUCUGUUUGCCCUUGGACAGGAUUACCUGACAAUGCAAAACUU
ACUAUAAGGUAUAUACCCCACAAAAAACUUGUAGAACUUAAAUCCUUAAAAUAUUACCUUACAUCUU
AUAGGAAUGUAGGUAUAUUGCAAGAACAUGCAAUAAACAGAAUUUUAGAUGAUUUGGUGGAAUUCC
UGCAGCCAAAAUUUAUGGAAAUAAUAGGCGAAUUUCAGGAAAGAGGAGGAAUAGCUACAAGAAUUA
UAGCAAGGUAUGAAAAAGAGGAGUAUUAAACUUAAAAGGCUGCCUAAAAUUUUGUAGGCAGCUUUU
UUAUUCAUUUUAGUUUUUCUUCAAAAUGAGUUUACAGAGAACGAUGAGGAAGCAAGCU

***Tte* mRNA for *in vitro* translation:**

GGGCAGUGAGCAACAAAAUGCUCACCUGGGUCGCAGUAACCCCAGUUAACAAAACAAGGGAGGUAA
UUUUGUGCCCAAAAAAAGAAUAAAAGAUUUAGCUGAAAUUGCUCUUGUUGCAGCAAUUUAUUUCGC
ACUCACAAUUAUAUUUUCGUCCAUUUCGUUUUUACCCGUUCAAUUUCGAAUCGGGGAAAUUACGAA
AUCCAUUGUAGUAUUCAAUAAAAAAUAUGCUAUUUCCAUGAUGAUAGGAAAUUUUUUUGCAAAUUUG
UUUAGCCCAUUUGCUGGUGCAAUGGAAUUAAUUUUUAUGCCUCUUUCGAACUUAAUAGGCUGUACA
AUUGGAUACUACAUUGGAAGACUUACUCACAAAGCGAUAGGAGCUAUAUUCAUAGCCCUUUGGAUU
GCAGCAUCAGUUGCAAUUACUUUAAAGGUUUCUGCAGGCAUACCAUUUAUUCCGACUUUCUUAAGC
GUGGGAGUAGCGGAAACUGUACUUUUGGUAACUGGAUAUUUUUUGCUUUUCACAAUUGAAAAGAAA
GGAGUUGUGAAAUUUGACAGAUAAAUAUAAAGAGAGAAGAUUUGACAUUUACGGUUACGAAAAAAU
UGACAAAGAAGUUCUAGAAUCUAUUGAAUAUGAGUAUCCUGAAAAAAAUACUAUCGUGGAGUAUAUU
ACCGAUGAAUUUUCUUCUGUUUGCCCUUGGACAGGAUUACCUGACAAUGCAAAACUUACUAUAAGG
UAUAUACCCCACAAAAAACUUGUAGAACUUAAAUCCUUAAAAUAUUACCUUACAUCUUAUAGGAAUG
UAGGUAUAUUGCAAGAACAUGCAAUAAACAGAAUUUUAGAUGAUUUGGUGGAAUUCCUGCAGCCAA
AAUUUAUGGAAAUAAUAGGCGAAUUUCAGGAAAGAGGAGGAAUAGCUACAAGAAUUAUAGCAAGGU
AUGAAAAAGAGGAGUAUUAAACUUAAAAGGCUGCCUAAAAUUUUGUAGGCAGCUUUUUUAUUCAUU
UUAGUUUUUCUUCAAAAUGAGUUUACAGAGAACGAUGAGGAAGCAAGCU

CAT control mRNA for *in vitro* translation:

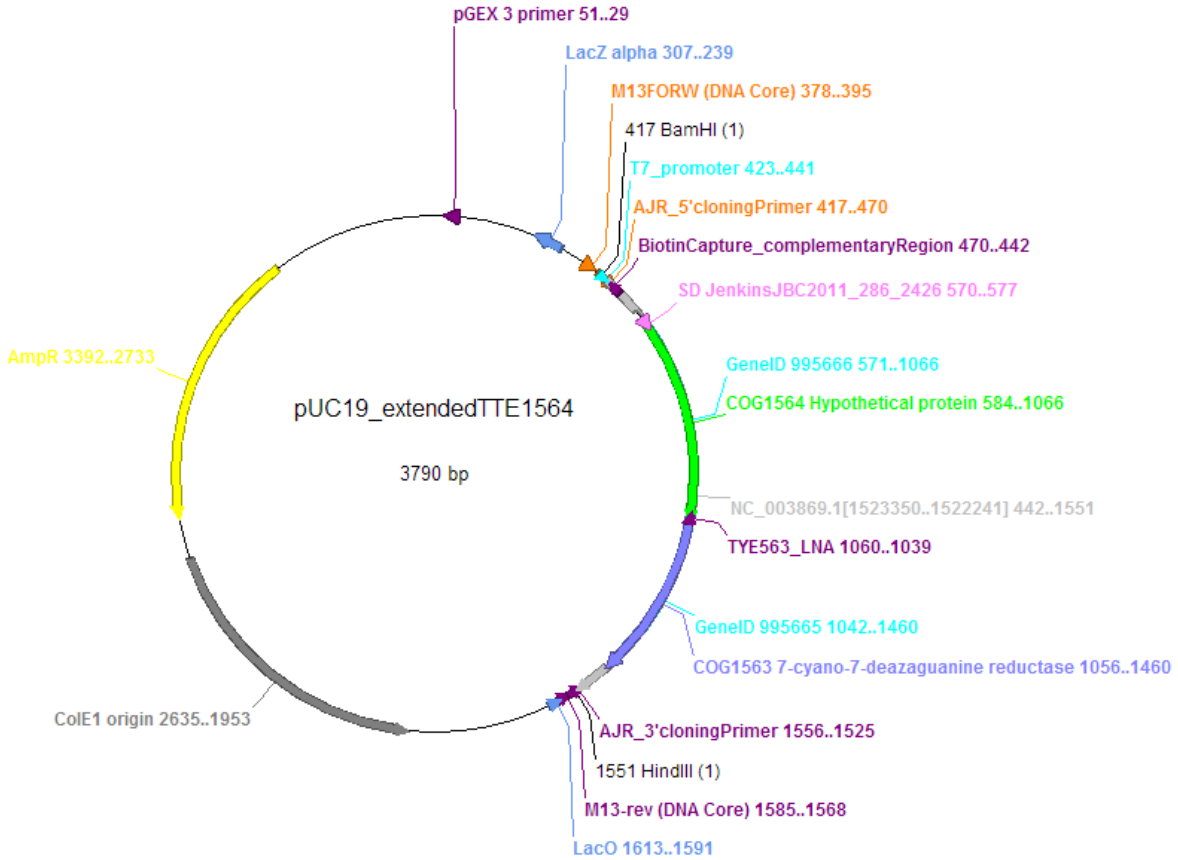
GGGAGACCACAACGGUUUCCCUCUAGAAAUAAUUUUGUUUAACUUUAAGAAGGAGAUAUACAUAUG
GAGAAAAAAAUCACUGGAUAUACCACCGUUGAUAUAUCCCAAUGGCAUCGUAAAGAACAUUUUGAG
GCAUUUCAGUCAGUUGCUCAAUGUACCUAUAACCAGACCGUUCAGCUGGAUAUUACGGCCUUUUUA
AAGACCGUAAAGAAAAAUAAGCACAAGUUUUAUCCGGCCUUUAUUCACAUUCUUGCCCGCCUGAUG
AAUGCUCAUCCGGAAUUCCGUAUGGCAAUGAAAGACGGUGAGCUGGUGAUAUGGGAUAGUGUUCA
CCCUUGUUACACCGUUUUCCAUGAGCAAACUGAAACGUUUUCAUCGCUCUGGAGUGAAUACCACGA
CGAUUUCCGGCAGUUUCUACACAUAUAUUCGCAAGAUGUGGCGUGUUACGGUGAAAACCUGGCCU
AUUUCCCUAAAGGGUUUAUUGAGAAUAUGUUUUUCGUCUCAGCCAAUCCCUGGGUGAGUUUCACCA
GUUUUGAUUUAAACGUGGCCAAUAUGGACAACUUCUUCGCCCCCGUUUUCACCAUGGGCAAAUAUU
AUACGCAAGGCGACAAGGUGCUGAUGCCGCUGGCGAUUCAGGUUCAUCAUGCCGUUUGUGAUGGC
UUCCAUGUCGGCAGAAUGCUUAAUGAAUUACAACAGUACUGCGAUGAGUGGCAGGGCGGGGCGCA
UCAUCAUCAUCAUCAUUAAGGAUCCGAAUUCGAGCUCCGUCGACAAGCUUGCGGCCGCACUCGAGC
ACCACCACCACCACCACUGAGAUCCGGCUGCUAACAAAGCCCGAAAGGAAGCUGAGUUGGCUGCUG

CCACCGCUGAGCAAUAACUAGCAUAACCCCUUGGGGCCUCUAAACGGGUCUUGAGGGGUUUUUUG
C

A.5 Plasmid maps and sequence information for plasmids used for in vitro *translation and SiM-KARTS experiments with Tte mRNA and CAT mRNA*

Plasmid sequences are in provided below in GenBank format. All plasmids used in this study have also been deposited with Addgene and are publicly available (www.addgene.com).

Figure A.5-1 pUC19_extended_TTE1564 plasmid for *in vitro* transcription of *Tte* mRNA for SiM-KARTS




```
LOCUS       pUC19_extendedTT        3790 bp ds-DNA     circular     21-OCT-2014
DEFINITION  Cloning vector pUC19c, complete sequence.
ACCESSION   L09137 X02514
VERSION     L09137.2  GI:20141090
KEYWORDS    .
SOURCE      Cloning vector pUC19c
  ORGANISM  Cloning vector pUC19c other sequences; artificial sequences;
            vectors.
REFERENCE   1  (bases 1 to 2686)
  AUTHORS   Yanisch-Perron,C., Vieira,J. and Messing,J.
  TITLE     Improved M13 phage cloning vectors and host strains: nucleotide
```

```
            sequences of the M13mp18 and pUC19 vectors
  JOURNAL   Gene 33 (1), 103-119 (1985)
   PUBMED   2985470
REFERENCE   2  (bases 1 to 2686)
  AUTHORS   Chambers,S.P., Prior,S.E., Barstow,D.A. and Minton,N.P.
  TITLE     The pMTL nic- cloning vectors. I. Improved pUC polylinker regions
            to facilitate the use of sonicated DNA for nucleotide sequencing
  JOURNAL   Gene 68 (1), 139-149 (1988)
   PUBMED   2851488
REFERENCE   3  (bases 1 to 2686)
  AUTHORS   Gilbert,W.
  TITLE     Obtained from VecBase 3.0
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1993) Department of Biochemistry, University of
            Minnesota, St. Paul, MN 55108, USA
REFERENCE   5  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Rutgers, The State University of New
            Jersey, Waksman Institute of Microbiology, 190 Frelinghuysen Road,
            Piscataway, NJ 08854-8020, USA
  REMARK    Sequence update by submitter
COMMENT     AJR_pUC19_extendedCOG1564 from 1 to 3790
COMMENT     pUC19_COG1564 from 1 to 3716
COMMENT     Alignment and annotation to NCBI reference for Tte MB4 genome -
            http://tinyurl.com/bn2w7cq
COMMENT
COMMENT     Capture Strand: GCCTCTTAGCAACTTGTAGTAGGAGTTCCAAAAAAAAAA-biotin 3?
COMMENT     LNA probe: TYE563-+GT+CAAATTT+CA+CAA+CT+C+CTTT+C
COMMENT     Anti-SD Probe: 5? Cy5-GAUCACCUCCUU 3?
COMMENT
COMMENT     On Apr 11, 2002 this sequence version replaced gi:209213. These
            data and their annotation were supplied to GenBank by Will Gilbert
            under the auspices of the GenBank Currator Program. pUC19c -
            Cloning vector (beta-galactosidase mRNA on complementary strand)
            ENTRY PUC19C                    #TYPE DNA CIRCULAR TITLE pUC19c -
            Cloning vector (beta-galactosidase mRNA on complementary strand)
            DATE 03-FEB-1986    #sequence  16-DEC-1986 ACCESSION VB0033 SOURCE
            artificial COLLECTION  ATCC 37254 REFERENCE #number    #authors
            Norrander J., Kempe T., Messing J.    #journal Gene (1983) 26:
            101-106 REFERENCE    #number  1    #authors Yanisch-Perron C.,
            Vieira J., Messing J.    #journal Gene (1985) 33: 103-119 #comment
            shows the complete compiled sequence REFERENCE #number 2 #authors
            Chambers,S.P., et al.    #journal Gene (1988) 68: 139-149
            #describes mutation at nt1308 and its effect on copy number
            REFERENCE    #number    #authors Pouwels P.H., Enger-Valk B.E.,
            Brammar W.J.    #book    Cloning Vectors, Elsvier 1985 and
            supplements    #comment  vector I-A-iv-20 COMMENT    This Sequence
            was obtained 3-MAR-1986 from J. Messing, Waksman Institute, NJ on
            floppy disk.    Revised 16-DEC-1986 by F. Pfeiffer:    1062/3 'AT'
            to 'TA' to match revised sequence of PBR322    The
            beta-galactosidase mRNA sequence including the multiple    cloning
            site of M13mp19 is on the strand complementary to that    shown.
            KEYWORDS CROSSREFERENCE    #complement VecBase(3):pUC19 #prerevised
            GenBank(50):M11662, EMBL(11):ARPuc19    #parent VecBase(3):pUC13,
            VecBase(3):M13mp19, VecSource(3):bGal19 PARENT Features of pUC19c
            (2686 bp) residue     source       1- 137 2074-2210 pBR322     138-
            237 2252-2351 pBR322     238- 395 1461-1304 (c) Lac-Operon     396-
            452 57-   1 (c) polylinker of M13mp19     455- 682  1298-1071 (c)
            Lac-Operon     683-2686 2352-4355 pBR322    Conflict (cfl) and
```

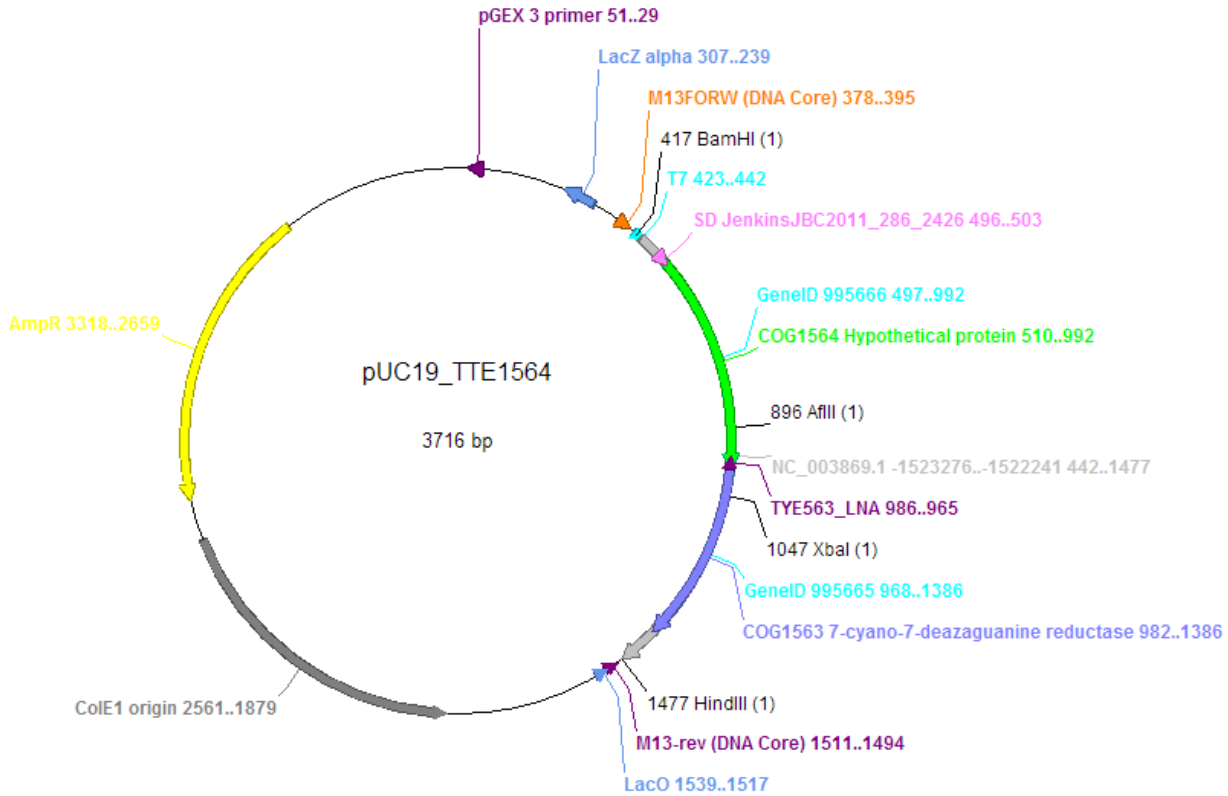
```
            Mutations (mut): pUC19c  source    mut 1308  A  G  2977   pBR322
            linked to increased copy number    mut 1942  A  G  3611   pBR322
            mut 2243 T  C  3912   pBR322 FEATURE    1629-2417 789-1 (c) Ap-R;
            b-lactamase POLYLINKER
            HindIII-SphI-PstI-SalI-XbaI-BamHI-SmaI-KpnI-SacI-EcoRI SELECTION
            #resistance  Ap    #indicator beta-galactosidase SUMMARY  pUC19c
            #length 2686   #checksum 4465.
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     primer_bind     complement(1568..1585)
                     /label=M13-rev (DNA Core)
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     complement(1525..1556)
                     /label=AJR_3'cloningPrimer
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     rep_origin      complement(1953..2635)
                     /label=ColE1 origin
                     /ApEinfo_fwdcolor=gray50
                     /ApEinfo_revcolor=gray50
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     417..470
                     /label=AJR_5'cloningPrimer
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     CDS             complement(239..307)
                     /label=LacZ alpha
                     /ApEinfo_fwdcolor=#6495ed
                     /ApEinfo_revcolor=#6495ed
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_binding    complement(1591..1613)
                     /label=LacO
                     /ApEinfo_fwdcolor=#6495ed
                     /ApEinfo_revcolor=#6495ed
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    join(442..534,1461..1551)
                     /label=NC_003869.1[1523350..1522241]
                     /ApEinfo_fwdcolor=#c0c0c0
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     CDS             complement(2733..3392)
                     /label=AmpR
                     /ApEinfo_fwdcolor=yellow
                     /ApEinfo_revcolor=yellow
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     complement(29..51)
                     /label=pGEX 3 primer
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
```

```
                 width 5 offset 0
primer_bind      complement(442..470)
                 /label=BiotinCapture_complementaryRegion
                 /ApEinfo_fwdcolor=#ff8000
                 /ApEinfo_revcolor=#800080
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
gene             1042..1460
                 /label=GeneID 995665
                 /ApEinfo_fwdcolor=cyan
                 /ApEinfo_revcolor=green
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
CDS              1056..1460
                 /label=COG1563 7-cyano-7-deazaguanine reductase
                 /ApEinfo_fwdcolor=#8080ff
                 /ApEinfo_revcolor=#ffff00
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
gene             571..1066
                 /label=GeneID 995666
                 /ApEinfo_fwdcolor=cyan
                 /ApEinfo_revcolor=green
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
CDS              584..1066
                 /locus_tag="TTE1564"
                 /note="Best Blastp hit =
                 gi|12723715|gb|AAK04893.1|AE006312_13 (AE006312)
                 HYPOTHETICAL PROTEIN [Lactococcus lactis subsp. lactis],
                 score 71.6, E-value 6.00E-12"
                 /codon_start=1
                 /transl_table=11
                 /product="hypothetical protein"
                 /protein_id="NP_623164.1"
                 /db_xref="GI:20807993"
                 /db_xref="GeneID:995666"
                 /label=COG1564 Hypothetical protein
                 /ApEinfo_fwdcolor=#00ff00
                 /ApEinfo_revcolor=#ffff00
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      378..395
                 /label=M13FORW (DNA Core)
                 /ApEinfo_fwdcolor=#ff8000
                 /ApEinfo_revcolor=#800080
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
misc_feature     complement(1039..1060)
                 /label=TYE563_LNA
                 /ApEinfo_fwdcolor=#ff8000
                 /ApEinfo_revcolor=#800080
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
RBS              570..577
                 /label=SD JenkinsJBC2011_286_2426
                 /ApEinfo_fwdcolor=#ff80ff
                 /ApEinfo_revcolor=green
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
promoter         423..441
                 /label=T7_promoter
                 /ApEinfo_fwdcolor=cyan
```

```
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
ORIGIN
        1 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
       61 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
      121 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
      181 accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc
      241 attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat
      301 tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt
      361 tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggt acccggggat
      421 cctaatacga ctcactatag gggaactcct actacaagtt gctaagaggc tatttttag
      481 ttcaaattac tcatacaatc atgttaaaat taatcgcagt gagcaacaaa atgctcacct
      541 gggtcgcagt aaccccagtt aacaaaacaa gggaggtaat tttGTGCCCA AAAAAAGAAT
      601 AAAAGATTTA GCTGAAATTG CTCTTGTTGC AGCAATTTAT TTCGCACTCA CAATTATATT
      661 TTCGTCCATT TCGTTTTTAC CCGTTCAATT TCGAATCGGG GAAATTACGA AATCCATTGT
      721 AGTATTCAAT AAAAAATATG CTATTTCCAT GATGATAGGA AATTTTTTTG CAAATTTGTT
      781 TAGCCCATTT GCTGGTGCAA TGGAATTAAT TTTTATGCCT CTTTCGAACT TAATAGGCTG
      841 TACAATTGGA TACTACATTG GAAGACTTAC TCACAAAGCG ATAGGAGCTA TATTCATAGC
      901 CCTTTGGATT GCAGCATCAG TTGCAATTAC TTTAAAGGTT TCTGCAGGCA TACCATTTAT
      961 TCCGACTTTC TTAAGCGTGG GAGTAGCGGA AACTGTACTT TTGGTAACTG GATATTTTTT
     1021 GCTTTTCACA ATTGAAAAGA AAGGAGTTGT GAAATTTGAC AGATAAATAT AAAGAGAGAA
     1081 GATTTGACAT TTACGGTTAC GAAAAAATTG ACAAAGAAGT TCTAGAATCT ATTGAATATG
     1141 AGTATCCTGA AAAAAATACT ATCGTGGAGT ATATTACCGA TGAATTTTCT TCTGTTTGCC
     1201 CTTGGACAGG ATTACCTGAC AATGCAAAAC TTACTATAAG GTATATACCC CACAAAAAAC
     1261 TTGTAGAACT TAAATCCTTA AAATATTACC TTACATCTTA TAGGAATGTA GGTATATTGC
     1321 AAGAACATGC AATAAACAGA ATTTTAGATG ATTTGGTGGA ATTCCTGCAG CCAAAATTTA
     1381 TGGAAATAAT AGGCGAATTT CAGGAAAGAG GAGGAATAGC TACAAGAATT ATAGCAAGGT
     1441 ATGAAAAAGA GGAGTATTAA acttaaaagg ctgcctaaaa ttttgtaggc agctttttta
     1501 ttcattttag tttttcttca aaatgagttt acagagaacg atgaggaagc aagcttggcg
     1561 taatcatggt catagctgtt tcctgtgtga aattgttatc cgctcacaat tccacacaac
     1621 atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag ctaactcaca
     1681 ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg ccagctgcat
     1741 taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ttccgcttcc
     1801 tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca
     1861 aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca
     1921 aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcgtt tttccatagg
     1981 ctccgccccc ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg
     2041 acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtgcg ctctcctgtt
     2101 ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcgggaag cgtggcgctt
     2161 tctcatagct cacgctgtag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc
     2221 tgtgtgcacg aaccccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt
     2281 gagtccaacc cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt
     2341 agcagagcga ggtatgtagg cggtgctaca gagttcttga agtggtggcc taactacggc
     2401 tacactagaa gaacagtatt tggtatctgc gctctgctga agccagttac cttcggaaaa
     2461 agagttggta gctcttgatc cggcaaacaa accaccgctg gtagcggtgg tttttttgtt
     2521 tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag aagatccttt gatcttttct
     2581 acggggtctg acgctcagtg gaacgaaaac tcacgttaag ggattttggt catgagatta
     2641 tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa
     2701 agtatatatg agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc
     2761 tcagcgatct gtctatttcg ttcatccata gttgcctgac tccccgtcgt gtagataact
     2821 acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcg agacccacgc
     2881 tcaccggctc cagatttatc agcaataaac cagccagccg gaagggccga gcgcagaagt
     2941 ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta
     3001 agtagttcgc cagttaatag tttgcgcaac gttgttgcca ttgctacagg catcgtggtg
     3061 tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgatc aaggcgagtt
     3121 acatgatccc ccatgttgtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgtc
     3181 agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt
     3241 actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc
     3301 tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg ggataataacc
     3361 gcgccacata gcagaacttt aaaagtgctc atcattggaa aacgttcttc ggggcgaaaa
     3421 ctctcaagga tcttaccgct gttgagatcc agttcgatgt aacccactcg tgcacccaac
     3481 tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa
```

```
     3541 aatgccgcaa aaaagggaat aagggcgaca cggaaatgtt gaatactcat actcttcctt
     3601 tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa
     3661 tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccacct
     3721 gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg
     3781 ccctttcgtc
//
```

Figure A.5-2 pUC19_TTE1564 plasmid for *in vitro* transcription of *Tte* mRNA for *in vitro* translation




```
LOCUS       pUC19_TTE1564           3716 bp ds-DNA     circular     21-OCT-2014
DEFINITION  Cloning vector pUC19c, complete sequence.
ACCESSION   L09137 X02514
VERSION     L09137.2  GI:20141090
KEYWORDS    .
SOURCE      Cloning vector pUC19c
  ORGANISM  Cloning vector pUC19c other sequences; artificial sequences;
            vectors.
REFERENCE   1  (bases 1 to 2686)
  AUTHORS   Yanisch-Perron,C., Vieira,J. and Messing,J.
  TITLE     Improved M13 phage cloning vectors and host strains: nucleotide
            sequences of the M13mp18 and pUC19 vectors
  JOURNAL   Gene 33 (1), 103-119 (1985)
   PUBMED   2985470
REFERENCE   2  (bases 1 to 2686)
  AUTHORS   Chambers,S.P., Prior,S.E., Barstow,D.A. and Minton,N.P.
  TITLE     The pMTL nic- cloning vectors. I. Improved pUC polylinker regions
            to facilitate the use of sonicated DNA for nucleotide sequencing
  JOURNAL   Gene 68 (1), 139-149 (1988)
   PUBMED   2851488
REFERENCE   3  (bases 1 to 2686)
```

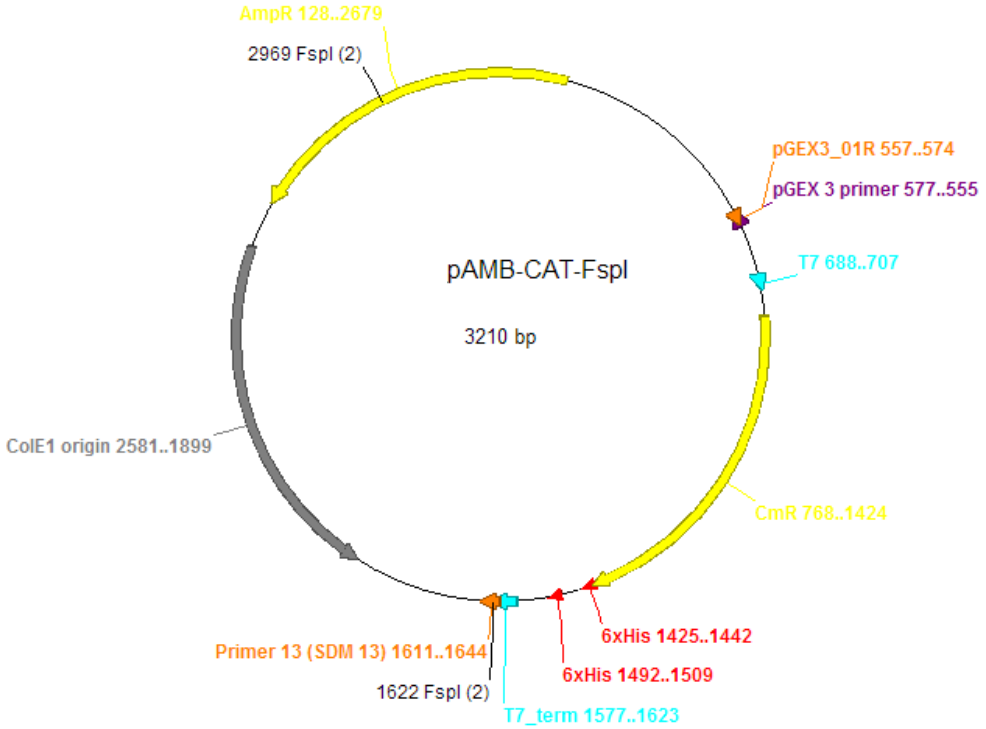
```
  AUTHORS   Gilbert,W.
  TITLE     Obtained from VecBase 3.0
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1993) Department of Biochemistry, University of
            Minnesota, St. Paul, MN 55108, USA
REFERENCE   5  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Rutgers, The State University of New
            Jersey, Waksman Institute of Microbiology, 190 Frelinghuysen Road,
            Piscataway, NJ 08854-8020, USA
  REMARK    Sequence update by submitter
COMMENT     pUC19_COG1564 from 1 to 3716
COMMENT     Alignment and annotation to NCBI reference for Tte MB4 genome -
            http://tinyurl.com/bn2w7cq  On Apr 11, 2002 this sequence version
            replaced gi:209213. These data and their annotation were supplied
            to GenBank by Will Gilbert under the auspices of the GenBank
            Currator Program. pUC19c - Cloning vector (beta-galactosidase mRNA
            on complementary strand) ENTRY PUC19C                    #TYPE DNA
            CIRCULAR TITLE pUC19c - Cloning vector (beta-galactosidase mRNA on
            complementary strand) DATE 03-FEB-1986    #sequence  16-DEC-1986
            ACCESSION VB0033 SOURCE artificial COLLECTION  ATCC 37254 REFERENCE
            #number    #authors Norrander J., Kempe T., Messing J.    #journal
            Gene (1983) 26: 101-106 REFERENCE    #number  1    #authors
            Yanisch-Perron C., Vieira J., Messing J.    #journal Gene (1985)
            33: 103-119 #comment shows the complete compiled sequence REFERENCE
            #number 2    #authors  Chambers,S.P., et al.    #journal Gene
            (1988) 68: 139-149    #describes mutation at nt1308 and its effect
            on copy number REFERENCE    #number    #authors Pouwels P.H.,
            Enger-Valk B.E., Brammar W.J.    #book    Cloning Vectors, Elsvier
            1985 and supplements    #comment  vector I-A-iv-20 COMMENT    This
            Sequence was obtained 3-MAR-1986 from J. Messing, Waksman
            Institute, NJ on floppy disk.    Revised 16-DEC-1986 by F.
            Pfeiffer:    1062/3 'AT' to 'TA' to match revised sequence of
            PBR322    The beta-galactosidase mRNA sequence including the
            multiple    cloning site of M13mp19 is on the strand complementary
            to that    shown. KEYWORDS CROSSREFERENCE    #complement
            VecBase(3):pUC19 #prerevised      GenBank(50):M11662,
            EMBL(11):ARPuc19    #parent VecBase(3):pUC13, VecBase(3):M13mp19,
            VecSource(3):bGal19 PARENT Features of pUC19c (2686 bp) residue
            source       1- 137 2074-2210 pBR322     138- 237 2252-2351 pBR322
            238- 395 1461-1304 (c) Lac-Operon     396- 452 57-   1 (c)
            polylinker of M13mp19     455- 682  1298-1071 (c) Lac-Operon
            683-2686 2352-4355 pBR322    Conflict (cfl) and Mutations (mut):
            pUC19c  source    mut 1308  A  G  2977   pBR322 linked to increased
            copy number    mut 1942  A  G  3611   pBR322 mut 2243 T  C  3912
            pBR322 FEATURE    1629-2417 789-1 (c) Ap-R; b-lactamase POLYLINKER
            HindIII-SphI-PstI-SalI-XbaI-BamHI-SmaI-KpnI-SacI-EcoRI SELECTION
            #resistance  Ap    #indicator beta-galactosidase SUMMARY  pUC19c
            #length 2686   #checksum 4465.
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     primer_bind     423..442
                     /label=T7
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     complement(1494..1511)
```

```
                /label=M13-rev (DNA Core)
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
rep_origin      complement(1879..2561)
                /label=ColE1 origin
                /ApEinfo_fwdcolor=gray50
                /ApEinfo_revcolor=gray50
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             complement(239..307)
                /label=LacZ alpha
                /ApEinfo_fwdcolor=#6495ed
                /ApEinfo_revcolor=#6495ed
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
misc_binding    complement(1517..1539)
                /label=LacO
                /ApEinfo_fwdcolor=#6495ed
                /ApEinfo_revcolor=#6495ed
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             complement(2659..3318)
                /label=AmpR
                /ApEinfo_fwdcolor=yellow
                /ApEinfo_revcolor=yellow
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
primer_bind     complement(29..51)
                /label=pGEX 3 primer
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
source          442..1477
                /label=NC_003869.1 -1523276..-1522241
                /ApEinfo_fwdcolor=#c0c0c0
                /ApEinfo_revcolor=#ffffff
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
gene            968..1386
                /label=GeneID 995665
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             982..1386
                /label=COG1563 7-cyano-7-deazaguanine reductase
                /ApEinfo_fwdcolor=#8080ff
                /ApEinfo_revcolor=#ffff00
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
gene            497..992
                /label=GeneID 995666
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             510..992
                /locus_tag="TTE1564"
                /note="Best Blastp hit =
                gi|12723715|gb|AAK04893.1|AE006312_13 (AE006312)
```

```
                     HYPOTHETICAL PROTEIN [Lactococcus lactis subsp. lactis],
                     score 71.6, E-value 6.00E-12"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="NP_623164.1"
                     /db_xref="GI:20807993"
                     /db_xref="GeneID:995666"
                     /label=COG1564 Hypothetical protein
                     /ApEinfo_fwdcolor=#00ff00
                     /ApEinfo_revcolor=#ffff00
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     RBS             496..503
                     /label=SD JenkinsJBC2011_286_2426
                     /ApEinfo_fwdcolor=#ff80ff
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     378..395
                     /label=M13FORW (DNA Core)
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    complement(965..986)
                     /label=TYE563_LNA
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
ORIGIN
        1 tcgcgcgttt cggtgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca
       61 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg
      121 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc
      181 accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc
      241 attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat
      301 tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt
      361 tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggt acccggggat
      421 cctaatacga ctcactatag ggcagtgagc aacaaaatgc tcacctgggt cgcagtaacc
      481 ccagttaaca aaacaaggga ggtaattttG TGCCCAAAAA AAGAATAAAA GATTTAGCTG
      541 AAATTGCTCT TGTTGCAGCA ATTTATTTCG CACTCACAAT TATATTTTCG TCCATTTCGT
      601 TTTTACCCGT TCAATTTCGA ATCGGGGAAA TTACGAAATC CATTGTAGTA TTCAATAAAA
      661 AATATGCTAT TTCCATGATG ATAGGAAATT TTTTTGCAAA TTTGTTTAGC CCATTTGCTG
      721 GTGCAATGGA ATTAATTTTT ATGCCTCTTT CGAACTTAAT AGGCTGTACA ATTGGATACT
      781 ACATTGGAAG ACTTACTCAC AAAGCGATAG GAGCTATATT CATAGCCCTT TGGATTGCAG
      841 CATCAGTTGC AATTACTTTA AAGGTTTCTG CAGGCATACC ATTTATTCCG ACTTTCTTAA
      901 GCGTGGGAGT AGCGGAAACT GTACTTTTGG TAACTGGATA TTTTTTGCTT TTCACAATTG
      961 AAAAGAAAGG AGTTGTGAAA TTTGACAGAT AAATATAAAG AGAGAAGATT TGACATTTAC
     1021 GGTTACGAAA AAATTGACAA AGAAGTTCTA GAATCTATTG AATATGAGTA TCCTGAAAAA
     1081 AATACTATCG TGGAGTATAT TACCGATGAA TTTTCTTCTG TTTGCCCTTG GACAGGATTA
     1141 CCTGACAATG CAAAACTTAC TATAAGGTAT ATACCCCACA AAAAACTTGT AGAACTTAAA
     1201 TCCTTAAAAT ATTACCTTAC ATCTTATAGG AATGTAGGTA TATTGCAAGA ACATGCAATA
     1261 AACAGAATTT TAGATGATTT GGTGGAATTC CTGCAGCCAA AATTTATGGA AATAATAGGC
     1321 GAATTTCAGG AAAGAGGAGG AATAGCTACA AGAATTATAG CAAGGTATGA AAAAGAGGAG
     1381 TATTAAactt aaaaggctgc ctaaaatttt gtaggcagct tttttattca ttttagtttt
     1441 tcttcaaaat gagtttacag agaacgatga ggaagcaagc ttggcgtaat catggtcata
     1501 gctgtttcct gtgtgaaatt gttatccgct cacaattcca cacaacatac gagccggaag
     1561 cataaagtgt aaagcctggg gtgcctaatg agtgagctaa ctcacattaa ttgcgttgcg
     1621 ctcactgccc gctttccagt cgggaaacct gtcgtgccag ctgcattaat gaatcggcca
     1681 acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc gcttcctcgc tcactgactc
     1741 gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatacg
     1801 gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag gccagcaaaa
```

```
1861 ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc gcccccctga
1921 cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag gactataaag
1981 ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga ccctgccgct
2041 taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc atagctcacg
2101 ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc
2161 ccccgttcag cccgaccgct gcgccttatc cggtaactat cgtcttgagt ccaacccggt
2221 aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca gagcgaggta
2281 tgtaggcggt gctacagagt tcttgaagtg gtggcctaac tacggctaca ctagaagaac
2341 agtatttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag ttggtagctc
2401 ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca agcagcagat
2461 tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc
2521 tcagtggaac gaaaactcac gttaagggat tttggtcatg agattatcaa aaaggatctt
2581 cacctagatc cttttaaatt aaaaatgaag ttttaaatca atctaaagta tatatgagta
2641 aacttggtct gacagttacc aatgcttaat cagtgaggca cctatctcag cgatctgtct
2701 atttcgttca tccatagttg cctgactccc cgtcgtgtag ataactacga tacgggaggg
2761 cttaccatct ggccccagtg ctgcaatgat accgcgagac ccacgctcac cggctccaga
2821 tttatcagca ataaaccagc cagccggaag ggccgagcgc agaagtggtc ctgcaacttt
2881 atccgcctcc atccagtcta ttaattgttg ccgggaagct agagtaagta gttcgccagt
2941 taatagtttg cgcaacgttg ttgccattgc tacaggcatc gtggtgtcac gctcgtcgtt
3001 tggtatggct tcattcagct ccggttccca acgatcaagg cgagttacat gatcccccat
3061 gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc gttgtcagaa gtaagttggc
3121 cgcagtgtta tcactcatgg ttatggcagc actgcataat tctcttactg tcatgccatc
3181 cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat
3241 gcggcgaccg agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag
3301 aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt
3361 accgctgttg agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc
3421 ttttactttc accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa
3481 gggaataagg gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg
3541 aagcatttat cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa
3601 taaacaaata ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac
3661 cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtc
//
```

Figure A.5-3 pAMB_CAT-FspI plasmid for *in vitro* transcription of CAT control mRNA for *in vitro* translation




```
LOCUS       New_DNA                 3210 bp ds-DNA     circular     21-OCT-2014
DEFINITION  .
ACCESSION
VERSION
SOURCE      .
  ORGANISM  .
COMMENT     pR15_pAMB-CAT-FspI from 1 to 3210
COMMENT     //nwalter-serv.chem.lsa.umich.edu/data/data2/PaulLund/Plasmids-ApeA
            nnotated/pUC19c/pAMB-CAT(pUC19 CAT).ape from 1 to 3210
COMMENT     control vector provided with the ActivePro? In Vitro Translation
            Kit (Ambion) , FspI site introduced using SMD 13 primer: GAG GGG
            TTT TTT GCG CAA AGG AGG AAC TAT ATC C
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     misc_feature    1577..1623
                     /label=T7_term
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_binding    1425..1442
                     /label=6xHis
                     /ApEinfo_fwdcolor=#ff0000
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     1611..1644
                     /label=Primer 13 (SDM 13)
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
```

```
                     width 5 offset 0
     rep_origin      complement(1899..2581)
                     /label=ColE1 origin
                     /ApEinfo_fwdcolor=gray50
                     /ApEinfo_revcolor=gray50
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     CDS             complement(join(2679..3210,1..128))
                     /label=AmpR
                     /ApEinfo_fwdcolor=yellow
                     /ApEinfo_revcolor=yellow
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     CDS             768..1424
                     /label=CmR
                     /ApEinfo_fwdcolor=yellow
                     /ApEinfo_revcolor=yellow
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     complement(555..577)
                     /label=pGEX 3 primer
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     protein_bind    1492..1509
                     /label=6xHis(1)
                     /ApEinfo_label=6xHis
                     /ApEinfo_fwdcolor=red
                     /ApEinfo_revcolor=red
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     290..307
                     /label=Amp_01R
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     557..574
                     /label=pGEX3_01R
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
ORIGIN
        1 ttttctgtga ctggtgagta ctcaaccaag tcattctgag aatagtgtat gcggcgaccg
       61 agttgctctt gcccggcgtc aatacgggat aataccgcgc cacatagcag aactttaaaa
      121 gtgctcatca ttggaaaacg ttcttcgggg cgaaaactct caaggatctt accgctgttg
      181 agatccagtt cgatgtaacc cactcgtgca cccaactgat cttcagcatc ttttactttc
      241 accagcgttt ctgggtgagc aaaaacagga aggcaaaatg ccgcaaaaaa gggaataagg
      301 gcgacacgga aatgttgaat actcatactc ttcctttttc aatattattg aagcatttat
      361 cagggttatt gtctcatgag cggatacata tttgaatgta tttagaaaaa taaacaaata
      421 ggggttccgc gcacatttcc ccgaaaagtg ccacctgacg tctaagaaac cattattatc
      481 atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtcTCGC GCGTTTCGGT
      541 GATGACGGTG AAAACCTCTG ACACATGCAG CTCCCGGAGA CGGTCACAGC TTGTCTGTAA
      601 GCGGATGCCG GGAGCAGACA AGCCCGTCAG GGCGCGTCAG CGGGTGTTGG CGGGTGTCGG
      661 GGCTGGCAGA TCTCGATCCC GCGAAATTAA TACGACTCAC TATAGGGAGA CCACAACGGT
      721 TTCCCTCTAG AAATAATTTT GTTTAACTTT AAGAAGGAGA TATACATATG GAGAAAAAAA
      781 TCACTGGATA TACCACCGTT GATATATCCC AATGGCATCG TAAAGAACAT TTTGAGGCAT
      841 TTCAGTCAGT TGCTCAATGT ACCTATAACC AGACCGTTCA GCTGGATATT ACGGCCTTTT
      901 TAAAGACCGT AAAGAAAAAT AAGCACAAGT TTTATCCGGC CTTTATTCAC ATTCTTGCCC
      961 GCCTGATGAA TGCTCATCCG GAATTCCGTA TGGCAATGAA AGACGGTGAG CTGGTGATAT
     1021 GGGATAGTGT TCACCCTTGT TACACCGTTT TCCATGAGCA AACTGAAACG TTTTCATCGC
```

```
     1081 TCTGGAGTGA ATACCACGAC GATTTCCGGC AGTTTCTACA CATATATTCG CAAGATGTGG
     1141 CGTGTTACGG TGAAAACCTG GCCTATTTCC CTAAAGGGTT TATTGAGAAT ATGTTTTTCG
     1201 TCTCAGCCAA TCCCTGGGTG AGTTTCACCA GTTTTGATTT AAACGTGGCC AATATGGACA
     1261 ACTTCTTCGC CCCCGTTTTC ACCATGGGCA AATATTATAC GCAAGGCGAC AAGGTGCTGA
     1321 TGCCGCTGGC GATTCAGGTT CATCATGCCG TTTGTGATGG CTTCCATGTC GGCAGAATGC
     1381 TTAATGAATT ACAACAGTAC TGCGATGAGT GGCAGGGCGG GGCGCATCAT CATCATCATC
     1441 ATTAAGGATC CGAATTCGAG CTCCGTCGAC AAGCTTGCGG CCGCACTCGA GCACCACCAC
     1501 CACCACCACT GAGATCCGGC TGCTAACAAA GCCCGAAAGG AAGCTGAGTT GGCTGCTGCC
     1561 ACCGCTGAGC AATAACTAGC ATAACCCCTT GGGGCCTCTA AACGGGTCTT GAGGGGTTTT
     1621 TTGCGCAAAG GAGGAACTAT ATCCGGACCC GCTTTCCAGT CGGGAAACCT GTCGTGCCAG
     1681 CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC
     1741 GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT
     1801 CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG
     1861 TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC
     1921 CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA
     1981 AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT
     2041 CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG
     2101 GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG
     2161 CTGGGCTGTG TGCACGAAcc ccccgttcag cccgaccgct gcgccttatc cggtaactat
     2221 cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac
     2281 aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac
     2341 tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc agttaccttc
     2401 ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt
     2461 tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc
     2521 ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg
     2581 agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca
     2641 atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca
     2701 cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag
     2761 ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac
     2821 ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc
     2881 agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct
     2941 agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc
     3001 gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg
     3061 cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc
     3121 gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat
     3181 tctcttactg tcatgccatc cgtaagatgc
//
```

A.6 Matlab codes for SiM-KARTS-related analyses

AcceptorAveraging_truncate_Qub.m

Description

Pre-process single molecule trace data prior to idealization with QuB. Provides rough normalization of trace intensity. Modified from existing script from Mario Blanco. Paul Lund 06/2015 added trace length truncation to make trace length data uniform when working with movies of different length.

```
clear;
close all;
warning off MATLAB:divideByZero;
```

Get Input File Names

```
fprintf(1,'Please select the files you wish to Average\n')
[TestFileName, TestPathName] = uigetfile('*.dat','Please select all your Path
files','Multiselect','on');
```

```
nfiles = length(TestFileName);
```

Get Normalization factor (multiples of StDev) and number of frames to truncate

```
factor = input ('Multiples of stDev to use for averaging (4)? ');
if length(factor)==0
    factor = 4;
end
truncLen = input('How many frames do you wish to sample? (entire trace) ');
if length(truncLen)==0
    trunc = false;
else
    trunc = true;
end
```

This program gets the average acceptor value and stDev, then normalizes the acceptor values by the avg + n*StDev

```
    for n=1:nfiles
        FileName = strcat(TestPathName,TestFileName{n});
        fid = fopen(FileName,'rt');
        raw = fscanf(fid,'  %e %e %e', [3 inf]);
        fclose(fid);
        raw = raw';
        if trunc  %determine whether to take first n frames or entire trace based on
trunc flag
            len = truncLen;
        else
            len = length(raw);
        end

        Acceptor = raw(1:len,3);
        m = mean(Acceptor,1);
        sd = std(Acceptor);
        Norm = m+factor*sd;
        AccNorm = Acceptor/Norm;
        Donor = 1-AccNorm;
        %C = horzcat(Donor, AccNorm);
        time=1:len;
        time=time';
        C = horzcat(time, Donor, AccNorm);
        %NameMod = strtok(FileName,'.');
        NameMod = deblank(strjoin(strsplit(FileName,'.dat'))); %Get file name w/o .dat
file extension

        if trunc  % Change file naming to indcate whether trace has been truncated
            NameMod = strcat(NameMod, '_truncFr', num2str(truncLen),'_Accept.dat');
        else
            NameMod = strcat(NameMod,'_Accept.dat');
        end
        fid = fopen(NameMod,'wt');
        fprintf(fid,'%e %e %e\n', C');
        fclose(fid);
    end
```

Published with MATLAB® R2015a

Fano_Calc.m

Contents

Description

```
% Calculates Fano factor values for a given set of time windows (Ln 59).
% Uses a defined random stream to select a time
% window of speicifed length from each trace in the data set. Re-running the
% script with the same set of traces and number of time windows will select
% the same winodws from those traces and thus return the same fano factor
% values for each time window.
```

Mario Blanco, 06/2015

Put the path files (5 column data with time/donor/acceptor/FRET/ideal) into a folder called ideal (Ln 23) Example: C:\Users\mrblanc\Dropbox\ArlieCluster (1)\For_Clustering\100uMQ\ideal Add the path files into ideal, then when the dialog box comes up select the 100uMQ folder (one up from ideal). The code will search for a folder named ideal, then create a new folder called remainder and spit out the 95% confidence interval and Fano factors in the command window in Matlab.

Paul Lund, 06/2015

Add option to calculate Fano values a user-specified number of times for a given set of time windows using randomly selected window, then calculate average value and std dev of the Fano factor for each time window in the set. Remove remainder calculation.

Import Data

```
clear all;
close all;

folder_name = uigetdir('Z:\PaulLund\QuB\datasets\test\ideal');
%find subdirectories; requires subdir.m function from Elmar Tarajan
[Elmar.Tarajan@Mathworks.de]
[subdirect files]=subdir(folder_name);

for n=1:size(subdirect,2)
    Tokens = {};
    A=subdirect{1,n};
    while ~isempty(A)
        [Tokens{end+1} A] = strtok(A,'\');
    end

    if strcmp(Tokens{end},'ideal')
        %         mkdir(subdirect{1,n},'remainder');
        fprintf('Time to Analyze\n');
        sprintf([Tokens{1,end-3:end}],'\n')
        listingall=dir(fullfile(subdirect{1,n},'*.dat'));
        for j=1:size(listingall,1)
            %             if
exist(strcat(subdirect{1,n},'\remainder\',strtok(listingall(j,1).name,'.'),'_remain.da
t'), 'file')
            %                 fprintf('Already analyzed\n')
            %             else
            RAW=[];
            RAW=importdata(strcat(subdirect{1,n},'\',listingall(j).name));
```

```
            RAWnormstore{j,n}=RAW;
        end
    end
end
```

Setup options for Fano calculations

```
Count=[];

lengthvect=[50 100 200 400]; %List of time windows in frames for which to calculate
Fano factor
trial = input ('Number of trials if not using defined random stream? [none, use
defined stream]');
if isempty(trial)
    myStream=RandStream('mt19937ar');
    RandStream.setGlobalStream(myStream);
    trial = 1;
    definedStream=1;
else
    definedStream=0;
    Fanotrials = cell(length(lengthvect),1);
    mkdir(folder_name,'ideal\trials');
    verbose = input ('Save result from each trial? [n]','s');
    if  strcmp(verbose,'y')
        verbose = 1;
    else
        verbose = 0;
    end
end
```

Calculate Fano factors

```
for  q=1:trial
    for j=1:size(RAWnormstore,2)
        for n=1:size(RAWnormstore,1)
            if ~isempty(RAWnormstore{n,j})
                for p=1:length(lengthvect)
                    states=unique(RAWnormstore{n,j}(:,5));
                    states=sort(states);
                    if (size(RAWnormstore{n,j},1)-lengthvect(p)-1)>0
                        Int = randi([1 size(RAWnormstore{n,j},1)-lengthvect(p)-
1],1,1);
                        if numel(states)<2
                            Count = 0;
                        else
                            %Count{n,p} =
numel(find(RAWnormstore{n,j}(Int:Int+lengthvect(p),5)==states(2)));
                            %figure;
                            %plot(RAWnormstore{n,j}(Int:Int+lengthvect(p),5))
                            %hold on;

Count{n,p}=round(numel(find(diff(RAWnormstore{n,j}(Int:Int+lengthvect(p),5))~=0))/2);
                            %title(num2str(Count{n,p}));
                            %hold off;
                        end
                    else
                    end
                end
            else
            end
        end
    end
```

```
    Fano=[];
    for n=1:size(Count,2)
        Fano(n,1)=var(cat(1,Count{:,n}))./mean(cat(1,Count{:,n}));
        Fano(n,2)=mean(cat(1,Count{:,n}));
        Fano(n,3)=var(cat(1,Count{:,n}));
    end

  figure;
    plot(lengthvect,Fano(:,1),'o');
    hold on;
    bounds=gaminv([.025,.975],(size(Count,1)-1)/2,2/(size(Count,1)-1));
    display(bounds)
    Fanoout=horzcat(lengthvect'./10,Fano(:,1));
    display(Fanoout);
    axis([min(lengthvect) max(lengthvect) 0 2.5]);
    plot(lengthvect,repmat(bounds(1),1,length(lengthvect)),'r--');
    plot(lengthvect,repmat(bounds(2),1,length(lengthvect)),'r--');
    ylabel('Fano Factor')
    xlabel('Window length (frames)')
    title(['Fano vs Poisson: lengthvect = ' char(num2str(lengthvect))]);

    if definedStream==1
        dlmwrite(strcat(folder_name,'\ideal\','Fano.txt'),Fano,'delimiter','\t');
        % save picture of graph
        h =gcf;
        fname = [folder_name '\ideal\FanoPlot' strrep(char(num2str(lengthvect)),'
','-') '.jpg'];
        print(h,'-r150','-djpeg',fname)
    else
        Fanotrials{q} = Fanoout;
        if verbose == 1
            % save result
            output = vertcat(bounds, [Fanoout(:,1) Fanoout(:,2)]);

save(strcat(folder_name,'\ideal\trials\','Fano', '_trial',num2str(q),'.txt'),'output',
'-ascii');

            %save picture of plot
            h =gcf;
            fname = [folder_name '\ideal\trials\FanoPlot' '_trial',num2str(q),'.jpg'];
            print(h,'-r150','-djpeg',fname)
        else
        end
        close(gcf);
    end
end

% calculate avergage fano value from trials and std deviation
if definedStream ==0
    avgFanoValues = zeros(length(lengthvect),3);
    for k=1:length(lengthvect)
        b=[];
        for m=1:length(Fanotrials) %collect all fano values for time window k
            b = [b Fanotrials{m}(k,2)];
        end
        avgFanoValues(k,1)=lengthvect(k); % store time window
        avgFanoValues(k,2)=mean(b); % store mean Fano value from trials
        avgFanoValues(k,3)=std(b); % store std of Fano values from trials
    end
    close all;
    %plot avg Fano values with std dev
```

```
    figure;
    errorbar(avgFanoValues(:,1),avgFanoValues(:,2),avgFanoValues(:,3),'d');
    hold on;
    display(bounds)
    axis([min(lengthvect)*0.95 max(lengthvect)*1.05 0 2.5]);
    plot(lengthvect,repmat(bounds(1),1,length(lengthvect)),'r--');
    plot(lengthvect,repmat(bounds(2),1,length(lengthvect)),'r--');
    ylabel('Fano factor')
    xlabel('Window length (frames)')
    title(strcat('Average Fano factor:', num2str(trial),' trials'));

    %save plot
    h =gcf;
    fname2 =
[folder_name '\ideal\trials\avg_FanoPlot_',num2str(trial), '_trials.jpg'];
    print(h,'-r150','-djpeg',fname2)

    %save result
    boundsOut = [bounds 0];
    output = vertcat([avgFanoValues(:,1)./10 avgFanoValues(:,2) avgFanoValues(:,3)],
boundsOut);
    save(strcat(folder_name,'\ideal\trials\','avgFano_',
num2str(trial),'_trials.txt'),'output','-ascii');
end
```

Published with MATLAB® R2015a

subdir.m (accessory function necessary for Fano_Calc.m; provided by Elmar Tarajan)

```
function [sub,fls] = subdir(CurrPath)
%   SUBDIR  lists (recursive) all subfolders and files under given folder
%%   SUBDIR
%        returns all subfolder under current path.
%
%   P = SUBDIR('directory_name')
%       stores all subfolders under given directory into a variable 'P'
%
%   [P F] = SUBDIR('directory_name')
%       stores all subfolders under given directory into a
%       variable 'P' and all filenames into a variable 'F'.
%       use sort([F{:}]) to get sorted list of all filenames.
%
%   See also DIR, CD
%   author:  Elmar Tarajan [Elmar.Tarajan@Mathworks.de]
%   version: 2.0
%   date:    07-Dez-2004
%
if nargin == 0
   CurrPath = cd;
end% if
if nargout == 1
   sub = subfolder(CurrPath,'');
else
   [sub fls] = subfolder(CurrPath,'','');
end% if
  %
  %
function [sub,fls] = subfolder(CurrPath,sub,fls)
%-------------------------------------------------
tmp = dir(CurrPath);
```

```
tmp = tmp(~ismember({tmp.name},{'.' '..'}));
for i = {tmp([tmp.isdir]).name}
   sub{end+1} = [CurrPath '\' i{:}];
   if nargin==2
      sub = subfolder(sub{end},sub);
   else
      tmp = dir(sub{end});
      fls{end+1} = {tmp(~[tmp.isdir]).name};
      [sub fls] = subfolder(sub{end},sub,fls);
   end% if
end% if
```

Published with MATLAB® R2015a

ParseDWT_avgDwellTime.m

Contents

Description

Repurpose original code to calculate avg dwell time in state 1 and state 2 for each molecule from idealized traces data in *.dwt format from QuB. Capable of handling >2 states. Paul Lund 06/2015

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Data Processing Post-QuB            %
% Franklin Fuller 08/2008             %
%                                     %
% Paul Lund 06/2015                   %
% Repurpose to caluclate avg dwell    %
% time in state 1 and state 2 for each%
% molecule.                           %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%  Opens a dialog box where you can select all your *.dwt files.
%  It will read all of them in and parse them into HaMMy-like path files,
%  for use with post-filtering software.

%  As always, you'll have to set the current directory in Matlab to where
%  this file is located, but the *.dwt files can be in any directory.

clear;
close all;
warning off MATLAB:divideByZero;
```

Get the File Names & User Preferences

```
fprintf(1,'Select all your *.dwt files (from a single directory only)\n')

% [TestFileName, TestPathName] = uigetfile('*.dwt','Please select all your *.dwt
files','Multiselect','on'); %

%*** Currently, Multiselect option doesn't work with code b/c the variable nseg goes
out of bounds when there are multiple files
```

```
[TestFileName, TestPathName] = uigetfile('*.dwt','Please select  your *.dwt
file','Multiselect','off');
% This is just in case the user selects only one file (converts array ->
% cell array), because the rest of the code expects a cell array.
if ~iscell(TestFileName)
   G = cell(1);
   G{1} = TestFileName;
   TestFileName = G;
end
nfiles = size(TestFileName);
nfiles = nfiles(1,2);

% userchoice2 = input('Diagnostic Output? (y/n) [n]: ','s');
% if strcmp(userchoice2,'n') || isempty(userchoice2)
%     Verbose = 0;
%     refmax = 1;
% else
    Verbose = 1;
    refmax = 3;
% end
dropFirstLast = input('Discard first and last dwell times from trace? (y/n) [y]:
','s');
if strcmp(dropFirstLast,'y') || isempty(dropFirstLast)
    dropFirstLast = 1;
else
    dropFirstLast = 0;
end
```

Parse QuB *.dwt files

```
%Regular Expressions Used
seg = '\S+\s(?<seg>\d+)\s'; %returns segment #
dwl = '\S+\s(?<dwl>\d+)\s'; %returns #dwells found in seg #
smp = '\S+\s(?<smp>\d+)\s'; %returns sampling rate used in the seg
srt = '\S+\s(?<srt>\d+)\s'; %returns start pos relative to total file
csc = '\S+\s(?<csc>\d+)\s'; %returns # of classes fit in the seg
dat = '(?<css>\d+)\s(?<dur>\d+)'; %returns class (*.css) and duration (*.dur)
aml = '\d+\s(?<dnr>\S+)\s(?<acp>\S+)\s(?<frt>\S+)'; %reads amal lines, returns donor,
acceptor, & fret

%Note to the potential reader: I strongly recommend reading the matlab help
%files on Regular Expressions and the Structure Data type to better
%understand this code.  Also, look at the output *.dwt files from QuB in
%Notepad to understand the apriori formatting assumptions.


 for n=1:nfiles

    FileName = strcat(TestPathName,TestFileName{n}); %get *.dwt file name
    fid = fopen(FileName,'rt'); %open the *.dwt file
    trigger = 0; %determines presence of header in current read line
    datac = cell(1); %non-contiguously stores incoming data pieces
    index = 1; %records read location for internal referencing
    sindex = 0; %temporarily records header location for internal reference
    drec = 0; %records whether or not data since the last header has been rec'd
    while ~feof(fid)
        read = fgetl(fid);
        exprh = strcat(seg,dwl,smp,srt,csc); %compose the header regexp
        parseh = regexp(read,exprh,'names'); %parse the header
        if size(parseh,1)~=0 %if header is found, size of parseh not = zero
            if (drec == 1)   %if data has been rec'd and new
```

```
                datao{str2double(parseh(1,1).seg)-1} = cat(1,datac{:});    %header is
found, vertcat rec'd data
                len = size(datao{str2double(parseh(1,1).seg)-1},1); %Add a time column
(units of frames).
                Time = 1:len;
                datao{str2double(parseh(1,1).seg)-1} =
horzcat(Time',datac{str2double(parseh(1,1).seg)-1},datao{str2double(parseh(1,1).seg)-
1}(:,1:refmax));
                datac = cell(1); %flush datac contents
                fprintf(1,'Flushed datac! Seg %g\n',nseg)
                stor = parseh(1,1).seg;
            end
            nseg = str2double(parseh(1,1).seg); %store header values as numbers
            ndwl = str2double(parseh(1,1).dwl); %(used for decoding the DWT)
            nsmp = str2double(parseh(1,1).smp);
            nsrt = str2double(parseh(1,1).srt);
            ncsc = str2double(parseh(1,1).csc);
            if ncsc>=1
                %Warning to reader: this portion of the code is EXTREMELY
                %ghetto and confusing.  All this stuff is here to decrypt
                %the amplitudes and deviations in the general case that
                %some amps or devs may be integers (as opposed to floats).
                adv1 = '\s(?<';
                adv2 = '>[-]?\d*[.]?\d*[E]?[-]?\d*)\s(?<';
                adv3 = '>[-]?\d*[.]?\d*[E]?[-]?\d*)';
                for q=1:ncsc
                    if q==1
                        madv =
strcat('\d',adv1,'amp',num2str(q),adv2,'dev',num2str(q),adv3);
                    else
                        madv =
strcat(madv,adv1,'amp',num2str(q),adv2,'dev',num2str(q),adv3);
                    end
                end
                %Now that we've constructed a regular expression specific
                %to the header....
                x = regexp(read,madv,'names'); %reads in amplitude and dev
                gamps = cell(1);
                gdevs = cell(1);
                for q=1:ncsc
                    gamps{q} = str2double(x.(strcat('amp',num2str(q)))); %note the
dynamic field name syntax.
                    gdevs{q} = str2double(x.(strcat('dev',num2str(q))));
                end
                z.amps = cat(1,gamps{:}); z.devs = cat(1,gdevs{:});
                trigger = 1;                    %values for data associated
                sindex = index;                 %with this header
                drec = 0;
            else
                fprintf(1,'Error Detected in Modeling Parameters: Null Class Set')
                return;
            end
        end
        if (trigger == 1 && sindex ~= index)
            v = regexp(read,dat,'names');     %parse in data
            if isempty(v)
                break; %sometimes (maybe all the time) there are extra blank lines
before eof
            end
            datac{index} = [z.amps(str2double(v.css)+1) str2double(v.dur)];
            if Verbose == 1
```

```
                datac{index} = horzcat(datac{index},[str2double(v.css)
str2double(v.dur)]);
            end
%             datac{index} = repmat(datac{index},(str2double(v.dur)/nsmp),1);
            drec = 1;
        end
        index = index+1;
    end
    fprintf(1,'Compiling...\n')
    fclose(fid);
    datao{nseg} = cat(1,datac{:}); %vertcat final data found
    len = size(datao{nseg},1); %Add a time column (units of frames).
    Time = 1:len;
    datao{nseg} = horzcat(Time',datac{nseg},datao{nseg}(:,1:refmax));
    FileName = strtok(FileName,'.');
        dsize = size(datao,2);
```

Calculate dwell times in each state

```
     % Paul Lund 06/2015
     % For each segment, calculate the average dwell time in each state (i.e., class)
and save as file
    segAvgDwells = cell(1,dsize);

    allData = vertcat(datao{:});
    globalCscMax = max(allData(:,4));

    if dropFirstLast ==1; %set whether to take all dwell times
        startOffset = 1;
        endOffset = 1;
    else
        startOffset = 0;
        endOffset = 0;
    end

    for k=1:dsize % load seqments one at a time from datao
        temp = datao{k};
        startPos = 1+startOffset;
        endPos = size(temp,1)-endOffset;
        caseAvgs = zeros(1,globalCscMax+1);

        if endPos>size(temp,1)&& dropFirstLast ==1% check if there are only 2 dwell
times in the trace
            %do nothing and save zeros for avg dwell times
        else
            cscMax = max(temp(startPos:endPos,4)); %determine max number of classes
present in current segment


            for csc=0:cscMax %iterate through all classes
                multiplier = temp(startPos:endPos,4) == csc; % create binary list for
every dwell time: 1 if matches the class, else 0
                if sum(multiplier)~=0 %if class is present
                    caseAvgs(csc+1) = dot(temp(startPos:endPos,3),
multiplier)/sum(multiplier); %multiply binary list by dwell time list and find average
                else
                    caseAvgs(csc+1) = 0;
                end
            end
        end
        segAvgDwells{k} = [caseAvgs(:)];
```

```
    end
    outputFormat = '%1.7f';
    for t =1:globalCscMax  %make output rule for number of observed classes
        outputFormat = strcat('%1.7f\t',outputFormat);
    end
    NameMod = strcat(FileName,'dwellAvgs_Segs1-',num2str(dsize),'.txt');
    fid = fopen(NameMod,'wt');
    for i=1:dsize
        fprintf(fid,outputFormat,segAvgDwells{1,i}');
        if i~=dsize; fprintf(fid,'\n'); end
    end
        fclose(fid);
end
```

Published with MATLAB® R2015a

Appendix B: Supplementary material for characterization of S1 interactions with pseudoknot RNAs

B.1 Concentration measurement of preQ$_1$ stock solution

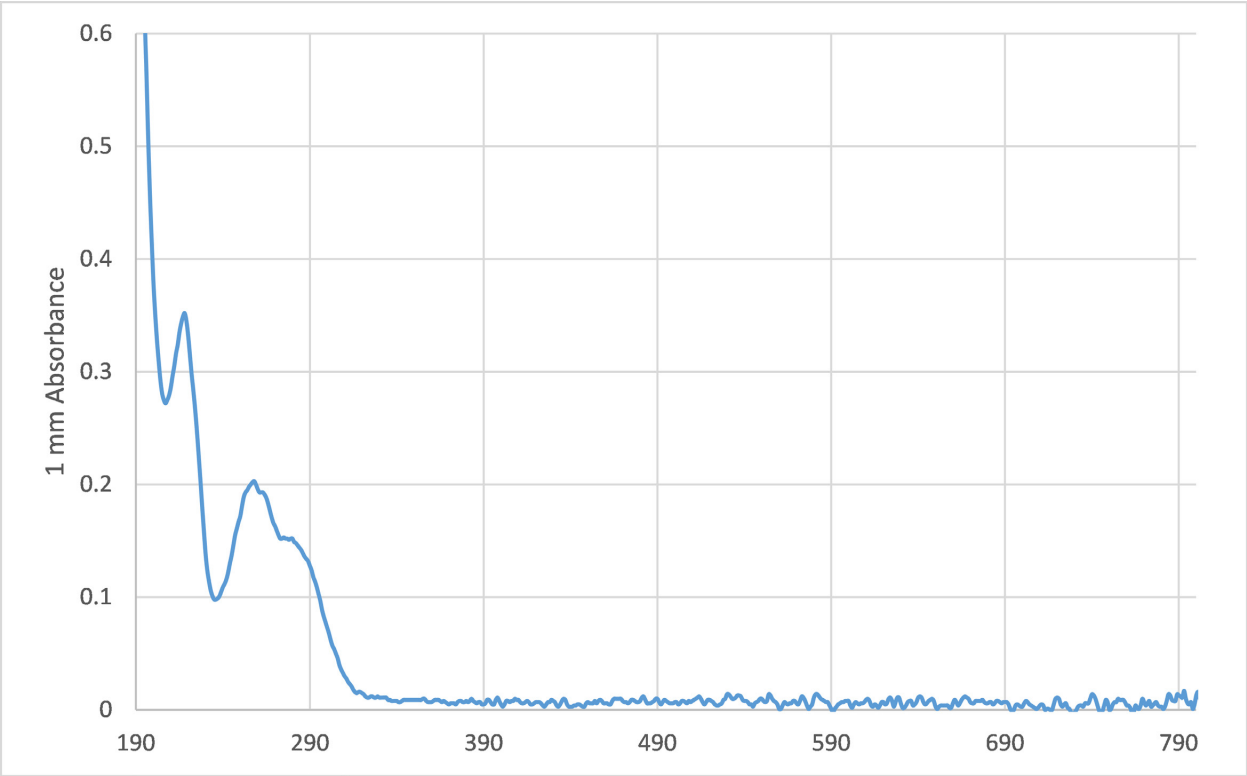


Figure B.1-1 UV-vis spectrum of a preQ$_1$ stock solution.

UV-vis spectrum of preQ$_1$ in milliQ water, obtained using a Nanodrop2000 spectrophotometer.

Appendix Table B.2-1 Site-directed mutagenesis primers for generating mutant aptamer plasmids.

Table of site-directed mutagenesis primers used to generate mutant *Tte* aptamer series of plasmids. Mutation sites are indicated in bold lower case. Nucleotide numbering is relative to the start of the aptamer sequence (see **Figure 3-1**).

Plasmid name	Aptamer mutation	Primer sequence (5′ to 3′)
pUC19_TTE1564	None (wild-type)	--
pUC19_TTE1564_C15U	C15U	GCTCACCTGGGTCGCAGTAA**t**CCCAGTTAACAA
pUC19_TTE1564_C15A	C15A	GCTCACCTGGGTCGCAGTAA**a**CCCAGTTAACAA
pUC19_TTE1564_AU	G5U, G11U, C16A, C30A	CCTGG**t**TCGCA**t**TAAC**a**CCAGTTAACAAAA**a**AAGGGAGGTAATT
pUC19_TTE1564_CG	G5C, G11C, C16G, C30G	CCTGG**c**TCGCA**c**TAAC**g**CCAGTTAACAAAA**g**AAGGGAGGTAATT
pUC19_TTE1564_UUCG-loop	ΔU6-C15, insert UUCG loop	AACAAAATGCTCACCTGGG**ttcg**CCCAGTTAACAAAACAAGG
pUC19_TTE1564_GAAA-loop[a]	ΔU6-C15, insert GAAA loop	CCTTGTTTTGTTAACTGGG**tttc**CCCAGGTGAGCATTTTGTT

[a.] Mutation is introduced through mutagenesis on the (-) strand.

Appendix Table B.2-2 DNA Primers and oligonucleotides for generating *in vitro* transcription templates: +6 and minimal series pseudoknot constructs.

Name	ε_{260} (M^{-1} cm^{-1})	Sequence (5′ to 3′)
PCR template primers:		
Forward: pUC19_Univ_01F	181 500[a]	TTTCCCAGTCACGACGTT
Reverse: Tte1564_+6_01R	192 100[a]	GGGCACAAAATTACCTC
Annealed template oligonucleotides:		
(+) strand: T7_leader	259 808[b]	AATTTAATACGACTCACTATAGG
(-) strand: Tte_IVTxn_rev	626 566[b]	CCCTTGTTTTGTTAACTGGGGTTACTGCGACCCAGGACCTATAGTGAGTCGTATTAAATT
(-) strand: Tte_tailMut_IVTxn_rev	623 053[b]	CCCTTGTTTT**t**TTAACTGGGGTTACTGCGACCCAGGACCTATAGTGAGTCGTATTAAATT
(-) strand: Bsu_C12U_IVTxn_rev	643 915[b]	CCTTAGTTTTTTATAGAGGGTGTAACTAGAACCTCTGCCTATAGTGAGTCGTATTAAATT
(-) strand: Bsu_C12U_tailMut_IVTxn_rev	647 668[b]	CCTTAGTTT**g**TTATAGAGGGTGTA**a**CTAGAACCTCTGCCTATAGTGAGTCGTATTAAATT

[a.] Extinction coefficient provided by Life Technologies

[b.] Extinction coefficient calculated using OligoCalc[140]

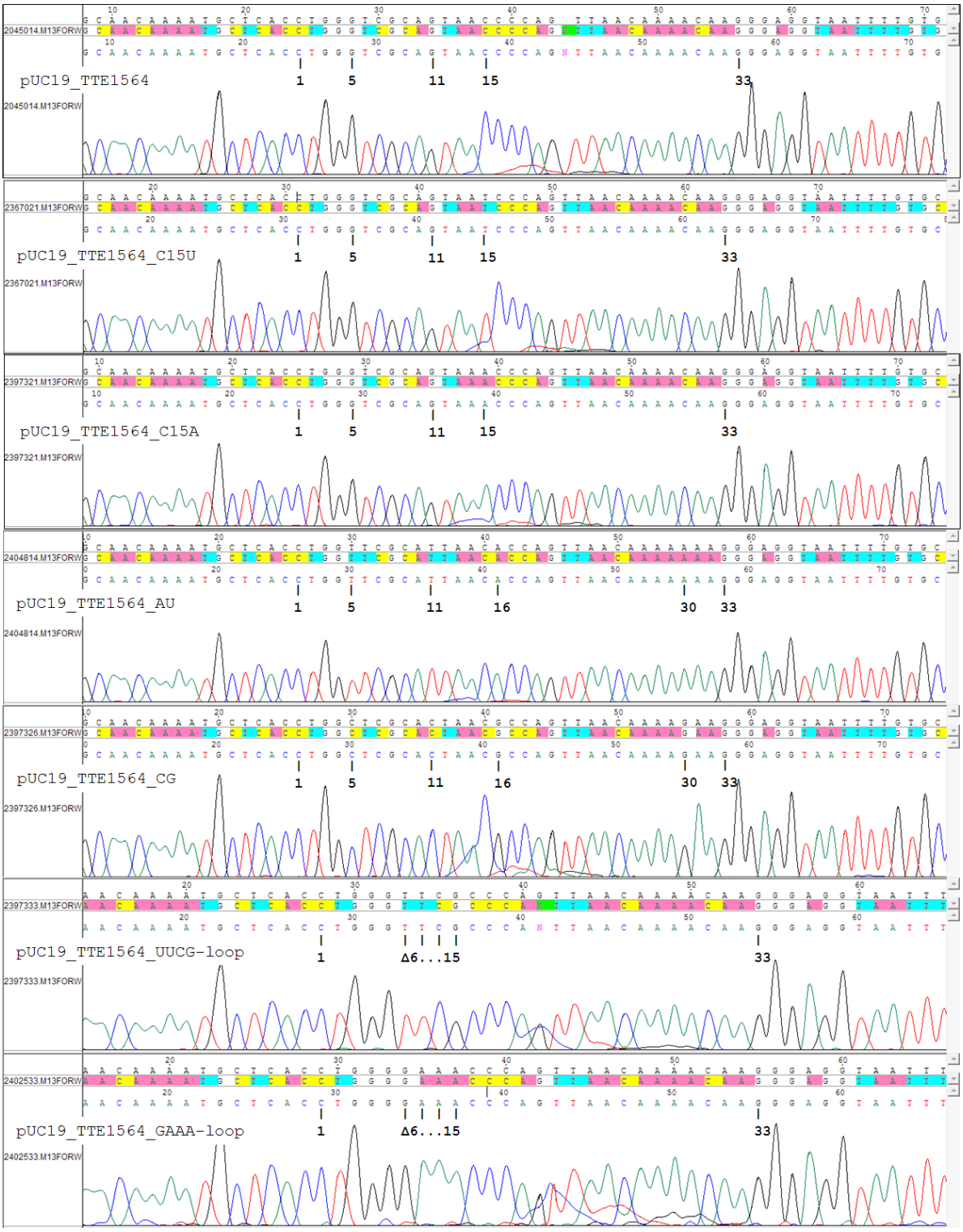


Figure B.2-1 Sequencing Chromatograms for *Tte* mutant aptamer plasmids

Mutated nucleotides, as well the first and last nucleotides of the riboswitch aptamer, are indicated.

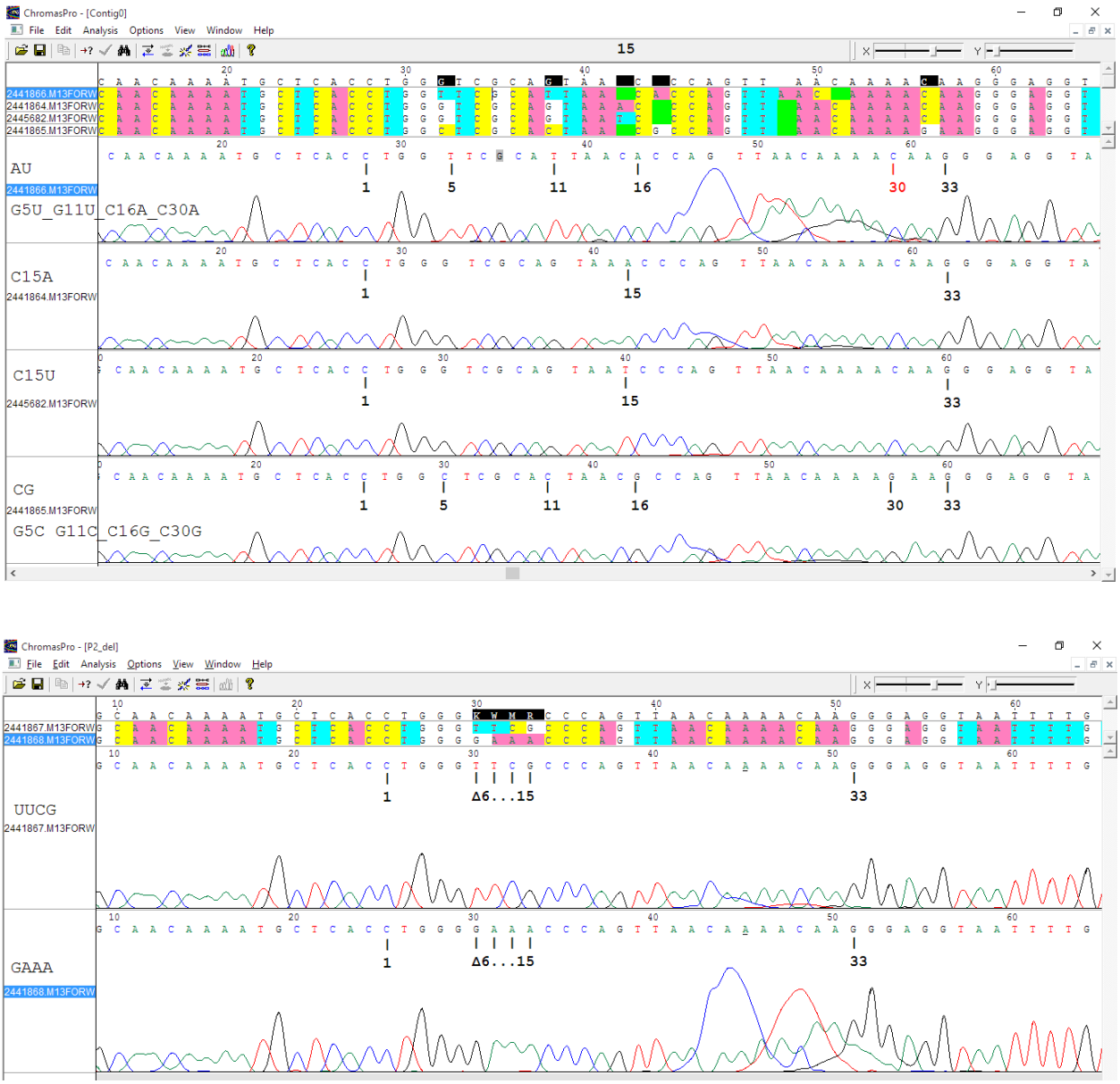


Figure B.2-2 Sequencing chromatograms for PCR products used as templates for *in vitro* transcription of +6 pseudoknot variants

PCR products were sequenced using the M13 forward primer (5'-TGTAAAACGACGGCCAGT-3'). Mutated nucleotides, as well the first and last nucleotides of the riboswitch aptamer, are indicated. Due to inherent difficulty in sequencing very short sequences, the free dye peaks appear prominently in some chromatograms. It is important to note that in the PCR product that served as template for AU, the C30A mutation present in the template plasmid(See **Figure B.2-1**) reverted to the wild-type nucleotide in the PCR product (red), possibly due to amplification bias during PCR.

Appendix Table B.2-3 Complete sequences for RNA pseudoknot constructs.

Mutation sites are indicated in bold lower case. Underlining denotes nucleotides in the open reading frame. Nucleotide numbering is relative to the start of the respective riboswitch aptamer sequence (see **Figure 3-1a**). The molar extinction coefficients reported refer only to the RNA component.

Construct	Aptamer mutation	ε_{260} (M^{-1} cm^{-1})	Length (nt)	Sequence (5′ to 3′)
+6 Series:				
WT	None (wild-type)	945 180	76	GGGCAGUGAGCAACAAAAUGCUCACCUGGGUCGCAGUAACCCCAGUUAACAAAAC AAGGGAGGUAAUUUUGUGCCC
C15U	C15U	946 970	76	GGGCAGUGAGCAACAAAAUGCUCACCUGGGUCGCAGUAA**u**CCCAGUUAACAAAAC AAGGGAGGUAAUUUUGUGCCC
C15A	C15A	952 381	76	GGGCAGUGAGCAACAAAAUGCUCACCUGGGUCGCAGUAA**a**CCCAGUUAACAAAAC AAGGGAGGUAAUUUUGUGCCC
AU	G5U, G11U, C16A	950 570	76	GGGCAGUGAGCAACAAAAUGCUCACCUGG**u**UCGCA**u**UAAC**a**CCAGUUAACAAAAC AAGGGAGGUAAUUUUGUGCCC
CG	G5C, G11C, C16G, C30G	945 180	76	GGGCAGUGAGCAACAAAAUGCUCACCUGG**c**UCGCA**c**UAAC**g**CCAGUUAACAAAA**g** AAGGGAGGUAAUUUUGUGCCC
UUCG	ΔU6-C15, insert UUCG loop	868 056	70	GGGCAGUGAGCAACAAAAUGCUCACCUGGG**uucg**CCCAGUUAACAAAACAAGGGA GGUAAUUUUGUGCCC
GAAA	ΔU6-C15, insert GAAA loop	884 956	70	GGGCAGUGAGCAACAAAAUGCUCACCUGGG**gaaa**CCCAGUUAACAAAACAAGGGA GGUAAUUUUGUGCCC
Minimal Series:				
Tte^{min}	None (wild-type)	485 437	39	GGUCCUGGGUCGCAGUAACCCCAGUUAACAAAACAAGGG
Tte^{rigid}	C25A (rigid L3)	491 884	39	GGUCCUGGGUCGCAGUAACCCCAGUUAAA**a**AAACAAGGG
Bsu^{min}	None (wild-type, *Bsu* sequence)	487 092	39	GGCAGAGGUUCUAGUUACACCCUCUAUAAAAAACUAAGG
Bsu^{floppy}	A27C (flexible L3)	480 769	39	GGCAGAGGUUCUAGUUACACCCUCUAUAA**c**AAACUAAGG
smFRET construct:				
Tte^{smFRET}	None (wild-type)	442 087	39	biotin-UCACCUGGGUCGCAG(U-Cy5)AACCCCAGUUAACAAAACAAGGG-Dy547

B.3 Quantification of preQ$_1$ and S1 binding experiments

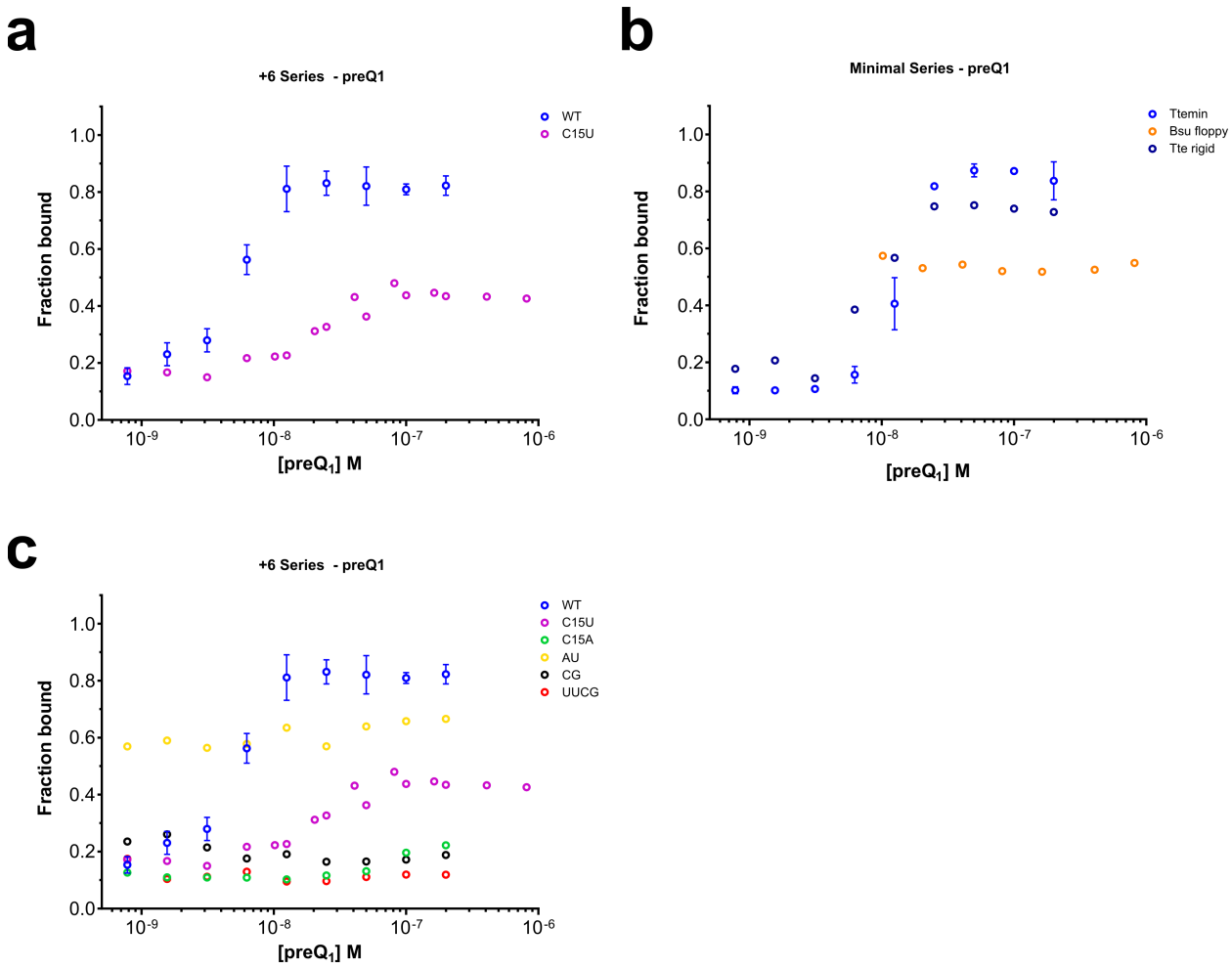


Figure B.3-1 Quantification of EMSA data from preQ$_1$ binding experiments

(**a**) Quantification of +6 series WT and C15U constructs that showed a clear ligand-bound band in preQ$_1$ EMSA experiments. (**b**) Quantification of data for in preQ$_1$ EMSA experiments with minimal series constructs. No clear shift was seen with the addition of preQ$_1$ for the *Bsu*^{floppy} construct. (**c**) Quantification for +6 series constructs. The majority showed no stable binding of preQ$_1$ that results in a mobility shift. Error bars represent the standard deviation of at least 2 independent measurements.

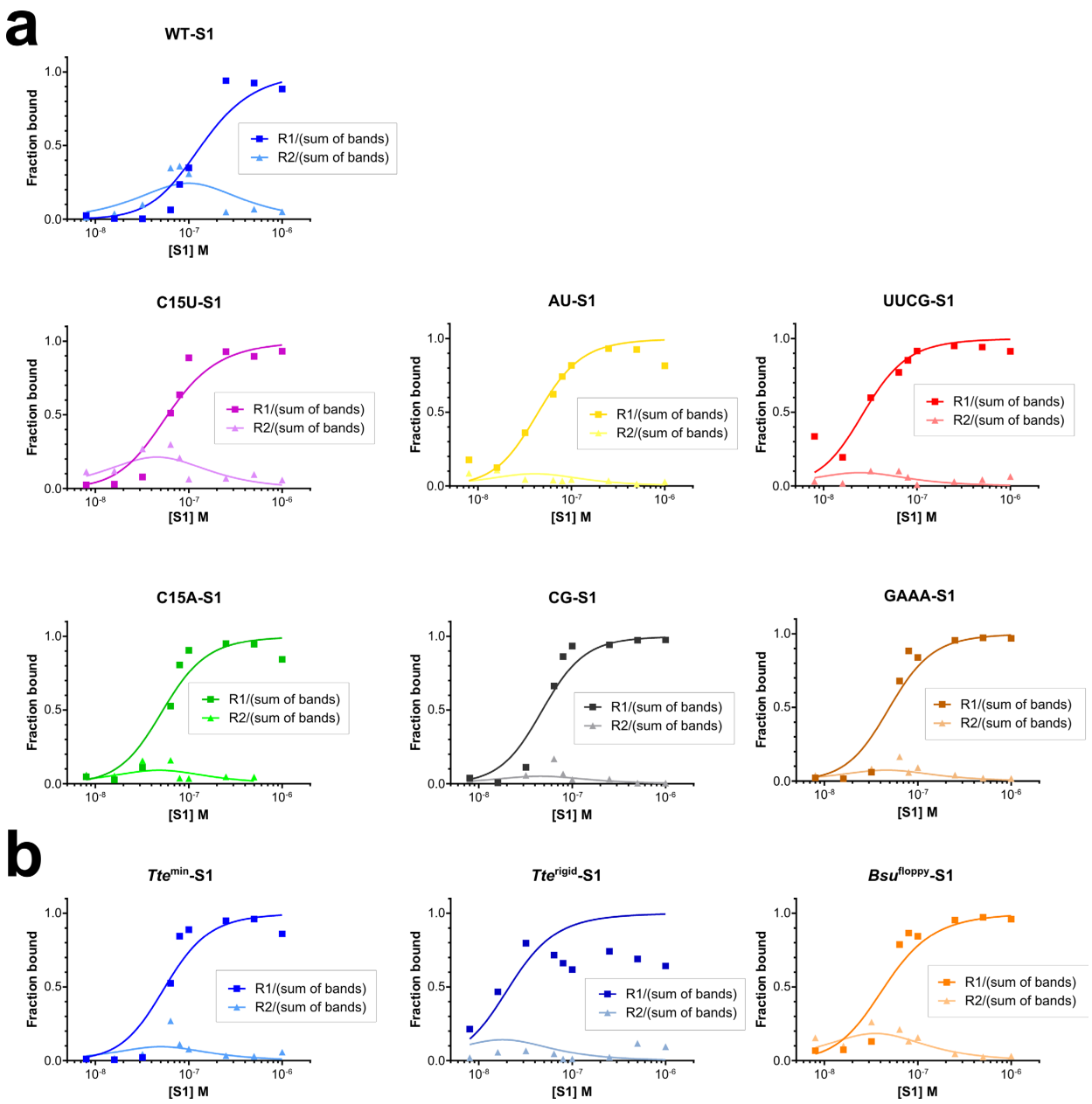


Figure B.3-2 Quantification of EMSA data from S1 binding experiments

Quantification of data for in S1 EMSA experiments with (**a**) +6 series and (**b**) minimal series constructs. Band intensities of S1-RNA complexes were used to calculate the fraction of total RNA bound as a function of S1 concentration, which was then fit using a 2-site binding model as described in **Materials and Methods 3.5.7**. Note that in some cases only a single complex was clearly observed (see main text) and so the fraction bound quantified for Complex 2 represent the intensity in the area of the gel where Complex 2 migrates in variants where a second complex is clearly apparent (R1, band intensity for Complex 1; R2, band intensity for Complex 2; sum of bands, sum of R1, R2, and free RNA band intensities).

B.4 Folding heterogeneity in pseudoknot variants

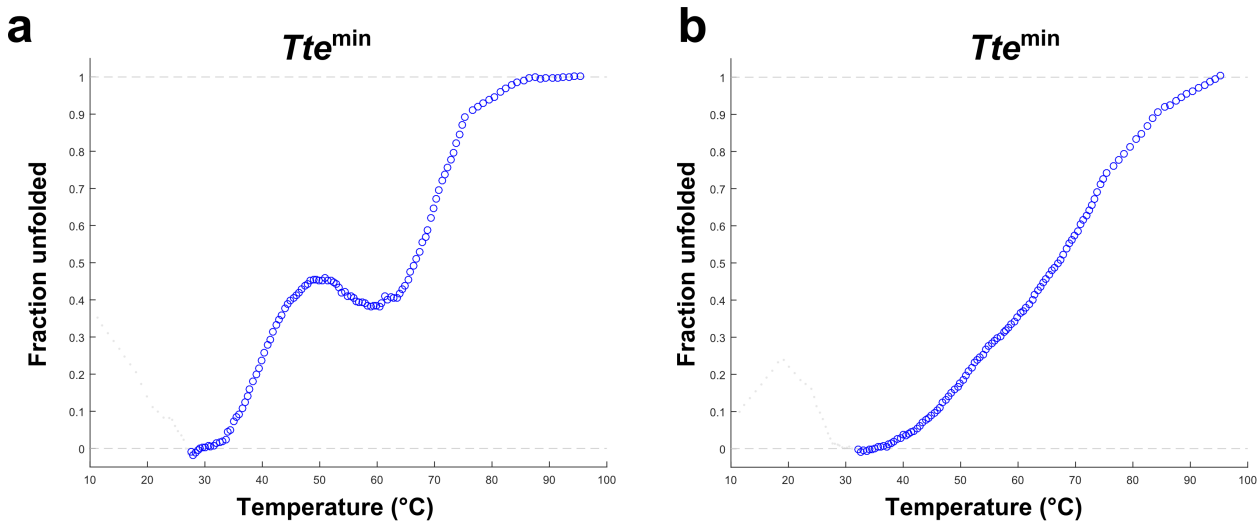


Figure B.4-1 Folding heterogeneity in pseudoknot RNAs.

Initial (**a**) and subsequent (**b**) melting curve profiles obtained for the same sample in the same cuvette. In some (but not all) instances, the melting curve profile obtained in subsequent melting experiments on the sample of pseudoknot RNA show similar transitions as observed initially, but are overall less dramatic. Because the RNA was found to be intact, as assessed by denaturing PAGE, after up to four sequential melting experiments, the differences in melting profiles suggest significant differences in the heterogeneity of the starting fold of the RNA between experiments.

B.5 Preparation E. coli *ribosomal protein S1*

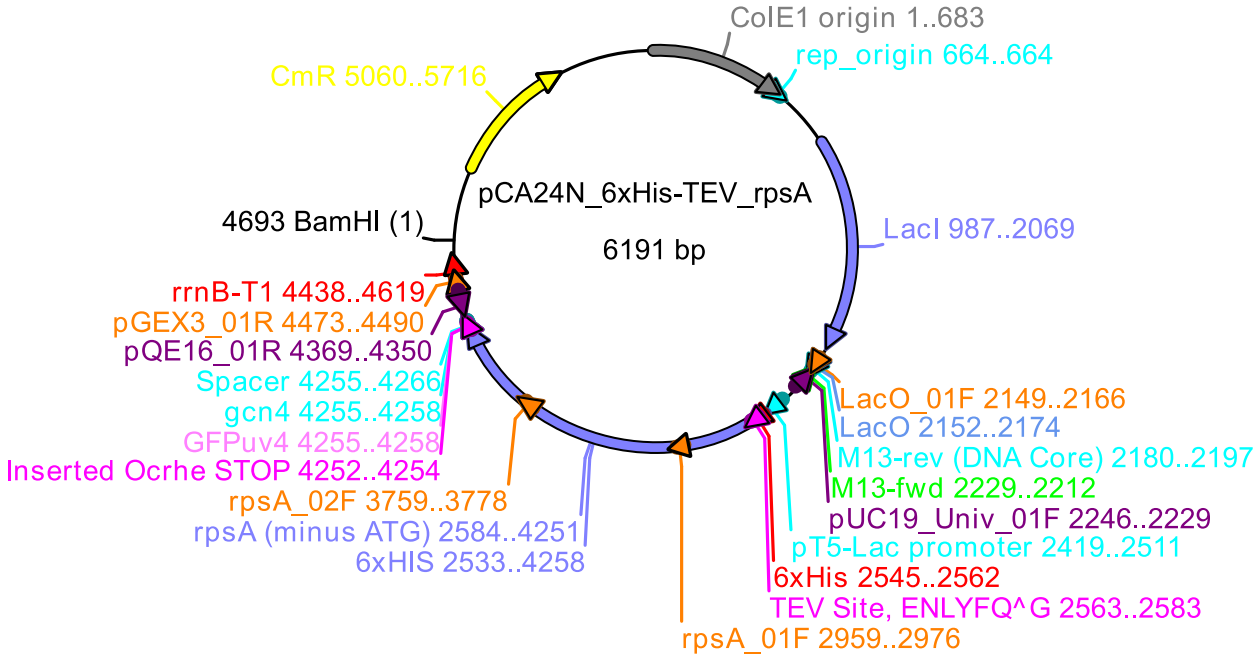
Escherichia coli [gbbct]: 8087 CDS's (2330943 codons)
http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=37762&aa=1&style=N

```
            11122233344455666777
Tln         ThrAspProAlaLeuArgAla
pDNA   5'-..ACGGATCCGGCCCTGAGGGCC..-3'
New Seq     GAGAATCTGTATTTTCAGGGC   Matching base, optimal codon
Tln         GluAsnLeuTyrPheGlnGly   Matching base, sub-optimal codon
            GAAAACCTGTATTTTCAGGGC   Optimal codon usage
```

Figure B.5-1 Selection of the sequence encoding the TEV cleavage site

The 7 amino-acid linker sequence in the pCA24N_6xHis-rpsA plasmid (ASKA(-) clone JW0894) was mutated to the TEV protease recognition sequence ENLYFQ^G based on which degenerate codons resulted in the fewest required number of base changes while still employing codons that appear with high frequency in the ORFeome.

Figure B.5-2 Plasmid map and sequence information for S1 expression plasmid




```
LOCUS       pCA24N_6xHis_TEV        6191 bp ds-DNA     circular     01-JUL-2013
DEFINITION  Cloning vector pCA24N DNA, complete genome.
ACCESSION   AB052891
VERSION     AB052891.2  GI:63147361
KEYWORDS    .
SOURCE      Cloning vector pCA24N
  ORGANISM  Cloning vector pCA24N other sequences; artificial sequences;
            vectors.
REFERENCE   1
  AUTHORS   Kitagawa,M.
  TITLE     Archive Vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 5240)
  AUTHORS   Kitagawa,M., Hirai,A. and Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2000) Hirotada Mori, Nara Institute of Science &
            Technology, Research & Education Center of Genetic Information;
            Takayama 8916-5, Ikoma, Nara 630-0101, Japan
            (E-mail:hmori@gtc.naist.jp, Tel:81-743-72-5662, Fax:81-743-72-5669)
COMMENT     Expected sequence after SDM using  ASKA-rpsA_inTEV_01R  and
            ASKA-rpsA_inOchre_01F to insert TEV cleavage site and earlier stop
            codon.
COMMENT
COMMENT
COMMENT     pCA24N_rpsA from 1 to 6191
COMMENT     On May 10, 2005 this sequence version replaced gi:12082329.
COMMENT
COMMENT     ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     promoter        2419..2511
                     /standard_name="p-T5-lac"
                     /label=pT5-Lac promoter
```

```
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             join(2533..2562,4255..4258)
                /gene="his-tag"
                /codon_start=1
                /transl_table=11
                /product="6xHIS"
                /protein_id="BAB20664.1"
                /db_xref="GI:12082330"
                /translation="MRGSHHHHHHTDPALRPMRPQ"
                /label=6xHIS
                /ApEinfo_fwdcolor=#8080ff
                /ApEinfo_revcolor=#ffff00
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
gene            4255..4258
                /gene="gcn4"
                /label=gcn4
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
misc_feature    4255..4258
                /gene="gcn4"
                /product="GFPuv4"
                /label=GFPuv4
                /ApEinfo_fwdcolor=#ff80ff
                /ApEinfo_revcolor=#690f96
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
terminator      4438..4619
                /gene="rrnB-T1"
                /label=rrnB-T1
                /ApEinfo_fwdcolor=#ff0000
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             5060..5719
                /gene="cat"
                /codon_start=1
                /transl_table=11
                /product="CAT"
                /protein_id="BAB20666.1"
                /db_xref="GI:12082332"
                /translation="MEKKITGYTTVDISQWHRKEHFEAFQSVAQCTYNQTVQLDITAF
                LKTVKKNKHKFYPAFIHILARLMNAHPEFRMAMKDGELVIWDSVHPCYTVFHEQTETF
                SSLWSEYHDDFRQFLHIYSQDVACYGENLAYFPKGFIENMFFVSANPWVSFTSFDLNV
                ANMDNFFAPVFTMGKYYTQGDKVLMPLAIQVHHAVCDGFHVGRMLNELQQYCDEWQGG
                A"
                /label=CAT
                /ApEinfo_fwdcolor=#8080ff
                /ApEinfo_revcolor=#ffff00
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
rep_origin      664..664
                /label=rep_origin
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
                /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                width 5 offset 0
CDS             987..2069
```

```
                 /gene="lacI^q"
                 /codon_start=1
                 /transl_table=11
                 /product="LacI"
                 /protein_id="BAB20667.1"
                 /db_xref="GI:12082333"
                 /translation="MKPVTLYDVAEYAGVSYQTVSRVVNQASHVSAKTREKVEAAMAE
                 LNYIPNRVAQQLAGKQSLLIGVATSSLALHAPSQIVAAIKSRADQLGASVVVSMVERS
                 GVEACKAAVHNLLAQRVSGLIINYPLDDQDAIAVEAACTNVPALFLDVSDQTPINSII
                 FSHEDGTRLGVEHLVALGHQQIALLAGPLSSVSARLRLAGWHKYLTRNQIQPIAEREG
                 DWSAMSGFQQTMQMLNEGIVPTAMLVANDQMALGAMRAITESGLRVGADISVVGYDDT
                 EDSSCYIPPLTTIKQDFRLLGQTSVDRLLQLSQGQAVKGNQLLPVSLVKRKTTLAPNT
                 QTASPRALADSLMQLARQVSRLESGQ"
                 /label=LacI
                 /ApEinfo_fwdcolor=#8080ff
                 /ApEinfo_revcolor=#ffff00
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      2180..2197
                 /label=M13-rev (DNA Core)
                 /ApEinfo_fwdcolor=cyan
                 /ApEinfo_revcolor=green
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      complement(2212..2229)
                 /label=M13-fwd
                 /ApEinfo_fwdcolor=cyan
                 /ApEinfo_revcolor=green
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
rep_origin       1..683
                 /label=ColE1 origin
                 /ApEinfo_fwdcolor=gray50
                 /ApEinfo_revcolor=gray50
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
misc_binding     2152..2174
                 /label=LacO
                 /ApEinfo_fwdcolor=#6495ed
                 /ApEinfo_revcolor=#6495ed
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
CDS              5060..5716
                 /label=CmR
                 /ApEinfo_fwdcolor=yellow
                 /ApEinfo_revcolor=yellow
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      4473..4490
                 /label=pGEX3_01R
                 /ApEinfo_fwdcolor=#ff8000
                 /ApEinfo_revcolor=#800080
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      complement(2229..2246)
                 /label=pUC19_Univ_01F
                 /ApEinfo_fwdcolor=#ff8000
                 /ApEinfo_revcolor=#800080
                 /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                 width 5 offset 0
primer_bind      2149..2166
                 /label=LacO_01F
                 /ApEinfo_fwdcolor=#ff8000
```

```
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     CDS             2584..4251
                     /label=rpsA (minus ATG)
                     /ApEinfo_fwdcolor=#8080ff
                     /ApEinfo_revcolor=#ffff00
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    4255..4266
                     /label=Spacer
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    2545..2562
                     /label=6xHis
                     /ApEinfo_fwdcolor=#ff0000
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_binding    2563..2583
                     /label=TEV Site, ENLYFQ^G
                     /ApEinfo_fwdcolor=#ff00ff
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    4252..4254
                     /label=Inserted Ocrhe STOP
                     /ApEinfo_fwdcolor=#ff00ff
                     /ApEinfo_revcolor=green
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     misc_feature    complement(4350..4369)
                     /label=pQE16_01R
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     2959..2976
                     /label=rpsA_01F
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
     primer_bind     3759..3778
                     /label=rpsA_02F
                     /ApEinfo_fwdcolor=#ff8000
                     /ApEinfo_revcolor=#800080
                     /ApEinfo_graphicformat=arrow_data {{0 1 2 0 0 -1} {} 0}
                     width 5 offset 0
ORIGIN
        1 tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac cccgtagaaa
       61 agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc ttgcaaacaa
      121 aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca actctttttc
      181 cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta gtgtagccgt
      241 agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct ctgctaatcc
      301 tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg gactcaagac
      361 gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc acacagccca
      421 gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcta tgagaaagcg
      481 ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg gtcggaacag
      541 gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt cctgtcgggt
      601 ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcaggggggg cggagcctat
```

661 ggaaaaacgc cagcaacgcg gcctttttac ggttcctggc cttttgctgg ccttttgctc
721 acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc gcctttgagt
781 gagctgatac cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg agcgaggaag
841 cggaagagcg cctgatgcgg tattttctcc ttacgcatct gtgcggtatt tcacaccgca
901 taaattccga caccatcgaa tggtgcaaaa cctttcgcgg tatggcatga tagcgcccgg
961 aagagagtca attcagggtg gtgaatgtga aaccagtaac gttatacgat gtcgcagagt
1021 atgccggtgt ctcttatcag accgtttccc gcgtggtgaa ccaggccagc cacgtttctg
1081 cgaaaacgcg ggaaaaagtg gaagcggcga tggcggagct gaattacatt cccaaccgcg
1141 tggcacaaca actggcgggc aaacagtcgt tgctgattgg cgttgccacc tccagtctgg
1201 ccctgcacgc gccgtcgcaa attgtcgcgg cgattaaatc tcgcgccgat caactgggtg
1261 ccagcgtggt ggtgtcgatg gtagaacgaa gcggcgtcga agcctgtaaa gcggcggtgc
1321 acaatcttct cgcgcaacgc gtcagtgggc tgatcattaa ctatccgctg gatgaccagg
1381 atgccattgc tgtggaagct gcctgcacta atgttccggc gttatttctt gatgtctctg
1441 accagacacc catcaacagt attattttct cccatgaaga cggtacgcga ctgggcgtgg
1501 agcatctggt cgcattgggt caccagcaaa tcgcgctgtt agcgggccca ttaagttctg
1561 tctcggcgcg tctgcgtctg gctggctggc ataaatatct cactcgcaat caaattcagc
1621 cgatagcgga acgggaaggc gactggagtg ccatgtccgg ttttcaacaa accatgcaaa
1681 tgctgaatga gggcatcgtt cccactgcga tgctggttgc caacgatcag atggcgctgg
1741 gcgcaatgcg cgccattacc gagtccgggc tgcgcgttgg tgcggatatc tcggtagtgg
1801 gatacgacga taccgaagac agctcatgtt atatcccgcc gttaaccacc atcaaacagg
1861 attttcgcct gctggggcaa accagcgtgg accgcttgct gcaactctct cagggccagg
1921 cggtgaaggg caatcagctg ttgcccgtct cactggtgaa aagaaaaacc accctggcgc
1981 ccaatacgca aaccgcctct ccccgcgcgt tggccgattc attaatgcag ctggcacgac
2041 aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag ttagctcact
2101 cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg tggaattgtg
2161 agcggataac aatttcacac aggaaacagc tatgaccatg attacggatt cactggccgt
2221 cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc
2281 acatccccct ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca
2341 acagttgcct aagaaaccat tattatcatg acattaacct ataaaaatag gcgtatcacg
2401 aggccctttc gtcttcacct cgagaaatca taaaaaattt atttgctttg tgagcggata
2461 acaattataa tagattcaat tgtgagcgga taacaatttc acacagaatt cattaaagag
2521 gagaaattaa ctatgagagg atctcaccat caccatcacc atGAGAATCT GTATTTTCAG
2581 GGCACTGAAT CTTTTGCTCA ACTCTTTGAA GAGTCCTTAA AAGAAATCGA AACCCGCCCG
2641 GGTTCTATCG TTCGTGGCGT TGTTGTTGCT ATCGACAAAG ACGTAGTACT GGTTGACGCT
2701 GGTCTGAAAT CTGAGTCCGC CATCCCGGCT GAGCAGTTCA AAAACGCCCA GGGCGAGCTG
2761 GAAATCCAGG TAGGTGACGA AGTTGACGTT GCTCTGGACG CAGTAGAAGA CGGCTTCGGT
2821 GAAACTCTGC TGTCCCGTGA GAAAGCTAAA CGTCACGAAG CCTGGATCAC GCTGGAAAAA
2881 GCTTACGAAG ATGCTGAAAC TGTTACCGGT GTTATCAACG GCAAAGTTAA GGGCGGCTTC
2941 ACTGTTGAGC TGAACGGTAT TCGTGCGTTC CTGCCAGGTT CTCTGGTAGA CGTTCGTCCG
3001 GTGCGTGACA CTCTGCACCT GGAAGGCAAA GAGCTTGAAT TTAAAGTAAT CAAGCTGGAT
3061 CAGAAGCGCA ACAACGTTGT TGTTTCTCGT CGTGCCGTTA TCGAATCCGA AAACAGCGCA
3121 GAGCGCGATC AGCTGCTGGA AAACCTGCAG GAAGGCATGG AAGTTAAAGG TATCGTTAAG
3181 AACCTCACTG ACTACGGTGC ATTCGTTGAT CTGGGCGGCG TTGACGGCCT GCTGCACATC
3241 ACTGACATGG CCTGGAAACG CGTTAAGCAT CCGAGCGAAA TCGTCAACGT GGGCGACGAA
3301 ATCACTGTTA AAGTGCTGAA GTTCGACCGC GAACGTACCC GTGTATCCCT GGGCCTGAAA
3361 CAGCTGGGCG AAGATCCGTG GGTAGCTATC GCTAAACGTT ATCCGGAAGG TACCAAACTG
3421 ACTGGTCGCG TGACCAACCT GACCGACTAC GGCTGCTTCG TTGAAATCGA AGAAGGCGTT
3481 GAAGGCCTGG TACACGTTTC CGAAATGGAC TGGACCAACA AAAACATCCA CCCGTCCAAA
3541 GTTGTTAACG TTGGCGATGT AGTGGAAGTT ATGGTTCTGG ATATCGACGA AGAACGTCGT
3601 CGTATCTCCC TGGGTCTGAA ACAGTGCAAA GCTAACCCGT GGCAGCAGTT CGCGGAAACC
3661 CACAACAAGG GCGACCGTGT TGAAGGTAAA ATCAAGTCTA TCACTGACTT CGGTATCTTC
3721 ATCGGCTTGG ACGGCGGCAT CGACGGCCTG GTTCACCTGT CTGACATCTC CTGGAACGTT
3781 GCAGGCGAAG AAGCAGTTCG TGAATACAAA AAAGGCGACG AAATCGCTGC AGTTGTTCTG
3841 CAGGTTGACG CAGAACGTGA ACGTATCTCC CTGGGCGTTA AACAGCTCGC AGAAGATCCG
3901 TTCAACAACT GGGTTGCTCT GAACAAGAAA GGCGCTATCG TAACCGGTAA AGTAACTGCA
3961 GTTGACGCTA AAGGCGCAAC CGTAGAACTG GCTGACGGCG TTGAAGGTTA CCTGCGTGCT
4021 TCTGAAGCAT CCCGTGACCG CGTTGAAGAC GCTACCCTGG TTCTGAGCGT TGGCGACGAA
4081 GTTGAAGCTA AATTCACCGG CGTTGATCGT AAAAACCGCG CAATCAGCCT GTCTGTTCGT
4141 GCGAAAGACG AAGCTGACGA GAAAGATGCA ATCGCAACTG TTAACAAACA GGAAGATGCA
4201 AACTTCTCCA ACAACGCAAT GGCTGAAGCT TTCAAAGCAG CTAAAGGCGA GTAActatgc
4261 ggccgctaag ggtcgacctg cagccaagct taattagctg agcttggact cctgttgata
4321 gatccagtaa tgacctcaga actccatctg gatttgttca gaacgctcgg ttgccgccgg
4381 gcgtttttta ttggtgagaa tccaagctag gggaattcat cgtgactgac tgacgatctg

```
4441 cctcgcgcgt ttcggtgatg acggtgaaaa cctctgacac atgcagctcc cggagacggt
4501 cacagcttgt ctgtaagcgg atgccgggag cagacaagcc cgtcagggcg cgtcagcggg
4561 tgttggcggg tgtcggggcg cagccatgac ccagtcacgt agcgatagcg gagtgtataa
4621 ttcttgaaga cgaaagggcc tcgtgatacg cctattttta taggttaatg tcatgataat
4681 aatggtttct taggatcctg atgtccggcg gtgcttttgc cgttacgcac caccccgtca
4741 gtagctgaac aggagggaca gctgatagaa acagaagcca ctggagcacc tcaaaaacac
4801 catcatacac taaatcagta agttggcagc atcacccgac gcactttgcg ccgaataaat
4861 acctgtgacg gaagatcact tcgcagaata aataaatcct ggtgtccctg ttgataccgg
4921 gaagccctgg gccaactttt ggcgaaaatg agacgttgat cggcacgtaa gaggttccaa
4981 ctttcaccat aatgaaataa gatcactacc gggcgtattt tttgagttat cgagattttc
5041 aggagctaag gaagctaaaa tggagaaaaa aatcactgga tataccaccg ttgatatatc
5101 ccaatggcat cgtaaagaac attttgaggc atttcagtca gttgctcaat gtacctataa
5161 ccagaccgtt cagctggata ttacggcctt tttaaagacc gtaaagaaaa ataagcacaa
5221 gttttatccg gcctttattc acattcttgc ccgcctgatg aatgctcatc cggaattccg
5281 tatggcaatg aaagacggtg agctggtgat atgggatagt gttcaccctt gttacaccgt
5341 tttccatgag caaactgaaa cgttttcatc gctctggagt gaataccacg acgatttccg
5401 gcagtttcta cacatatatt cgcaagatgt ggcgtgttac ggtgaaaacc tggcctattt
5461 ccctaaaggg tttattgaga atatgttttt cgtctcagcc aatccctggg tgagtttcac
5521 cagttttgat ttaaacgtgg ccaatatgga caacttcttc gcccccgttt tcaccatggg
5581 caaatattat acgcaaggcg acaaggtgct gatgccgctg gcgattcagg ttcatcatgc
5641 cgtctgtgat ggcttccatg tcggcagaat gcttaatgaa ttacaacagt actgcgatga
5701 gtggcagggc ggggcgtaat ttttttaagg cagttattgg tgcccttaaa cgcctggtgc
5761 tacgcctgaa taagtgataa taagcggatg aatggcagaa attcgaaagc aaattcgacc
5821 cggtcgtcgg ttcagggcag ggtcgttaaa tagccgctta tgtctattgc tggtttaccg
5881 gtttattgac taccggaagc agtgtgaccg tgtgcttctc aaatgcctga ggccagtttg
5941 ctcaggctct ccccgtggag gtaataattg acgatatgat catttattct gcctcccaga
6001 gcctgataaa aacggttagg atcggagtca ggcaactatg gatgaacgaa atagacagat
6061 cgctgagata ggtgcctcac tgattaagca ttggtaactg tcagaccaag tttactcata
6121 tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg tgaagatcct
6181 ttttgataat c
//
```

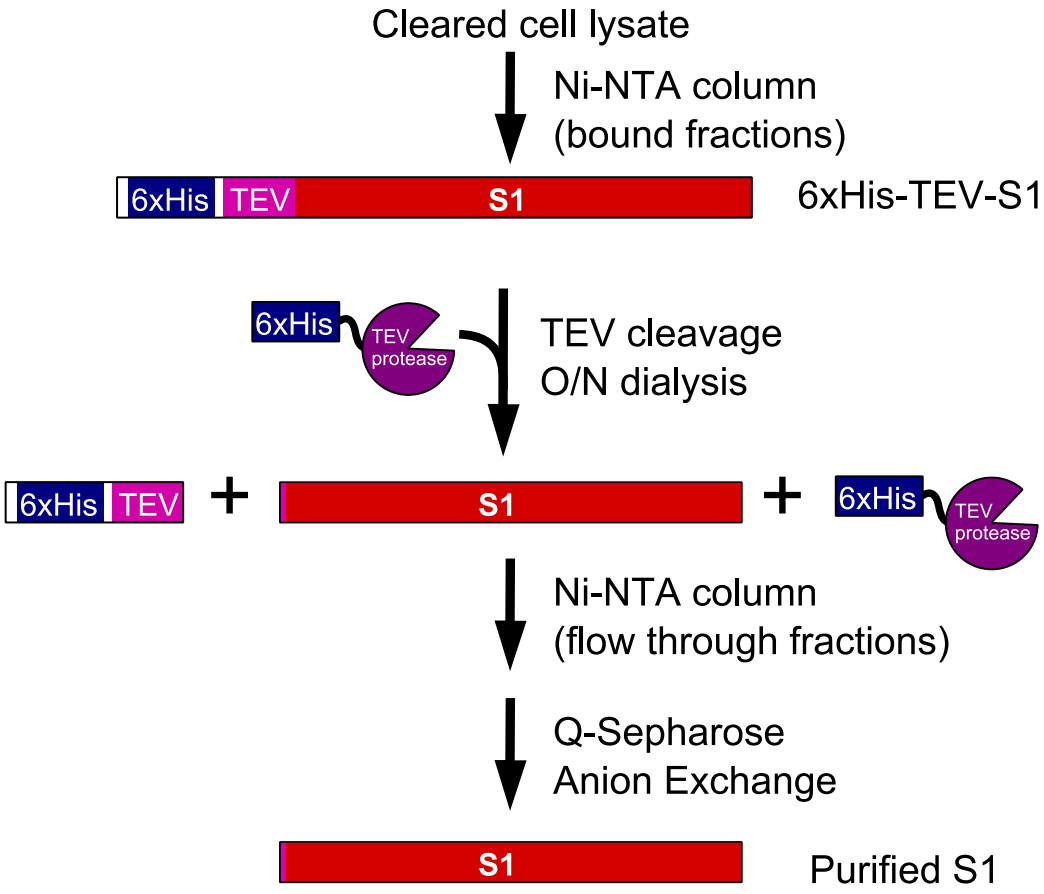


Figure B.5-3 Overview of S1 purification scheme

An N-terminally His-tagged variant of *E. coli* S1 protein is overexpressed in the BLR(DE3) strain of *E. coli* and purified over a nickel-affinity column. The N-terminal His-tag is cleaved during overnight dialysis by a TEV protease that is also His-tagged. After TEV cleavage, a glycine residue replaces the native methionine as the first amino acid in the protein (pink portion at the S1 N-terminus).The His-tag fragment and TEV protease are removed via a second nickel-affinity column, and the S1 protein present in the flow through is further purified of contaminants by anion exchange column chromatography.

B.6 Chromatographic separation and heterogeneity of RNA construct for smFRET experiments

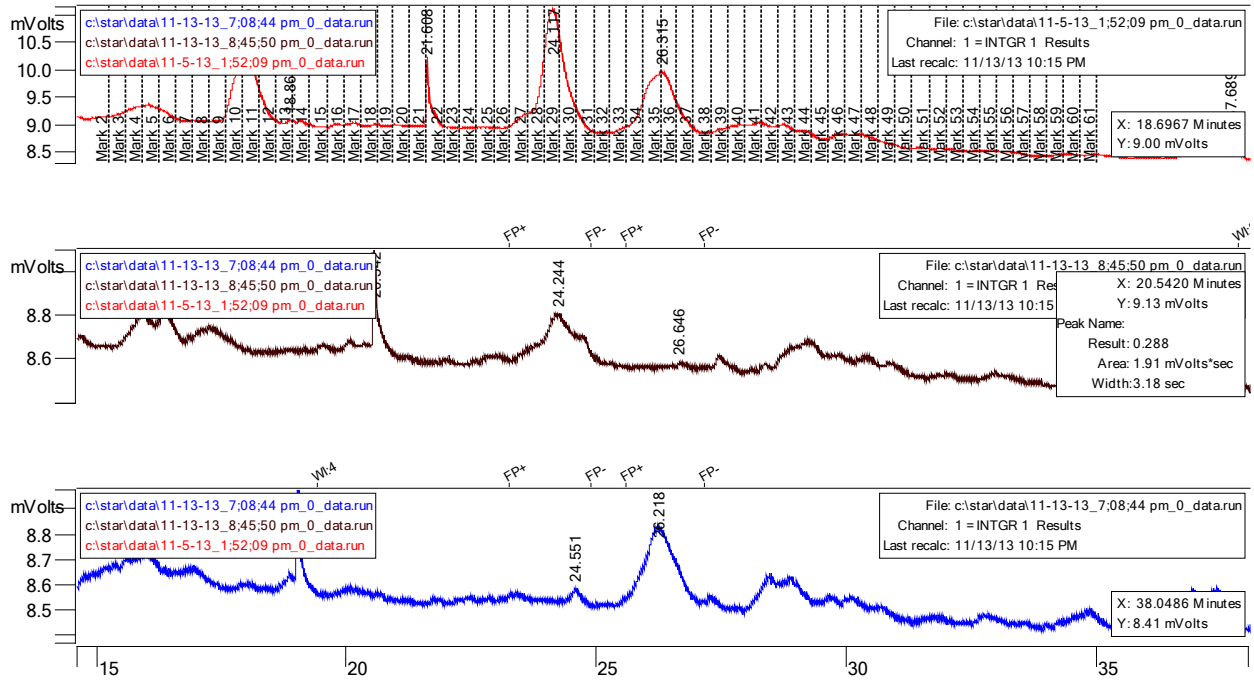


Figure B.6-1 HPLC chromatograms for *Tte*$^{\text{Dy547 only}}$

The singly-labeled RNA purchased from Dharmacon (*Tte*$^{\text{Dy547 only}}$) was deprotected according to the manufacturer's instruction and then further purified using reverse-phase HPLC on a Sunfire C_{18} column (Waters Corporation) with 0.1 M triethylamineacetate (buffer A) and acetonitrile (buffer B) as solvents, monitored at 260 nm. Approximately 5 µg of RNA was prepared in 40% formamide in milliQ water and heated for 1 min 40 sec in a 90 °C copper bead bath, then snap cooled on ice and used to make a HPLC single injection onto the column pre-equilibrated in buffer A. The material was eluted using a linear gradient (**Appendix Table B.6-1**) of 0 – 60% B over 50 min with a flow rate of 1 mL/min (red trace). Degraded RNA eluted at 18 min. *Tte*$^{\text{Dy547 only}}$ eluted in two peaks, with retention times of 24 min and 26 min. Fractions corresponding to the 24 min and 26 min peaks were pooled separately and brought to dryness under vacuum. The material was resuspended in water and re-injected using the same purification conditions. Elution profiles from the 24 min peak (black trace) and 26 min (blue trace) material did not redistribute into two peaks, but rather maintained their characteristic retention times, suggesting a persistent chemical or conformational difference, despite being heated and snap cooled in the presence of 40% formamide before injection.

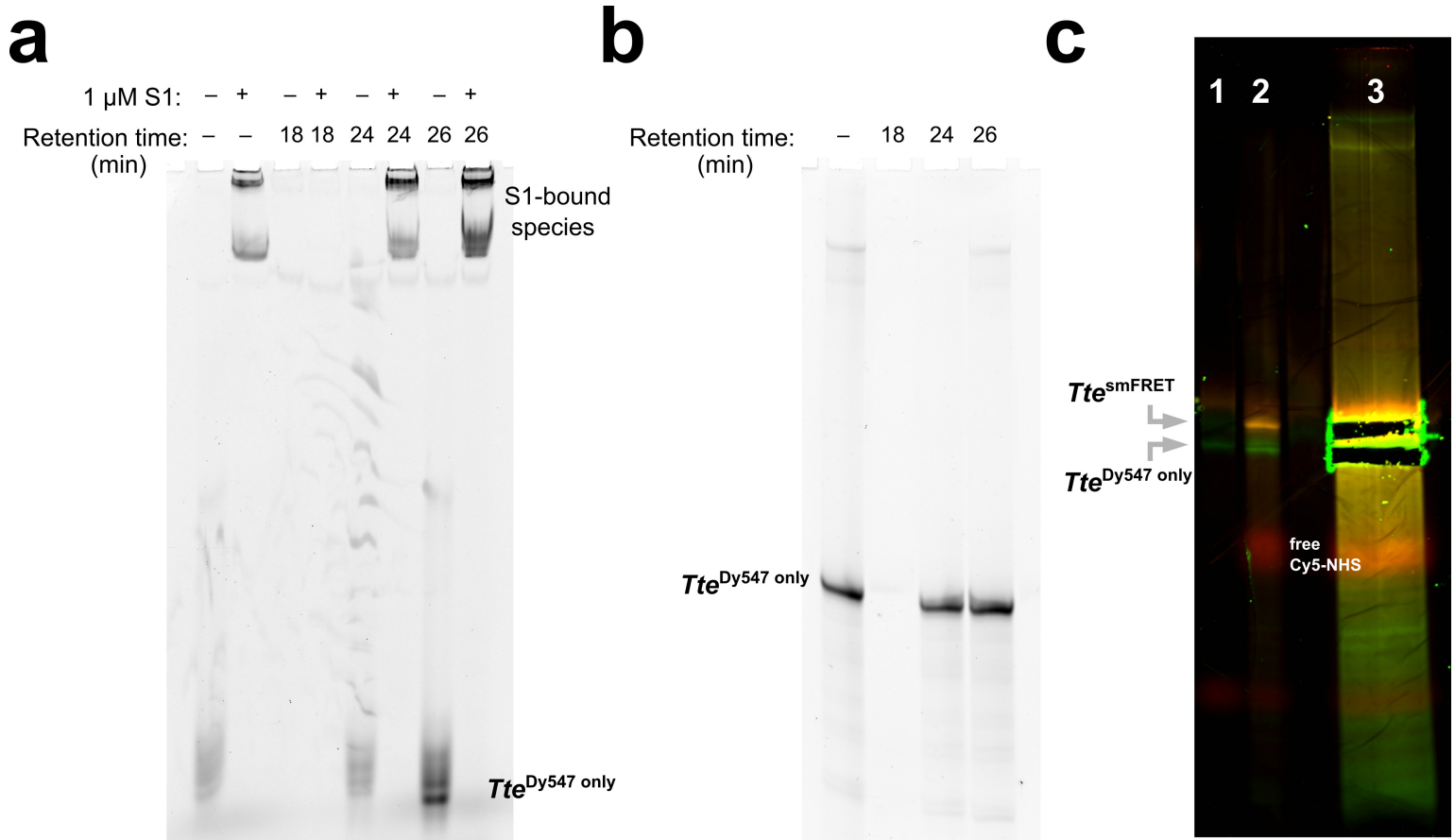


Figure B.6-2 Conformational characterization of *Tte*$^{\text{Dy547 only}}$ and *Tte*^{smFRET} RNAs

(**a**) The conformational heterogeneity of material that elutes with different characteristic HPLC retention times was assessed on a 12% native polyacrylamide gel (18 cm × 14.5 cm × 1.5 mm) in 1X TGE. Binding reactions containing 1 μL of 0.1 μM *Tte*$^{\text{Dy574 only}}$ RNA that had not been HPLC purified (–), or that eluted at 18, 24, or 26 min, were assembled essentially as described in **Materials and Methods 3.5.7** in the absence of preQ$_1$, and incubated in the presence or absence of 1 μM S1. Dy547 fluorescence was imaged using a Typhoon 9410 Variable Mode imager (Typhoon) as described. No signal was detected from the material eluting at 18 min, indicating that these RNA have degraded and lost the 3' Dy547 fluorophore. Multiple bands are observed for the 24 and 26 min peaks, however the 26 min peak has a more compact folded form (fastest migrating band) that

Figure B.6-2 Conformational characterization of *Tte*^{Dy547} only *and* *Tte*^{smFRET} RNAs (continued)

is absent in the 24 min peak. All RNA species are bound equally by S1. (**b**) The same stock RNA solutions used to prepare the binding reactions in (**a**) were electrophoresed on a 20% Urea-PAGE gel and imaged by Typhoon. A single band in the (–), 24 min and 26 min lanes indicates that the RNA is intact in each sample, and that the multiple bands observed in the native gel are due to stably folded conformations that are only disrupted under strongly denaturing conditions. An alternative explanation is that the RNA that elutes at different times by HPLC contain chemical differences that result in conformational differences but are too subtle to alter the RNA's migration under denaturing conditions. (**c**) Doubly fluorophore-labeled material prepared separately from 24 min and 26 min peaks was pooled and gel-purified on a 20% Urea-PAGE gel as described in **Materials and Methods 3.5.8**. The gel was imaged by Typhoon using both the Cy3 and Cy5 excitation/emission filter sets after the bands were cut from the gel to confirm successful isolation of singly from doubly fluorophore material (lane 3). Lane 1: Singly fluorophore-labeled material from the 26 min peak for reference, Lane 2: pooled doubly fluorophore-labeled material for reference.

Appendix Table B.6-1 HPLC Methods for purification of fluorophore-labeled RNAs.

The following HPLC methods were used with the *Tte*$^{DY547\ only}$ and *Tte*^{smFRET} RNAs discussed in this study. The gradient parameters are written exactly as entered in Method Builder, part of the Star chromatography workstation software, used with a ProStar chromatography system (Varian). Absorbance was monitored at 260 nm. HPLC solvents: (A) 0.1 M trimethylamine acetate, pH 7.0 (aqueous); B: acetonitrile (organic).

Method	Time (min:sec)	%A	%B	Flow rate (mL/min)
C18 for fluorophores2_slowEnd.mth	0:00	100	0	1
	25:00	70	30	1
	50:00	40	60	1
	55:00	0	100	1
	65:00	0	100	1
	75:00	100	0	1
	85:00	100	0	1
	86:00	100	0	0
C18_35minRamp_slowEnd.mth	0:00	100	0	1
	25:00	70	30	1
	35:00	58	42	1
	42:00	0	100	1
	55:00	0	100	1
	65:00	100	0	1
	75:00	100	0	1
	76:00	100	0	0

B.7 Procedure for quantification of gel images using ImageQuant v5.2

The following is a general step-wise protocol for quantification of *.gel files (generated by the Typhon™ 9410 variable mode imager) using the ImageQuant software.

Check Dynamic Range

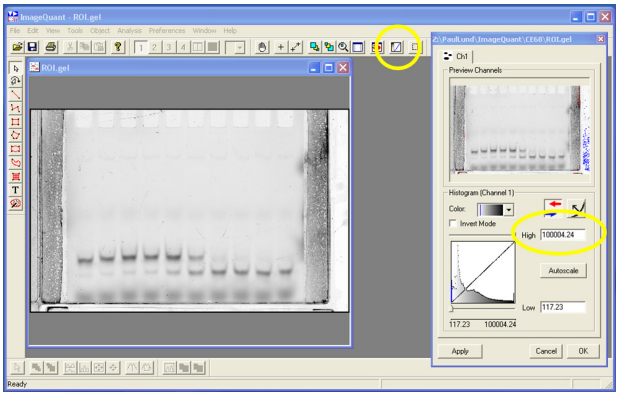
The dynamic range of the instrument is between 1-100,000 counts. Do the following to check whether your scan saturated the detector. If so, rescan with a lower PMT voltage.

1) Open the *.gel file in ImageQuant.

2) From the top menu, click on the Gray/Color Adjust button to view the pixel intensity distributions and thresholds.

   - If the High value is less than 100,000 move on to next section

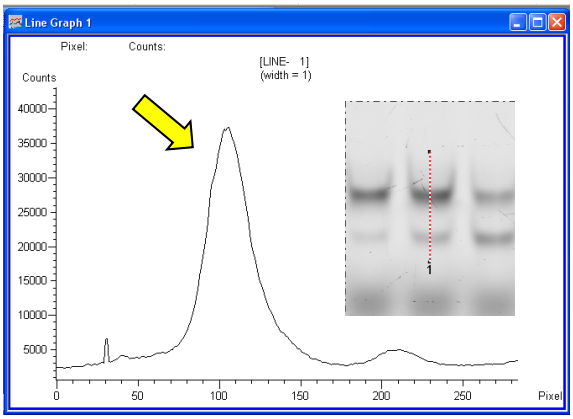
   - If High value is 100,000 or greater, determine whether bands of interest are saturated

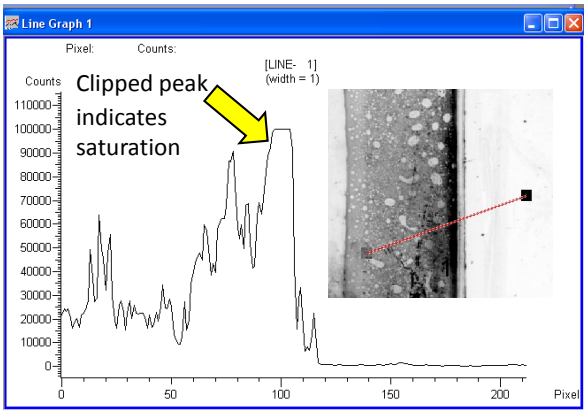


3) Draw a line through the darkest object and view its intensity profile by clicking  from the bottom toolbar.

   - If the darkest of your bands of interest show peak counts <100,000, the saturated portion of the image is likely something unimportant (e.g., a spacer).

   - If the intensity profile shows clipped peaks, then you've saturated the detector and should rescan with a lower PMT voltage.



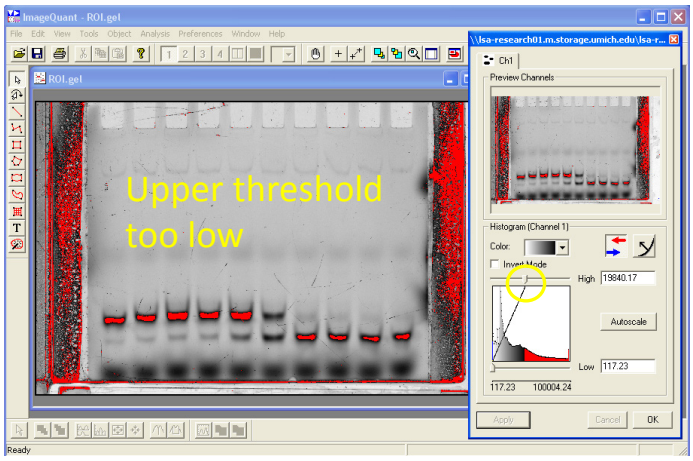


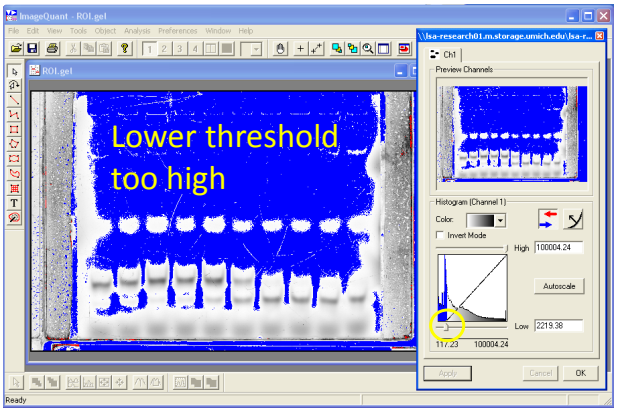
Adjust contrast

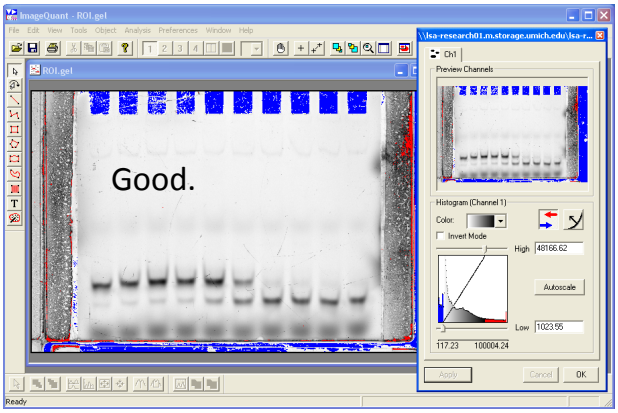
Adjusting the contrast does not change the quantification, it only makes it easier to box the bands and present a more visually pleasing image.

1) Adjust the upper and lower threshold value sliders in the histogram window and click Apply. Blue and Red areas show pixels that are getting clipped to white or to black, respectively with the current threshold settings.

2) Once the adjustment is finished, click OK.





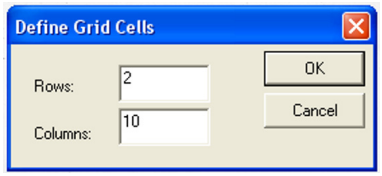


Box bands

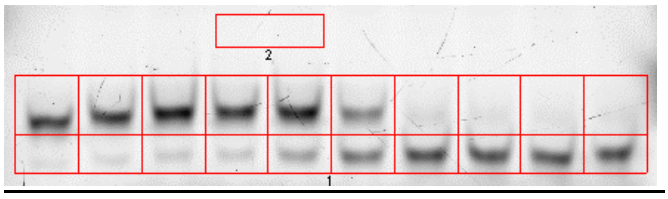
1) The Grid tool is easiest to use. Click  from the left toolbar and define the dimensions of your grid.

- Include additional rows if there is large vertical spacing between bands in a single lane

- Include columns for empty lanes



2) Draw a grid that spaces all lanes evenly, then adjust the row boundaries. Once the grid is properly adjusted, draw a rectangle in the area of the gel that is representative of background intensity.



Quantify bands

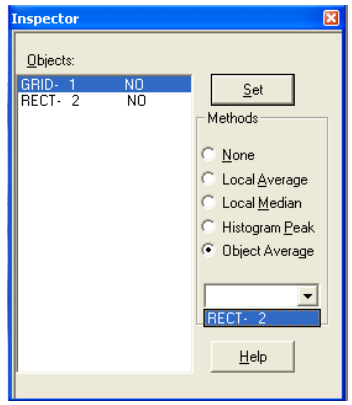
The method described below does not require equally sized boxes for each band, or that the boxes tightly fit the bands, so long as an appropriate background value is chosen and the background is even across lanes.

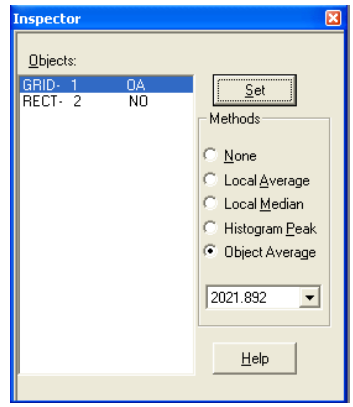
1) From the top menu bar, select Analysis> Background Correction….

2) Highlight the correct grid number and choose Object Average as the Method for background correction.

3) Select the appropriate rectangle that represents background for the corresponding grid.

4) Lastly, click set. The NO (None) next to GRID- # should change to OA (Object Average). The average intensity for the selected Object i.e. rectangle is displayed in the drop down box. Close the inspector window.





5) From the top menu bar, select Analysis > Auto Volume Report. Double click on the spreadsheet window to open and save the quantification in excel.

** The Volume Report contains a lot of values. Grid boxes are named by row (R) and column (C) from the top left. The Volume is calculated by summing (pixel value – background value) for all pixels in the box. This can be inaccurate if the background in the box contains pixels that have lower intensity than the background value (e.g., white scratches). This will result in pixels that contribute negative volume to the total volume.*

6) Instead, calculate the non-negative volume as shown:

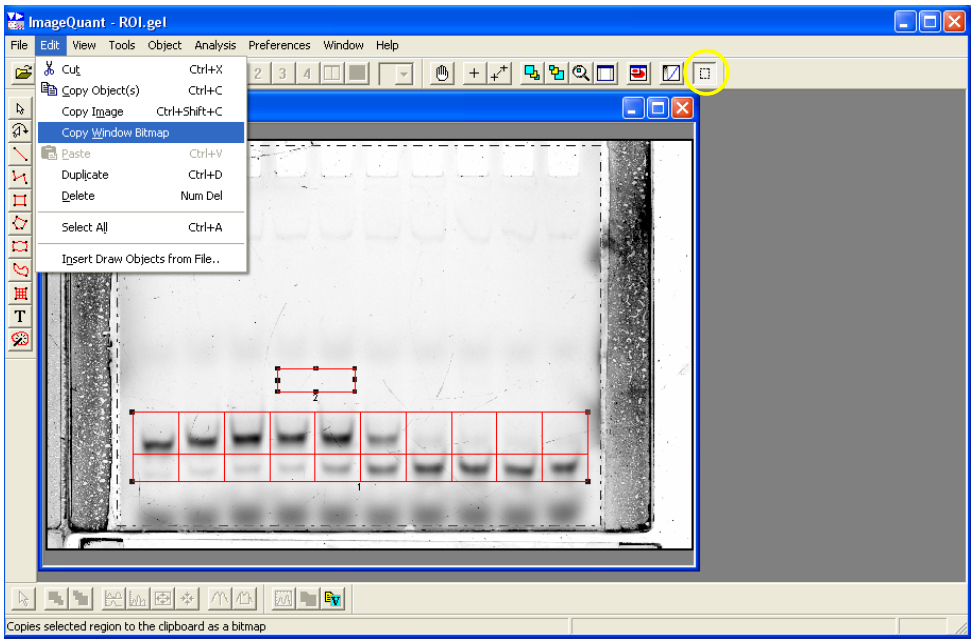
	A	B	C	D	V	W	X
1	Name	Volume	Bg. Value	Bg. Type	SumAboveBG	PixelsAboveBG	
2	GRID- 1-R1C1	1.06E+08	2021.892	ObjAverage	152147333.8	22638	=V2-W2*C2
3	GRID- 1-R2C1	20178191	2021.892	ObjAverage	50679734.74	15076	20197691
4	GRID- 1-R1C2	1.08E+08	2021.892	ObjAverage	152871829.1	22165	108056593
5	GRID- 1-R2C2	11945119	2021.892	ObjAverage	41041177.23	14352	12022983
6	GRID- 1-R1C3	1.51E+08	2021.892	ObjAverage	195241374.6	22091	150575758
22	RECT- 2	43891225	0	None	43891225.45	21708	

This takes the sum of all pixels whose intensity is greater than the background value (V), and then subtracts the corresponding background volume is subtracted for those pixels only (W*C). Pixels with intensities less than the background value do not contribute to the final number. Columns in the volume report that are not used are hidden for clarity.

Document images

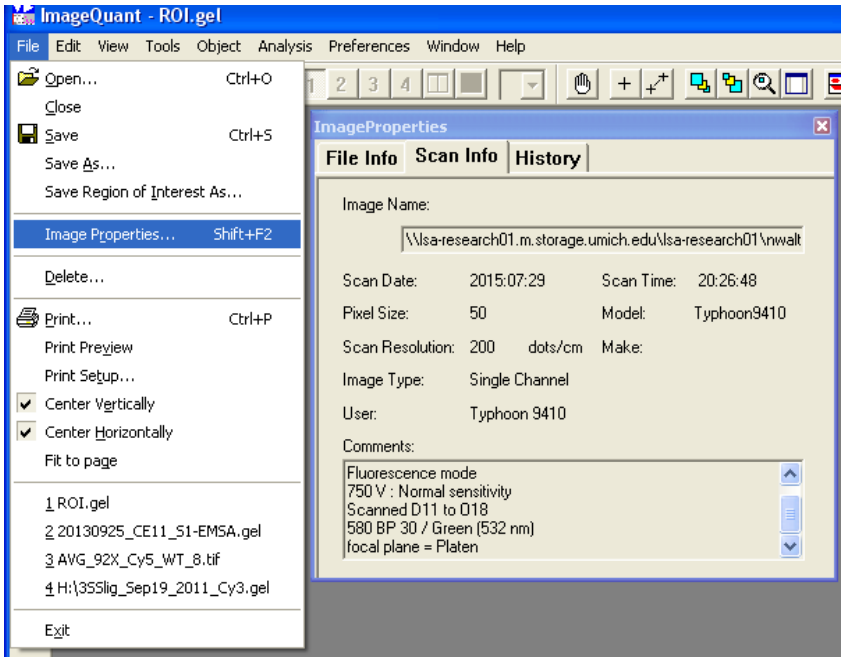
A major advantage of ImageQuant over ImageJ is the boxes drawn in ImageQuant can be saved, reopened, and adjusted later. This is particularly useful if you wish to later use a different rectangle is needed for background correction or box additional bands.

1) Save a copy of the gel image with the current grid and rectangle locations. Use the Region of Interest box to select the parts of the gel you wish to save, then select Edit> Copy Window Bitmap. Paste the image in a convenient place, e.g. the excel sheet containing the quantification.



2) To copy a high resolution image of the gel with the current contrast settings and without the grids and rectangle shown, Edit> Copy Image.

3) Lastly, view the parameters used in scan for later reference. Select File > Image Properties… and switch to the Scan Info tab.



B.8 Matlab code for processing UV-melting curves: `UVmelt_curveFit.m`

Contents

- Description
- Get Input File Names
- Create data structure called expt that stores the temperature range,
- Perform fitting for each sample
- Plot raw and smoothed data
- Find initial guesses for plateaus
- Do linear regression fitting on data and maximize adjusted r-square
- Smooth alpha, theta and plot 1st derivative
- Do 2-gaussian fitting first derivative d(alpha)/d(T^-1)
- Determine which Tm or Tms to save (i.e., peaks)
- Save expt data structure variable for use later

Description

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Processes data from Beckman DU-640B UV-vis melting curve data
%
% REQUIRED:
%    - Matlab Curve fitting Toolbox
%   Additional accessory function scripts in same directory
%    - poly8Fit.m
%    - gauss2Fit.m
%    - cursorBoxSelect
% INPUT:
%   Excel workbook containing worksheets converted from *.DUF
%   Each worksheet was originally one experiemnt on the instrument
%   Locations of sample data must be hard-coded on ln 60.
%   COOLING RAMPS: sort data in excel in order of INCREASING temperature.
% OUTPUT:
% 1) User-directed baseline correction:
%       plot approximate second derivative, zero-crossing points aid
%       user in selecting regions (between X-ing pts) that have
%       constant slope
% 2) Calculate fraction folded (alpha) and fraction unfolded (theta):
%       correct raw A260 w/ user-defined baselines, apply Savitzky-Golay
%       filtering to alpha, theta using user-defined span (window of pts)
% 3) Find Tm transitions in 1st derivative (dalpha/d(1/T)) vs T plot:
%       fit derivative plot with 2 gaussians; user can adjust initial
%       parameters for fitting ig fit is off. Tm are recorded only for
%       user-selected peaks in the derivative plot.
% 4) Save analyzed data as matlab variable that can be loaded again later
%
% Paul Lund 08/2015
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Clear all;
close all;
```

Get Input File Names

```
fprintf(1,'Please select the file you wish to Analyze\n')
[TestFileName, TestPathName] = uigetfile('*.*','Please select all your Path
files','Multiselect','off');
FileName = strcat(TestPathName,TestFileName);
[status, sheets] =xlsfinfo(FileName) %Get names of all sheets in workbook

% Dynamically select data from excel
```

```
% smplNames = xlsread(FileName,-1) %actively select datarange from excel
% num = xlsread(FileName,-1) %actively select datarange from excel

blank = input('Enter cuvette number of blank, comma between each if more than one.
Enter 0 if no blank. [1]','s'); % Check which cuvette;
multiBlanksFlag = 0;
if isempty(blank)
    blanks = 1;

elseif  length(blank)>1
    [mat,tok] = regexp(blank, '\d', 'match', 'tokens'); % mat{:}
    multiBlanksFlag = 1;

else length(blank) ==1
    blanks = str2double(blank);

end

% Hardcode data ranges for data in each sheet
nameRange = 'D30:H30';
tempRange = 'A51:A189';
absRange = 'Q51:T189';
```

Create data structure called expt that stores the temperature range,
sample names, and abs values for each worksheet in the book.

```
expt = [];




for j =1:numel(sheets)
    expt(j).name = sheets{j};
    expt(j).temps = xlsread(FileName, sheets{j},tempRange);
    [num,expt(j).smplNames] = xlsread(FileName, sheets{j}, nameRange);
    [expt(j).data] = xlsread(FileName, sheets{j},absRange);

    % duplicate missing value in first row of last sample
    [r,c] = size(expt(j).data);
    if isnan(expt(j).data(1,c))
        expt(j).data(1,c) = expt(j).data(2,c);
    end

    if multiBlanksFlag >0  % mutliple blanks used when true
        display('WARNING: Code can not handle this condition. Load data manually.')
        break;
    end
end
```

Perform fitting for each sample

```
% Iterate through each sample expt(m).data(:,q) for all expt(m)
% Choose which sample to start from
startPoint = input('Start from begining? (y/n) [y]','s')
if strcmp(startPoint, 'n')
    mstart = input('Sheet number:');
    qstart = input('Sample number:');
else
    mstart = 1;
    qstart = 1;
```

```
end

for m = mstart:numel({expt.name})
    currX = expt(m).temps;   %temps for current experiment
    expt(m).Tm = zeros(2,numel(expt(m).smplNames)-1); %Pre-allocate space for up to 2
Tm for each sample
    % Iterate through all samples in the expt
    for q = qstart:numel(expt(m).smplNames)-blanks
        close all;  % close all figure windows from previous sample
        currY = expt(m).data(:,q); %Corrected abs values for current sample
        currName = expt(m).smplNames(q+blanks);  % assumes that there's no or only 1
blank (cuvette 1)
        disp('Working on '); disp(currName);

        %Simple Moving average with delay correction
        window = input('moving average window?');
        coeffs = ones(1,window)/window;
        fDelay2 = (length(coeffs)-1)/2;
        movAvgY = filter(coeffs,1,currY);

        % Savitzky-Golay filter data w/ linear regression over window
        % ** Use these smoothed data to facilitate choosing baseline correction,
        % then go back and use baseline-corrected raw data and apply S. Golay fitler
**
        sgoY = sgolayfilt(currY,1,window);

        % Fit sgolay filtered data to 8th order polynomial
        [fitParam, gof] = poly8Fit(currX, sgoY);

        % Generate data pts for 1st and 2nd derivative of polynomial
        % poly8Fit uses normalized x values in fitting, so generate normalized X
        % values for plotting xnew = (x-mean(x))/std(x)
        meanT = mean(currX);
        stdT = std(currX);
        i = (currX(1)-meanT)/stdT;
        h = ((currX(length(currX))-meanT)/stdT-i)/500; %set norm'd stepsize to get
~500 pts
        n=1;
        deriv1 = [];
        deriv2 = [];
        deriv0 = [];
        derivx = [];
        while i<(currX(length(currX))-meanT)/stdT
            deriv0(n) = fitParam.p1*i^8 + fitParam.p2*i^7 + fitParam.p3*i^6 +
fitParam.p4*i^5 + fitParam.p5*i^4 + fitParam.p6*i^3 + fitParam.p7*i^2 + fitParam.p8*i
+ fitParam.p9;
            deriv1(n) = 8*fitParam.p1*i^7 + 7*fitParam.p2*i^6 + 6*fitParam.p3*i^5 +
5*fitParam.p4*i^4 + 4*fitParam.p5*i^3 + 3*fitParam.p6*i^2 + 2*fitParam.p7*i +
fitParam.p8;
            deriv2(n) = 56*fitParam.p1*i^6 + 42*fitParam.p2*i^5 + 30*fitParam.p3*i^4 +
20*fitParam.p4*i^3 + 12*fitParam.p5*i^2 + 6*fitParam.p6*i + 2*fitParam.p7;
            derivx(n)= i;
            n = n+1;    %increment loop counter, index
            i = i+h; % increment step size
        end
```

Plot raw and smoothed data

add sgolay filtering with same window as moving average

```
        f1 = figure; hold on;
```

```
        p2 = plot(currX, currY, ['o','k'], currX, sgoY,['r']);
        p2(2).LineWidth = 2;
        legend( p2, 'Raw Data', 'Smoothed Data', 'Location', 'NorthWest' );
        xlabel('Temperature (degrees C)')
        ylabel('Abs_2_6_0 _n_m (a.u.)')
        title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

        % Plot second derivative and zero line to easily visualize crossing
        % points
%         pYMA = plot(currX-fDelay2/window,movAvgY,'b');pYMA.LineWidth = 2;
%         f2 = figure; hold on;
%         plot(derivx*stdT+meanT,deriv2,'r');
%         plot([derivx(1)*stdT+meanT,derivx(length(derivx))*stdT+meanT], [0,0],'b--');
%         xlabel('Temperature (degrees C)');
%         ylabel('d^2A/dT^2');
%         title(strcat(strrep(expt(m).name,'_','\_'),':',currName,':  2nd
derivative'));
```

Find inital guesses for plateaus

Use coordinates where 2nd deriv = 0 a starting point, choose outermost pairs (1,2) and (n, last data point; Perform linear regression and plot, check with user.

```
        % Make overlay figure of raw data and 2nd deriv.
        f3 = figure;
        plot(currX, currY, ['o','k'], currX, sgoY,['r']);

        ax1 = gca;
        xlabel('Temperature (degrees C)');
        ylabel('Abs_2_6_0 _n_m (a.u.)');
        legend('Raw Data', 'Smoothed Data', 'Location', 'NorthWest' );
        set(ax1, 'XColor', 'k', 'YColor', 'k');
        ax1.XGrid = 'on';
        ax1.YGrid = 'off';
        hold(ax1, 'all');  %   <--------------------------------
        ax2 = axes('Position',get(ax1,'Position'),'XAxisLocation','top',...
            'YAxisLocation','right','Color','none','XColor','k','YColor','k');
        hold(ax2, 'all');  %   <-------------------------------

        plot(derivx*stdT+meanT,deriv2,'Parent', ax2);
        plot([derivx(1)*stdT+meanT,derivx(length(derivx))*stdT+meanT], [0,0],'b--');
        ax2.XGrid = 'on';
        ax2.YGrid = 'off';
        set(ax2, 'XColor', 'b', 'YColor', 'b');
        xlabel('Temperature (degrees C)');
        ylabel('d^2A/dT^2');
        legend('2nd deriv of poly8Fit', 'Location', 'SouthEast' );

        % Save graph of 2nd deiv and raw data
        fname = strcat(TestPathName,expt(m).name,'-',currName,'_',num2str(m),'-
',num2str(q),'_raw2ndDeriv');
        print(figure(f3),'-dsvg',fname{1});
        print(figure(f3),'-r150','-djpeg',fname{1});

        % Loop until user is satified with baseline corrections
        thetaCorrectionDone = 0;
        while thetaCorrectionDone ~= 1
            figure(f3);
```

```
            xLower = []; xUpper = []; %initialize/reset baseline variables.
            while numel(xLower)<3 % run until have at least 3 pts for each baseline
                % use cursor to get region for use in lower baseline
                [px1,px2] = cursorBoxSelect('LOWER BASELINE:')
                [val1,ind1] = min(abs(currX -px1)); %ind1 is the index of the nearest
temperature data pt to p1
                [val2,ind2] = min(abs(currX -px2)); %ind2 is the index of the nearest
temperature data pt to p2
                disp([currX(ind1),currX(ind2)])
                xLower = currX(ind1:ind2);  %% Extract datapoints in regions used for
lowerbaselines
                yLower = currY(ind1:ind2);  %% Extract datapoints in regions used for
lowerbaselines
                %             yLower = sgoY(ind1:ind2);  %% Extract SGOALY FILTERED
datapoints in regions used for upperbaselines
            end
            while numel(xUpper)<3  % run until have at least 3 pts for each baseline
                % use cursor to get region for use in upper baseline
                [px3,px4] = cursorBoxSelect('UPPER BASELINE:')
                [val3,ind3] = min(abs(currX -px3)); %ind3 is the index of the nearest
temperature data pt to p3
                [val4,ind4] = min(abs(currX -px4)); %ind4 is the index of the nearest
temperature data pt to p4
                disp([currX(ind3),currX(ind4)])
                xUpper = currX(ind3:ind4);  %% Extract datapoints in regions used for
upperbaselines
                yUpper = currY(ind3:ind4);  %% Extract datapoints in regions used for
upperbaselines
                %             yUpper = sgoY(ind3:ind4); %% Extract SGOALY FILTERED
datapoints in regions used for upperbaselines
            end
```

Do linear regression fitting on data and maximize adjusted r-square

use Robust(Bisquare) and Normal linear fitting, choose method w/ higher adjustedR^2

```
            [lowerFitR, LgofR] = fit( xLower, yLower, 'poly1', 'Robust', 'Bisquare');
            [lowerFit, Lgof] = fit( xLower, yLower, 'poly1');
            if LgofR.adjrsquare >Lgof.adjrsquare
                lowerFit = lowerFitR;
            end

            [upperFitR, UgofR] = fit( xUpper, yUpper, 'poly1', 'Robust', 'Bisquare');
            [upperFit, Ugof] = fit( xUpper, yUpper, 'poly1');
            if UgofR.adjrsquare > Ugof.adjrsquare
                upperFit = upperFitR;
            end

            % Plot baselines on top of raw data figure
            x = [currX(1), currX(length(currX))]
            yLower = lowerFit.p1*x+lowerFit.p2;
            yUpper = upperFit.p1*x+ upperFit.p2;

            figure (f1); hold off;
            p2 = plot(currX, currY, ['o','k'], currX, sgoY,['r']);
            p2(2).LineWidth = 2;
            legend( p2, 'Raw Data', 'Smoothed Data', 'Location', 'NorthWest' );
            xlabel('Temperature (degrees C)')
            ylabel('Abs_2_6_0 _n_m (a.u.)')
            title(strcat(strrep(expt(m).name,'_','\_'),':',currName));
            hold on;
```

```
            plot(x, yLower);  % Plot baselines
            plot(x, yUpper);

            % Generate baseline-corretion on the RAW data

            for t =1:length(currX)
                %theta = (A-Alower)/(Aupper-Alower) (Owczarzy)
                thetaTemp(t) = (currY(t)-(lowerFit.p1*t+lowerFit.p2))/((upperFit.p1*t+
upperFit.p2)-(lowerFit.p1*t+lowerFit.p2));

                %theta = (Aupper-A)/(Aupper-Alower) (Mergny; what Marky+Breslauer call
alpha, Mergny calls theta)
                alphaTemp(t) = ((upperFit.p1*t+ upperFit.p2)-
currY(t))/((upperFit.p1*t+ upperFit.p2)-(lowerFit.p1*t+lowerFit.p2));

            end
            % Make initial plot of alpha to help user evaluate baseline choice
            f5 = figure; hold on;
            alphaPlotTemp = plot(currX, alphaTemp);

            alphaPlotTemp(1).LineWidth = 2;
            legend(alphaPlotTemp, 'No smoothing','Location', 'SouthWest' );
            xlabel('Temperature (degrees C)')
            ylabel('\alpha')
            title(strcat(strrep(expt(m).name,'_','\_'),':',currName));


            % Check with user if correction is good
            ansTheta = input('Satisfied with baseline correction? y/n [n]','s');
            if strcmp(ansTheta,'y')
                thetaCorrectionDone = 1;
                % Save data for baselines to expt data structure
                expt(m).baseLow(1,q) = lowerFit.p1; % slope
                expt(m).baseLow(2,q) = lowerFit.p2; % intercept
                expt(m).baseUp(1,q) = upperFit.p1;
                expt(m).baseUp(2,q) = upperFit.p2;

                % Save basline corrected alpha, theta (no smoothing)
                expt(m).alpha(:,q) = alphaTemp;
                expt(m).theta(:,q) = thetaTemp;

                % Save graph of raw data with baselines
                fname1 = strcat(TestPathName,expt(m).name,'-',currName, '_',
num2str(m),'-',num2str(q),'_baselines');
                print(figure(f1),'-dsvg',fname1{1});
                print(figure(f1),'-r150','-djpeg',fname1{1});

                figure(f5);
                close(gcf);
            else
                thetaCorrectionDone = 0;
                figure(f5);
                close(gcf);
            end
        end
```

Smooth alpha, theta and plot 1st derivative

```
        fracFoldedOK = 0;


        window2 = window;
```

```
        while fracFoldedOK ~= 1
            % Apply SGOLAY FILTERING on alpha, Theta
            sgoTheta = sgolayfilt(thetaTemp,1,window2);
            sgoAlpha = sgolayfilt(alphaTemp,1,window2);

            f5 = figure; hold on;
            alphaPlot = plot(currX, sgoAlpha, ['r'],currX, alphaTemp);

            alphaPlot(1).LineWidth = 2;
            legend(alphaPlot, strcat('Smoothed \alpha, span= ', num2str(window2)), 'No
smoothing','Location', 'SouthWest' );
            xlabel('Temperature (degrees C)');
            ylabel('\alpha');
            title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

            % Find and plot 1st derivative of alpha
            f6 = figure; hold on;
            xlabel('Temperature (degrees C)');
            ylabel(['d','\alpha','/d','(1/T)']);
            dAlpha = diff(sgoAlpha)';
            dTminus1 = diff((1./currX));
            dAlpha_dTminus1 = dAlpha./dTminus1;
            % Remove last value from currX b/c taking numberical deriv by differences
            currXSmall = currX(1:(length(currX)-1));
            plot(currXSmall,dAlpha_dTminus1);
            legend(strcat('Smoothed \alpha, span= ',
num2str(window2)),'Location', 'West' );
            title(strcat(strrep(expt(m).name,'_','\_'),':',currName,': 1st
derivative'));
            ansTheta2 = input('Satisfied with smoothing? y/n [n]','s');
            if strcmp(ansTheta2,'y')
                fracFoldedOK = 1;

                % Save sgo filtered alpha, theta
                expt(m).sgoAlpha(:,q) = sgoAlpha;
                expt(m).sgoTheta(:,q) = sgoTheta;
                expt(m).sgo_span(q) = window2; % window for smoothing

                % Save first derivative
                expt(m).dAdTminus1(:,q) = dAlpha_dTminus1;

                % Save graph of smoothed alpha
                fname2 = strcat(TestPathName,expt(m).name,'-
',currName, '_',num2str(m),'-',num2str(q),'_smoothAlpha');
                print(figure(f5),'-dsvg',fname2{1});
                print(figure(f5),'-r150','-djpeg',fname2{1});

            else
                fracFoldedOK =0;
                window2 = input('Window for S.-Golay smoothing of \alpha? (odd #)');
                if mod(window2,2)~=1; window2 = window2+1;end %Make odd if even
                figure(f5);
                close gcf;
                figure(f6);
                close gcf;
            end
        end
```

Do 2-gaussian fitting first derivative d(alpha)/d(T^-1)

a1*exp(-((x-b1)/c1)^2) + a2*exp(-((x-b2)/c2)^2)

```
        % intialize parameters
        % peak 1
        a1 = max(dAlpha_dTminus1); % peak height
        b1 = currXSmall(round(2/3*length(currXSmall))); % peak center
        c1 =  (currXSmall(length(currXSmall))- currXSmall(1))/20; % peak width, 5% of
X-range
        % peak 2
        a2 = max(dAlpha_dTminus1)*0.5;
        b2 = currXSmall(round(3/4*length(currXSmall)));
        c2 = c1;
        gaussFitDone = 0;
        while gaussFitDone ~=1
            [derivFit, derivgof] =
gauss2Fit(currXSmall,dAlpha_dTminus1,a1,b1,c1,a2,b2,c2);
            figure(f6); hold off;
            xlabel('Temperature (degrees C)');
            ylabel(['d','\alpha','/d','(1/T)']);
            plot(currXSmall,dAlpha_dTminus1,'+b');
            hold on;
            % Plot gaussians and sum of gaussians
%             peakSum = plot( derivFit, currXSmall, dAlpha_dTminus1);

            peak1 = derivFit.a1*exp(-((currXSmall-derivFit.b1)/derivFit.c1).^2);
            peak2 = derivFit.a2*exp(-((currXSmall-derivFit.b2)/derivFit.c2).^2);
            peakSum = peak1 + peak2;

            plot(currXSmall,peakSum,'k--','LineWidth',2);
            plot(currXSmall,peak1,'g');
            plot(currXSmall,peak2,'m');
            xlabel('Temperature (degrees C)');
            ylabel(['d','\alpha','/d','(1/T)']);

            legend(strcat('Smoothed \alpha, span= ', num2str(window2)),'Sum','Peak
1', 'Peak 2','Location', 'West' );
            title(strcat(strrep(expt(m).name,'_','\_'),':',currName,': 1st
derivative'));
            grid on;
            ansGauss = input('Satified with guass fitting (y/n)? [y]','s');
            if strcmp(ansGauss,'n');
                b1 = input('Enter guess for peak 1 center: ');
                b2 = input('Enter guess for peak 2 center: ');
                c1 = input('Enter guess for peak 1 width: ');
                c1 = input('Enter guess for peak 2 width: ');
                gaussFitDone = 0;
            else
                gaussFitDone = 1;
                % Save graph of smoothed dalpha/d(1/T)with gauss fits to find Tm
                fname3 = strcat(TestPathName,expt(m).name,'-',currName, '_',
num2str(m),'-',num2str(q),'_dAdT-1_gauss');
                print(figure(f6),'-dsvg',fname3{1});
                print(figure(f6),'-r150','-djpeg',fname3{1});
            end
        end
```

Determine which Tm or Tms to save (i.e., peaks)

```
        TmSelectionDone =0;


        while TmSelectionDone ~= 1
           % Prompt user for which Tms to save
            [px5,px6] = cursorBoxSelect('MELTING TEMPS(peak centers):')
```

```
            %Save Tm (i.e. peak centers) for selectd peaks in 1st deriv

            if derivFit.b1>= px5 && derivFit.b2>= px5 && derivFit.b1<= px6 &&
derivFit.b2 <= px6 % Both peaks selected
                expt(m).Tm(1,q) = derivFit.b1;
                expt(m).Tm(2,q) = derivFit.b2;
            elseif derivFit.b1 >= px5 && derivFit.b1 <= px6 % Peak 1 only
                expt(m).Tm(1,q) = derivFit.b1;
                expt(m).Tm(2,q) = 0;
            elseif derivFit.b2 >= px5 && derivFit.b2 <= px6 % Peak 2 only
                expt(m).Tm(1,q) = derivFit.b2;
                expt(m).Tm(2,q) = 0;
            end
            disp(strcat('Tm1 =',num2str(expt(m).Tm(1,q)),'    Tm2
=',num2str(expt(m).Tm(2,q))));
            ansTm = input('Satified with selected Tms (y/n)? [n]','s');

            if strcmp(ansTm,'y')
                TmSelectionDone = 1;
            else
                TmSelectionDone = 0; expt(m).Tm(1,q) = 0; expt(m).Tm(2,q)=0;
            end
        end
    end
     if q == numel(expt(m).smplNames)-1; qstart = 1; end % Reset q to 1 if done with
last sample in current expt
end
```

Save expt data structure variable for use later

```
filePrefix = input('Name for save file: ','s');
fname4 = strcat(TestPathName, filePrefix, '_expt.mat');
save(fname4,'expt');
% Write results file to excel
% xlswrite('myfile.xlsx',expt(1).sgoAlpha,'Sheet2','B2')
```

Published with MATLAB® R2015a

B.9 Accessory functions and scripts for use with `UVmelt_curveFit.m`

List of functions:

- Poly8Fit.m
- Gauss2Fit.m
- cursorBoxSelect.m

List of Scripts:

- combineExptVars.m
- exptVar_2_xlsx.m
- normThetaAlpha.m

Poly8Fit.m

```
function [fitresult, gof] = poly8Fit(temps, sgoY)
%CREATEFIT(TEMPS,SGOY)
%  Create a fit.
```

```
%
%  Data for 'untitled fit 1' fit:
%      X Input : temps
%      Y Output: sgoY
%  Output:
%      fitresult : a fit object representing the fit.
%      gof : structure with goodness-of fit info.
%
%  See also FIT, CFIT, SFIT.
%  Auto-generated by MATLAB on 12-Aug-2015 15:59:40
%  Modified by Paul Lund 08/2015

 [xData, yData] = prepareCurveData( temps, sgoY );

% Set up fittype and options.
ft = fittype( 'poly8' );

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, 'Normalize', 'on' );

% Plot fit with data.
figure( 'Name', 'untitled fit 1' );
h = plot( fitresult, xData, yData);
legend( h, 'Smoothed Data', 'Polynomial Fit (n=8)', 'Location', 'NorthWest' );
% Label axes
xlabel('Temperature (degrees C)')
ylabel('Abs_2_6_0 _n_m (a.u.)')
grid on
```

Published with MATLAB® R2015a

cursorBoxSelect.m

Returns x-axis coordinates for box drawn with cursor

Modified from Haon_off.m Paul Lund 08/2015

```
function [point1,point2] = cursorBoxSelect(regionName)
fprintf(1,'Select a region of the graph to use as ')
disp(regionName)
k = waitforbuttonpress;                % Hold program until user selects region
point1 = get(gca,'CurrentPoint');      % button down detected
finalRect = rbbox;                     % return figure units
point2 = get(gca,'CurrentPoint');      % button up detected
point1 = point1(1,1);                  % extract x only
point2 = point2(1,1);
if point2 <point1                        % swap upper and lower value if box drawn
backwards
    temp = point1;
    point1 = point2;
    point2 = temp;
end
% p1 = min(point1,point2);               % calculate locations
% offset = abs(point1-point2);           % and dimensions
% f = nix;
% for i = [point1(1), point2(1)]
%     [val,ind] = min(abs(f-i));
    % ind is the index of the nearest value
```

```
% end
```

Published with MATLAB® R2015a

gauss2Fit.m

```
function [fitresult, gof] = gauss2Fit(currXSmall, dAlpha_dTminus1,a1,b1,c1,a2,b2,c2)
%CREATEFIT(CURRXSMALL,DALPHA_DTMINUS1)
%  Create a fit.
%
%  Data for 'untitled fit 1' fit:
%      X Input : currXSmall
%      Y Output: dAlpha_dTminus1
%      Starting paramters: a1, b1, c1, a2, b2, c2
%  Output:
%      fitresult : a fit object representing the fit.
%      gof : structure with goodness-of fit info.
%
%  See also FIT, CFIT, SFIT.
%  Auto-generated by MATLAB on 17-Aug-2015 16:38:42
%  Modified by Paul Lund 08/2015

 [xData, yData] = prepareCurveData( currXSmall, dAlpha_dTminus1 );


% Set up fittype and options.
ft = fittype( 'gauss2' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.Lower = [0 -Inf 0 -Inf -Inf 0];
opts.Robust = 'Bisquare';

opts.StartPoint = [a1 b1 c1 a2 b2 c2];

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );
```

Published with MATLAB® R2015a

combineExptVars.m

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Use to transfer analyzed data from one expt data structure to another
% useful when you need to re-analyze a single sample and then insert new
% analysis into existing expt data structure containing other analyzed data
% Paul Lund 08/2015
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Import data structures

```
clear all; close all;
fprintf(1,'Please select the file you wish to extract from\n')
[TestFileName, TestPathName] = uigetfile('*.mat','Please select the expt variable you
wish to extract from','Multiselect','off');
sourceFileName = strcat(TestPathName,TestFileName)
    source = load(sourceFileName,'expt');
```

```
fprintf(1,'Please select the file you wish to add to\n')
[TestFileName, TestPathName] = uigetfile('*.mat','Please select the expt variable you
wish to add to','Multiselect','off');
destFileName = strcat(TestPathName,TestFileName)
destination = load(destFileName,'expt');
```

Copy data from source to destination

EXCLUDING sheet names, sample names, temperatures and raw data

```
exptNums =input('Input sheet numbers in [# # #] format: ')
smplNums = input('Input sample numbers in [# #] format: ')
for m= exptNums % experiment numbers (xlsx sheet number)
    for q = smplNums  % sample numbers (1st sample = 1)
        destination.expt(m).Tm(:,q) = source.expt(m).Tm(:,q);
        destination.expt(m).baseLow(:,q) = source.expt(m).baseLow(:,q);
        destination.expt(m).baseUp(:,q) = source.expt(m).baseUp(:,q);
        destination.expt(m).alpha(:,q) = source.expt(m).alpha(:,q);
        destination.expt(m).theta(:,q) = source.expt(m).theta(:,q);
        destination.expt(m).sgoAlpha(:,q) = source.expt(m).sgoAlpha(:,q);
        destination.expt(m).sgoTheta(:,q) = source.expt(m).sgoTheta(:,q);
        destination.expt(m).sgo_span(q) = source.expt(m).sgo_span(q);
        destination.expt(m).dAdTminus1(:,q) = source.expt(m).dAdTminus1(:,q);
    end
end
```

Create new expt data structure and copy all fields

```
expt = [];

expt = destination.('expt');
```

Save updated expt data structure for use later

```
filePrefix = input('Name for save file: ','s');
fname4 = strcat(TestPathName, filePrefix, '_expt.mat');
save(fname4,'expt');
```

Published with MATLAB® R2015a

normThetaAlpha.m

Description

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% use to generate pretty 0-1 normalized plots of smoothed theta and alpha
% using previously analyzed expt data
% Paul Lund 08/2015
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Import data structures

```
clear all; close all;
fprintf(1,'Please select the file you wish to work with\n')
[TestFileName, TestPathName] = uigetfile('*.mat','Please select the expt variable you
wish to work on','Multiselect','off');
sourceFileName = strcat(TestPathName,TestFileName)
    source = load(sourceFileName,'expt');
```

```
%Create new expt data structure and copy all fields
expt = [];
expt = source.('expt');

% Choose which sample to start with
% startPoint = input('Start from begining? (y/n) [y]','s');
disp('Choose sample to nromalize: ');
% if strcmp(startPoint, 'n')
    mstart = input('Sheet number:');
    qstart = input('Sample number:');
% else
%    mstart = 1;
%    qstart = 1;
% end

m = mstart;
q = qstart;

blank = input('Enter cuvette number of blank, comma between each if more than one.
Enter 0 if no blank. [1]','s'); % Check which cuvette;
multiBlanksFlag = 0;
if isempty(blank)
    blanks = 1;

elseif  length(blank)>1
    [mat,tok] = regexp(blank, '\d', 'match', 'tokens'); % mat{:}
    multiBlanksFlag = 1;

else length(blank) ==1
    blanks = str2double(blank);
end
```

Load data

```
currX = expt(m).temps;   %temps for current experiment
currAlpha = expt(m).sgoAlpha(:,q);
currTheta = expt(m).sgoTheta(:,q);
currName = expt(m).smplNames(q+blanks);  % assumes that there's no or only 1 blank
(cuvette 1)
```

Plot SMOOTHED data

```
thetaCorrectionDone = 0;
while thetaCorrectionDone ~= 1
    f1 = figure; hold on;
    p1 = plot(currX, currAlpha, ['+','k'], currX, currAlpha, 'r');
    p1(2).LineWidth = 2;
    legend(strcat('Smoothed \alpha, span= ',
num2str(expt(m).sgo_span(q))), 'Location', 'SouthWest' );
    xlabel('Temperature (degrees C)')
    ylabel('\alpha')
    title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

    f2 = figure; hold on;
    p2 = plot(currX, currTheta, ['+','k'], currX, currTheta, 'b');
    p2(2).LineWidth = 2;
    legend(strcat('Smoothed \theta, span= ',
num2str(expt(m).sgo_span(q))), 'Location', 'NorthWest' );
    xlabel('Temperature (degrees C)')
    ylabel('\theta')
    title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

    % Loop until user is satified with normalization
```

```
    figure(f2);
    xZero = []; xOne = []; %initialize/reset baseline variables.
    while numel(xZero)==0 % run until have at least 1 pts for each baseline
        % use cursor to get region for use in lower baseline
        [px1,px2] = cursorBoxSelect('MINIMUM:')
        [val1,ind1] = min(abs(currX -px1)); %ind1 is the index of the nearest
temperature data pt to p1
        [val2,ind2] = min(abs(currX -px2)); %ind2 is the index of the nearest
temperature data pt to p2
        disp([currX(ind1),currX(ind2)])
        xZero = currX(ind1:ind2);  %% Extract datapoints in regions used for fully
folded
        yZeroTheta = currTheta(ind1:ind2);
        yOneAlpha = currAlpha(ind1:ind2);
    end
    while numel(xOne)==0 % run until have at least 1 pts for each baseline
        % use cursor to get region for use in upper baseline
        [px3,px4] = cursorBoxSelect('MAXIMUM:')
        [val3,ind3] = min(abs(currX -px3)); %ind3 is the index of the nearest
temperature data pt to p3
        [val4,ind4] = min(abs(currX -px4)); %ind4 is the index of the nearest
temperature data pt to p4
        disp([currX(ind3),currX(ind4)])
        xOne = currX(ind3:ind4);  %% Extract datapoints in regions used for fully
UNfolded
        yOneTheta = currTheta(ind3:ind4);
        yZeroAlpha = currAlpha(ind3:ind4);
    end
    %Clean up figures
    figure (f1); close gcf;
    figure (f2); close gcf;
```

Calculate and plot normalized data

```
    normCurrTheta = (currTheta - mean(yZeroTheta))./(mean(yOneTheta)-
mean(yZeroTheta));
    normCurrAlpha  = (currAlpha - mean(yZeroAlpha))./(mean(yOneAlpha)-
mean(yZeroAlpha));

    % Re-plot normalized data
    f1 = figure; hold on;
    p1 = plot(currX, normCurrAlpha, ['+','k'], currX, normCurrAlpha, 'r');
    p1(2).LineWidth = 2;
    legend(strcat('Smoothed \alpha, span= ',
num2str(expt(m).sgo_span(q))), 'Location', 'SouthWest' );
    xlabel('Temperature (degrees C)')
    ylabel('\alpha')
    % Plot 0 and 1 lines to aid user in evaluating normalization
    plot([currX(1),currX(length(currX))],[0,0],'--','Color',[.8 .8 .8]);
    plot([currX(1),currX(length(currX))],[1,1],'--','Color',[.8 .8 .8]);
    title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

    f2 = figure; hold on;
    p2 = plot(currX, normCurrTheta, ['+','k'], currX, normCurrTheta, 'b');
    p2(2).LineWidth = 2;
    legend(strcat('Smoothed \theta, span= ',
num2str(expt(m).sgo_span(q))), 'Location', 'NorthWest' );
    xlabel('Temperature (degrees C)')
    ylabel('\theta')
    % Plot 0 and 1 lines to aid user in evaluating normalization
    plot([currX(1),currX(length(currX))],[0,0],'--','Color',[.8 .8 .8]);
```

```
    plot([currX(1),currX(length(currX))],[1,1],'--','Color',[.8 .8 .8]);
    title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

    % Check with user if correction is good
    ansTheta = input('Satisfied with normalization? y/n [n]','s');
    if strcmp(ansTheta,'y')
        thetaCorrectionDone = 1;
    else
        thetaCorrectionDone = 0;
        close all;
    end
end
```

Select region of graph to save

set axes limits

```
figure(f1); ax1= gca;
axis([ax1.XLim(1),ax1.XLim(2),-0.05,1.1]);
figure (f2); ax2 = gca;
axis([ax1.XLim(1),ax1.XLim(2),-0.05,1.1]);
realRegionDone = 0;
while realRegionDone ~= 1
        xRange = []; %initialize/reset region selection
    while numel(xRange)==0 % run until have at least 1 pts for each baseline
        % use cursor to get region for saving
        [px1,px2] = cursorBoxSelect('GOOD DATA:')
        [val1,ind1] = min(abs(currX -px1)); %ind1 is the index of the nearest
temperature data pt to p1
        [val2,ind2] = min(abs(currX -px2)); %ind2 is the index of the nearest
temperature data pt to p2
        disp([currX(ind1),currX(ind2)])
        xRange = currX(ind1:ind2);  %% Extract datapoints in regions used for fully
folded
        yRangeTheta = normCurrTheta(ind1:ind2);
        yRangeAlpha = normCurrAlpha(ind1:ind2);
    end
    % Re-plot pretty figure w/ grayed out artifcats
    % alpha
    f3 = figure; hold on;
    plot(currX, normCurrAlpha,'.','Color',[.9 .9 .9]);
    ax3 = gca;
    axis([ax3.XLim(1),ax3.XLim(2),-0.05,1.05]);
    % Plot 0 and 1 lines to aid user in evaluating normalization
    plot([ax3.XLim(1),ax3.XLim(2)],[0,0],'--','Color',[.8 .8 .8]);
    plot([ax3.XLim(1),ax3.XLim(2)],[1,1],'--','Color',[.8 .8 .8]);

    p3 = plot(xRange, yRangeAlpha, ['o','r'],'MarkerFaceColor','r');
    legend(p3,strcat('Smoothed \alpha, span= ',
num2str(expt(m).sgo span(q))), 'Location', 'West' );
    xlabel('Temperature (degrees C)')
    ylabel('\alpha')

    title(strcat(strrep(expt(m).name,'_','\_'),':',currName));

        % theta
    f4 = figure; hold on;
    plot(currX, normCurrTheta,'.','Color',[.9 .9 .9]);
    ax4 = gca;
    axis([ax4.XLim(1),ax4.XLim(2),-0.05,1.05]);
    % Plot 0 and 1 lines to aid user in evaluating normalization
    plot([ax4.XLim(1),ax4.XLim(2)],[0,0],'--','Color',[.8 .8 .8]);
    plot([ax4.XLim(1),ax4.XLim(2)],[1,1],'--','Color',[.8 .8 .8]);
```

```
    p4 = plot(xRange, yRangeTheta, ['o','b'],'MarkerFaceColor','b');
    legend(p4,strcat('Smoothed \theta, span= ',
num2str(expt(m).sgo span(q))), 'Location', 'West' );
    xlabel('Temperature (degrees C)')
    ylabel('\theta')

    title(strcat(strrep(expt(m).name,' ','\ '),':',currName));

     % Check with user if correction is good
    ansGood = input('Satisfied with plot? y/n [n]','s');
    if strcmp(ansGood,'y')
        realRegionDone = 1;

    else
        realRegionDone = 0;
        figure(f3); close gcf;
        figure(f4); close gcf;
    end
end
```

Save plots and normalized data

```
%save alpha plot
fname3 = strcat(TestPathName,expt(m).name,'-',currName, '_',num2str(m),'-
',num2str(q),'_normAlpha');
print(figure(f3),'-dsvg',fname3{1});
print(figure(f3),'-r150','-djpeg',fname3{1});
%save theta plot
fname4 = strcat(TestPathName,expt(m).name,'-',currName, '_',num2str(m),'-
',num2str(q),'_normTheta');
print(figure(f4),'-dsvg',fname4{1});
print(figure(f4),'-r150','-djpeg',fname4{1});

% Prepare data
header =vertcat([{''} {''} {''} {'Norm Min'} {num2str(mean(yZeroAlpha))}
{num2str(mean(yZeroTheta))}],...
               [{''} {''} {''} {'Norm Max'} {num2str(mean(yOneAlpha))}
{num2str(mean(yOneTheta))}],...
               [{'Temp'} {'alpha'} {'theta'} {'Temp'} {'Norm alpha'} {'Norm theta'}]);

output5 = [currX, currAlpha, currTheta];
output6 = [xRange yRangeAlpha yRangeTheta];

% Write xlsx
fname5 = strcat(TestPathName,expt(m).name,'-',currName, '_',num2str(m),'-
',num2str(q),'_normData.xlsx');
xlswrite(fname5{1},header,'Sheet1','A1');
xlswrite(fname5{1},output5,'Sheet1','A4');
xlswrite(fname5{1},output6,'Sheet1','D4');
```

Published with MATLAB® R2015a

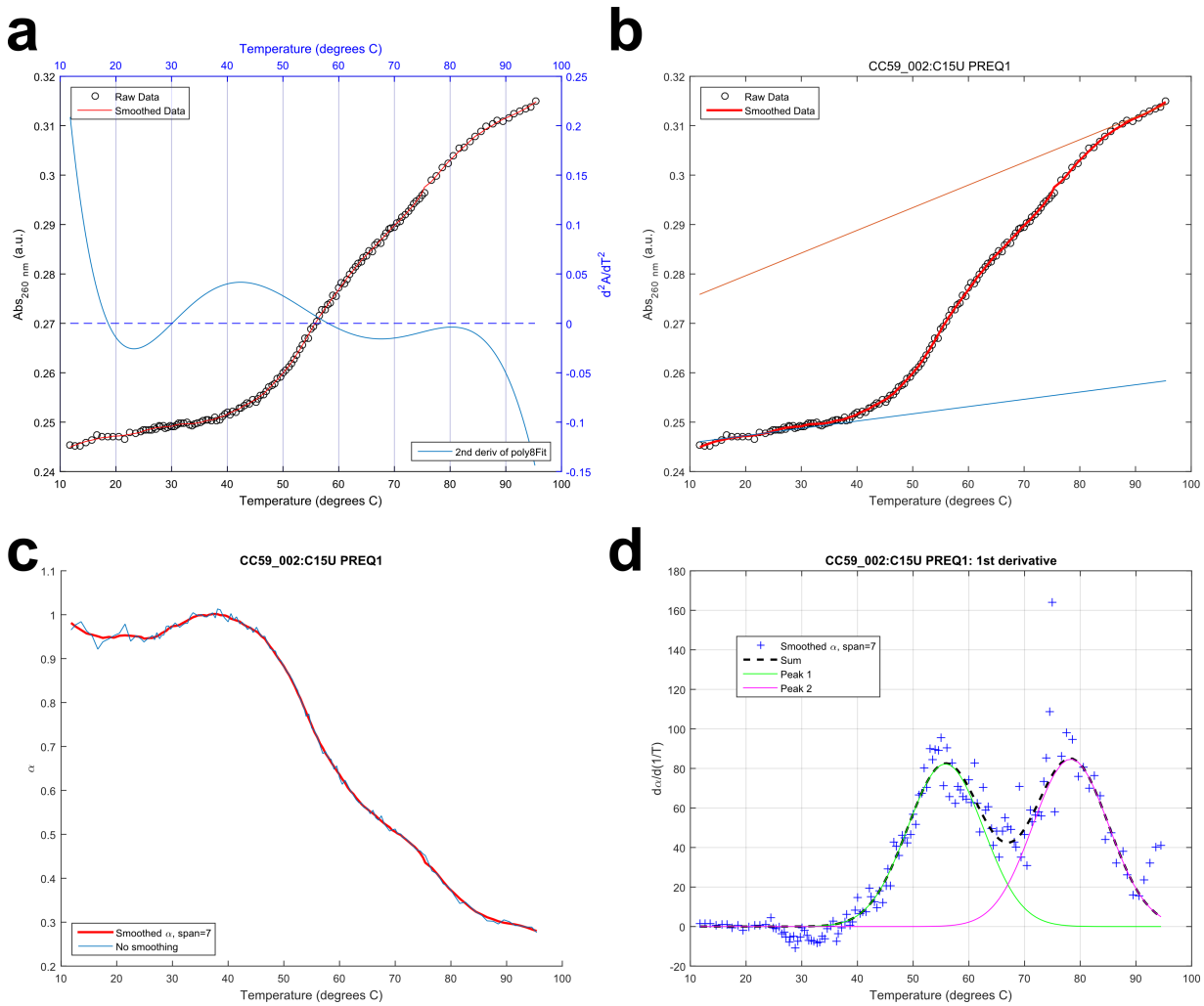
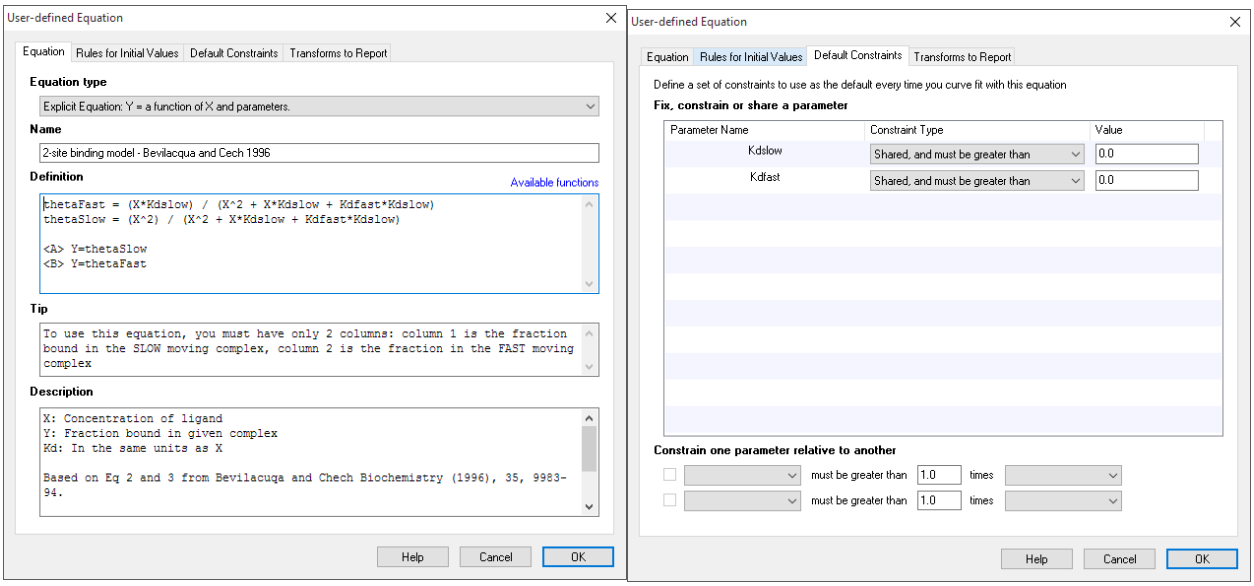


Figure B.9-1 Example output from `UVmelt_curveFit.m`

Example plots generated during the analysis of melting curve data for a sample display two distinct melting transitions. (**a**) Plot of raw absorbance data versus temperature overlaid with plot of approximated second derivative (d^2A/dT^2 *vs* T). The second derivative is approximated by taking the second derivative of the 8th order polynomial fit to Savitzky-Golay filtered raw absorbance values (filter order = 1, window = 7). Zero crossing points aid the user in determining regions of constant slope over which to choose the baselines. (**b**) Plot of user-selected baselines overlaid on top of raw absorbance data. (**c**) Plot of baseline-corrected, smoothed fraction-unfolded (α) data *vs* T. (**d**) Plot of $d\alpha/d(1/T)$ *vs* T fitted with the sum of Gaussians to determine T_m for each transition.

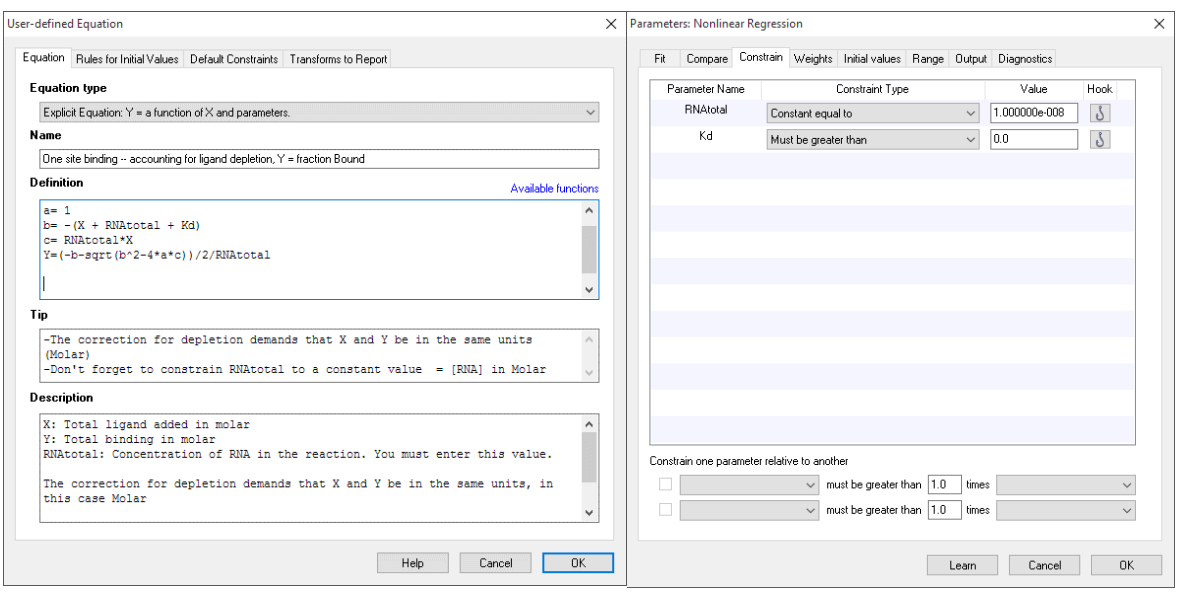
B.10 Implementation of 2-site binding model equations in GraphPad Prism

The following screen shots show the implementation used for simultaneous fitting of the fraction of total RNA bound (θ) in each of 2 distinct complexes (fast and slow) using GraphPad prism. Initial values for Kdfast and Kdslow are set to 0.1*(Value of X and YMID) and 5*(Value of X at YMID), respectively.



B.11 Implementation of 1-site binding model equation accounting for free ligand depletion in GraphPad Prism

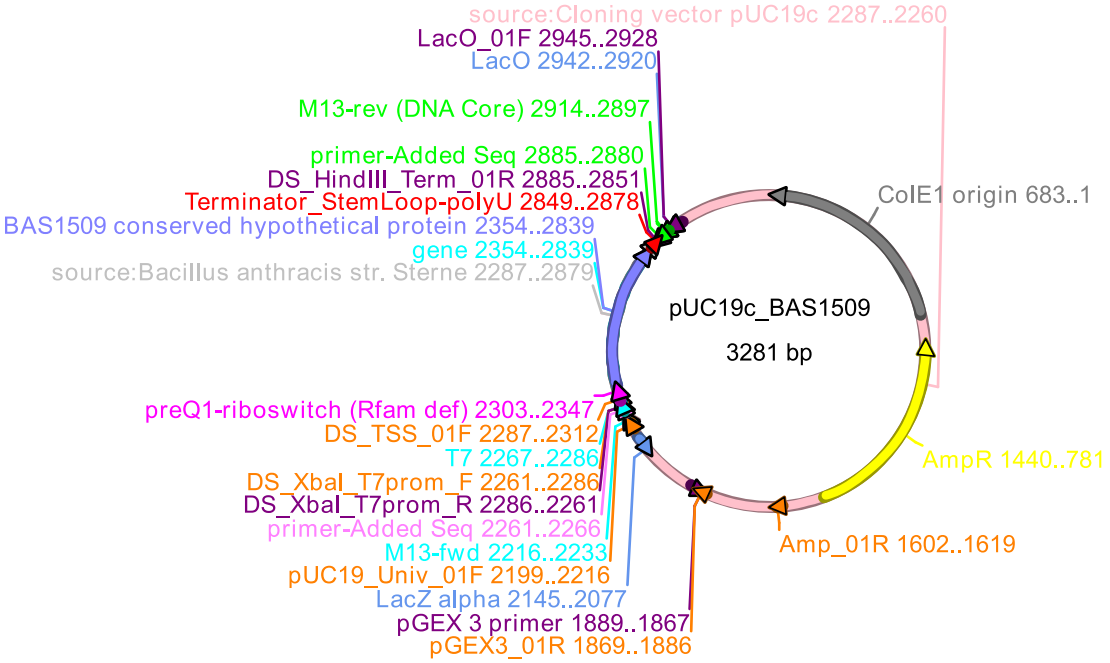
The concentration of RNA in the reaction in molar needs to be entered on the contraints tab when fitting.



Appendix C: Supplementary material for studies of initiation for *Tte* mRNA and the related preQ$_1$ riboswitch from *B. anthracis*

C.1 Sequence information for the pUC19c_BAS1509 plasmid for in vitro *transcription of* Bas *mRNA*

Figure C.1-1 pUC19c_BAS1509 plasmid map




```
LOCUS           pUC19c_BAS1509        3281 bp ds-DNA   circular     17-MAY-2013
DEFINITION  Cloning vector pUC19c, complete sequence.
ACCESSION   L09137 X02514
VERSION     L09137.2  GI:20141090
KEYWORDS    .
SOURCE      Cloning vector pUC19c
  ORGANISM  Cloning vector pUC19c other sequences; artificial sequences;
            vectors.
REFERENCE   1  (bases 1 to 2686)
  AUTHORS   Yanisch-Perron,C., Vieira,J. and Messing,J.
  TITLE     Improved M13 phage cloning vectors and host strains: nucleotide
            sequences of the M13mp18 and pUC19 vectors
  JOURNAL   Gene 33 (1), 103-119 (1985)
   PUBMED   2985470
REFERENCE   2  (bases 1 to 2686)
  AUTHORS   Chambers,S.P., Prior,S.E., Barstow,D.A. and Minton,N.P.
  TITLE     The pMTL nic- cloning vectors. I. Improved pUC polylinker regions
            to facilitate the use of sonicated DNA for nucleotide sequencing
  JOURNAL   Gene 68 (1), 139-149 (1988)
   PUBMED   2851488
REFERENCE   3  (bases 1 to 2686)
```

```
  AUTHORS   Gilbert,W.
  TITLE     Obtained from VecBase 3.0
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1993) Department of Biochemistry, University of
            Minnesota, St. Paul, MN 55108, USA
REFERENCE   5  (bases 1 to 2686)
  AUTHORS   Messing,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2002) Rutgers, The State University of New
            Jersey, Waksman Institute of Microbiology, 190 Frelinghuysen Road,
            Piscataway, NJ 08854-8020, USA
  REMARK    Sequence update by submitter
COMMENT     On Apr 11, 2002 this sequence version replaced gi:209213.
            These data and their annotation were supplied to GenBank by Will
            Gilbert under the auspices of the GenBank Currator Program. pUC19c
            - Cloning vector (beta-galactosidase mRNA on complementary strand)
            ENTRY PUC19C                    #TYPE DNA CIRCULAR TITLE pUC19c -
            Cloning vector
                          (beta-galactosidase mRNA on complementary strand)
            DATE      03-FEB-1986
               #sequence  16-DEC-1986
            ACCESSION VB0033
            SOURCE    artificial
            COLLECTION  ATCC 37254
            REFERENCE
               #number
               #authors  Norrander J., Kempe T., Messing J.
               #journal Gene (1983) 26: 101-106
            REFERENCE
               #number  1
               #authors  Yanisch-Perron C., Vieira J., Messing J.
               #journal Gene (1985) 33: 103-119
               #comment shows the complete compiled sequence
            REFERENCE
               #number  2
               #authors  Chambers,S.P., et al.
               #journal Gene (1988) 68: 139-149
               #describes mutation at nt1308 and its effect on copy number
            REFERENCE
               #number
               #authors Pouwels P.H., Enger-Valk B.E., Brammar W.J.
               #book    Cloning Vectors, Elsvier 1985 and supplements
               #comment  vector I-A-iv-20
            COMMENT
               This Sequence was obtained 3-MAR-1986 from J. Messing, Waksman
               Institute, NJ on floppy disk.
               Revised 16-DEC-1986 by F. Pfeiffer:
               1062/3 'AT' to 'TA' to match revised sequence of PBR322
               The beta-galactosidase mRNA sequence including the multiple
               cloning site of M13mp19 is on the strand complementary to that
               shown.
            KEYWORDS
            CROSSREFERENCE
               #complement
                 VecBase(3):pUC19
               #prerevised
                 GenBank(50):M11662, EMBL(11):ARPuc19
               #parent
                 VecBase(3):pUC13, VecBase(3):M13mp19, VecSource(3):bGal19
            PARENT
```

```
                Features of pUC19c (2686 bp)
                  residue     source
                   1- 137  2074-2210 pBR322
                 138- 237  2252-2351 pBR322
                 238- 395  1461-1304 (c) Lac-Operon
                 396- 452    57-   1 (c) polylinker of M13mp19
                 455- 682  1298-1071 (c) Lac-Operon
                 683-2686  2352-4355 pBR322
                Conflict (cfl) and Mutations (mut):
                     pUC19c  source
                mut 1308  A  G  2977   pBR322   linked to increased copy number
                mut 1942  A  G  3611   pBR322
                mut 2243  T  C  3912   pBR322
             FEATURE
                1629-2417 789-1 (c) Ap-R; b-lactamase
             POLYLINKER  HindIII-SphI-PstI-SalI-XbaI-BamHI-SmaI-KpnI-SacI-EcoRI
             SELECTION
                #resistance  Ap
                #indicator beta-galactosidase
             SUMMARY  pUC19c    #length 2686   #checksum 4465.
COMMENT      ApEinfo:methylated:1
FEATURES             Location/Qualifiers
     source          join(2287..2879,2886..3281,1..2260)
                     /organism="Cloning vector pUC19c"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:174689"
                     /label=source:Cloning vector pUC19c
                     /ApEinfo_fwdcolor=pink
                     /ApEinfo_revcolor=pink
     source          2287..2879
                     /organism="Bacillus anthracis str. Sterne"
                     /mol_type="genomic DNA"
                     /strain="Sterne"
                     /db_xref="taxon:260799"
                     /label=source:Bacillus anthracis str. Sterne
                     /ApEinfo_fwdcolor=#c0c0c0
                     /ApEinfo_revcolor=#c0c0c0
     primer_bind     complement(2897..2914)
                     /label=M13-rev (DNA Core)
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
     gene            2354..2839
                     /locus_tag="BAS1509"
                     /label=gene
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
     primer_bind     2216..2233
                     /label=M13-fwd
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
     CDS             2354..2839
                     /locus_tag="BAS1509"
                     /inference="non-experimental evidence, no additional
                     details recorded"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAT53827.1"
                     /db_xref="GI:49178451"
                     /translation="MNIRTLVGNGILAALYIAVSMLIQPFGFTNVQFRISEMFNHLVV
                     FNKKAIYGIVLGVFLTNLFFSPMIAYDLVFGVGQSILALVATIISMRFIKGVWARMIF
                     NTVIFTITMFMIAIELHLAFDLPFMLTWLTCAVGEFVVMAIGMPVMYWINKRVQFERF
                     M"
```

```
                /label=BAS1509 conserved hypothetical protein
                /ApEinfo_fwdcolor=#8080ff
                /ApEinfo_revcolor=#ffff00
rep_origin      complement(1..683)
                /label=ColE1 origin
                /ApEinfo_fwdcolor=gray50
                /ApEinfo_revcolor=gray50
CDS             complement(2077..2145)
                /label=LacZ alpha
                /ApEinfo_fwdcolor=#6495ed
                /ApEinfo_revcolor=#6495ed
misc_binding    complement(2920..2942)
                /label=LacO
                /ApEinfo_fwdcolor=#6495ed
                /ApEinfo_revcolor=#6495ed
primer_bind     complement(2851..2885)
                /label=DS_HindIII_Term_01R
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
CDS             complement(781..1440)
                /label=AmpR
                /ApEinfo_fwdcolor=yellow
                /ApEinfo_revcolor=yellow
primer_bind     2287..2312
                /label=DS_TSS_01F
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     complement(1867..1889)
                /label=pGEX 3 primer
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     2261..2286
                /label=DS_XbaI_T7prom_F
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     1869..1886
                /label=pGEX3_01R
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     complement(2261..2286)
                /label=DS_XbaI_T7prom_R
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     2199..2216
                /label=pUC19_Univ_01F
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     1602..1619
                /label=Amp_01R
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
primer_bind     complement(2928..2945)
                /label=LacO_01F
                /ApEinfo_fwdcolor=#ff8000
                /ApEinfo_revcolor=#800080
misc_feature    2261..2266
                /label=primer-Added Seq
                /ApEinfo_fwdcolor=#ff80ff
                /ApEinfo_revcolor=green
primer_bind     2267..2286
                /label=T7
                /ApEinfo_fwdcolor=cyan
                /ApEinfo_revcolor=green
```

```
     misc_feature    complement(2880..2885)
                     /label=primer-Added Seq(1)
                     /ApEinfo_label=primer-Added Seq
                     /ApEinfo_fwdcolor=cyan
                     /ApEinfo_revcolor=green
     terminator      2849..2878
                     /label=Terminator_StemLoop-polyU
                     /ApEinfo_fwdcolor=#ff0000
                     /ApEinfo_revcolor=green
     misc_structure  2303..2347
                     /label=preQ1-riboswitch (Rfam def)
                     /ApEinfo_fwdcolor=#ff00ff
                     /ApEinfo_revcolor=green
ORIGIN
        1 ggccgcgttg ctggcgtttt tccataggct ccgcccccct gacgagcatc acaaaaatcg
       61 acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc
      121 tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg cttaccggat acctgtccgc
      181 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc
      241 ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg
      301 ctgcgcctta tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc
      361 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga
      421 gttcttgaag tggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc
      481 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac
      541 caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg
      601 atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc
      661 acgttaaggg attttggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa
      721 ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta
      781 ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt
      841 tgcctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag
      901 tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag caataaacca
      961 gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc
     1021 tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt
     1081 tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg tttggtatgg cttcattcag
     1141 ctccggttcc caacgatcaa ggcgagttac atgatccccc atgttgtgca aaaaagcggt
     1201 tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat
     1261 ggttatggca gcactgcata attctcttac tgtcatgcca tccgtaagat gcttttctgt
     1321 gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc
     1381 ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat
     1441 cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag
     1501 ttcgatgtaa cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt
     1561 ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg
     1621 gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt atcagggtta
     1681 ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc
     1741 gcgcacattt ccccgaaaag tgccacctga cgtctaagaa accattatta tcatgacatt
     1801 aacctataaa aataggcgta tcacgaggcc ctttcgtctc gcgcgtttcg gtgatgacgg
     1861 tgaaaacctc tgacacatgc agctcccgga gacggtcaca gcttgtctgt aagcggatgc
     1921 cgggagcaga caagcccgtc agggcgcgtc agcgggtgtt ggcgggtgtc ggggctggct
     1981 taactatgcg gcatcagagc agattgtact gagagtgcac catatgcggt gtgaaatacc
     2041 gcacagatgc gtaaggagaa aataccgcat caggcgccat tcgccattca ggctgcgcaa
     2101 ctgttgggaa gggcgatcgg tgcgggcctc ttcgctatta cgccagctgg cgaaaggggg
     2161 atgtgctgca aggcgattaa gttgggtaac gccagggttt tcccagtcac gacgttgtaa
     2221 aacgacggcc agtgaattcg agctcggtac ccggggatcc tctagataat acgactcact
     2281 atagggtgtt gcttaaaaaa cgaataacgt ggttcgaaac catcccacgt aaaaaaacta
     2341 aggagatttt gtcgtgaata ttagaacatt agtcggtaat ggtattttag cggcattata
     2401 tattgctgtt tctatgctta ttcagccatt tggctttacg aatgtacagt ttcgtatttc
     2461 agagatgttt aatcatctcg ttgtatttaa taagaaagca atttacggaa ttgtattagg
     2521 tgtattttta acgaatctct ttttctcacc tatgattgct tacgatttag tatttggagt
     2581 agggcaatct attcttgcat tagttgcaac cattatttct atgcgattca ttaaaggtgt
     2641 ttgggctcgt atgattttta atacagttat ctttacaatt acaatgttta tgattgcaat
     2701 tgaacttcat cttgcatttg atttaccatt tatgttgact tggttaacat gtgcagtcgg
     2761 tgaatttgtt gtcatggcca ttggtatgcc tgtaatgtac tggattaata aacgagtaca
     2821 atttgaaaga tttatgtaat agatgaaaga gctattccta tagggatagc tctttttaa
     2881 agcttggcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc gctcacaatt
```

```
     2941 ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta atgagtgagc
     3001 taactcacat taattgcgtt gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc
     3061 cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct
     3121 tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg agcggtatca
     3181 gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac
     3241 atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa a
//
```

C.2 Sequence of the Bas *mRNA used for* in vitro *translation and filter binding assays*

Nucleotides in the 5' and 3' UTRs are shown in lower case.

ggguguugcuuaaaaaacgaauaacgugguucgaaaccaucccacguaaaaaaacuaaggagauuuugucGUGAAUAUUAGAAC
AUUAGUCGGUAAUGGUAUUUUAGCGGCAUUAUAUAUUGCUGUUUCUAUGCUUAUUCAGCCAUUUGGCUU
UACGAAUGUACAGUUUCGUAUUUCAGAGAUGUUUAAUCAUCUCGUUGUAUUUAAUAAGAAAGCAAUUUAC
GGAAUUGUAUUAGGUGUAUUUUUAACGAAUCUCUUUUUCUCACCUAUGAUUGCUUACGAUUUAGUAUUU
GGAGUAGGGCAAUCUAUUCUUGCAUUAGUUGCAACCAUUAUUUCUAUGCGAUUCAUUAAAGGUGUUUGG
GCUCGUAUGAUUUUUAAUACAGUUAUCUUUACAAUUACAAUGUUUAUGAUUGCAAUUGAACUUCAUCUUG
CAUUUGAUUUACCAUUUAUGUUGACUUGGUUAACAUGUGCAGUCGGUGAAUUUGUUGUCAUGGCCAUUG
GUAUGCCUGUAAUGUACUGGAUUAAUAAACGAGUACAAUUUGAAAGAUUUAUGUAAuagaugaaagagcuauu
ccuauagggauagcucuuuuuuaaagcu

C.3 Unix commands for analysis of SOLiD RNA transcriptome profiling data and related output

View personal webpage
http://www-personal.umich.edu/~palund/GSM343460.sorted.bam

sftp.itd.umich.edu

View CPU usage data and process list
top

Count occurrences of a string in a file
grep -o -c stringOfInterest fileName

NCBI GEO Data set weblink
http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13543
GSM341367 Mid-log_air_rep1_Illumina
GSM341368 Mid-log_CO2_rep1_Illumina
GSM343458 Early-log_air_SOLiD
GSM343459 Mid-log_CO2_SOLiD
GSM343460 Mid-log_air_SOLiD
GSM343461 Late-log_CO2_SOLiD
GSM343462 Late-sporulation_air_SOLiD
GSM343463 Late-sporulation_CO2_SOLiD

Download SOLiD seq data from NCBI's ftp server
palund% ftp ftp.ncbi.nlm.nih.gov
cd geo/series/GSE13nnn/GSE13543/suppl
get GSE13543_RAW.tar
bye

```
cd /sra/sra-instant/reads/ByStudy/sra/SRP/SRP001/SRP001274
get all the things
```

Extract and unzip files

```
tar -xf GSE13543_RAW.tar
gunzip GSM343460_5*
gunzip GSM343458_1*
```

Compress files when done

```
tar -cvf dirName.tar dirName/
gzip dirname.tar
```

Edit GMS343459.csfasta file to make usable

Remove first 17 commented lines from csfasta:

```
[nwalter-imac:~/Desktop/anthraxSeq/MLC-S] palund% sed -e
'1,17d' ../SOLiDreads/GSM343459_2A.csfasta.txt >../SOLiDreads/GSM343459_2A.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/LLC-S] palund% sed -e
'1,17d' ../SOLiDreads/GSM343459_2B.csfasta.txt >../SOLiDreads/GSM343459_2B.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e '1,17d'
GSM343461_6A.csfasta.txt > GSM343461_6A.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e '1,17d'
GSM343461_6B.csfasta.txt > GSM343461_6B.csfasta
```

Remove first 3 commented lines from qual file:

```
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e
'1,3d' ../SOLiDreads/GSM343459_2A.qual > ../SOLiDreads/GSM343459_2A.qual.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343459_2A.qual
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343459_2A.qual.tmp
GSM343459_2A.qual

[nwalter-imac:~/Desktop/anthraxSeq/MLC-S] palund% sed -e
'1,3d' ../SOLiDreads/GSM343459_2B.qual >../SOLiDreads/GSM343459_2B.qual.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343459_2B.qual
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343459_2B.qual.tmp
GSM343459_2B.qual

[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e '1,3d'
GSM343461_6A_QV.qual > GSM343461_6A_QV.qual.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343461_6A_QV.qual
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343461_6A_QV.qual.tmp
GSM343461_6A_QV.qual

[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e '1,3d'
GSM343461_6B_QV.qual > GSM343461_6B_QV.qual.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343461_6B_QV.qual
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343461_6B_QV.qual.tmp
GSM343461_6B_QV.qual
```

******Errors with read names in csfasta file containing extra characters. Need to edit files**

http://unix.stackexchange.com/questions/24140/return-only-the-portion-of-a-line-after-a-matching-pattern

Trim extra characters after the F3 in read name

Slow way

```
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% sed -e
's/\(^.*\)\(F3.*$\)/\1F3/' GSM343461_6A.csfasta > GSM343461_6A.csfasta.tmp
```

Fast way

```
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cut -d "," -f1
GSM343461_6B.csfasta>GSM343461_6B.csfasta.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cut -d "," -f1
GSM343461_6A.csfasta>GSM343461_6A.csfasta.tmp

[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cut -d "," -f1
GSM343459_2A.csfasta > GSM343459_2A.csfasta.tmp
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cut -d "," -f1
GSM343459_2B.csfasta > GSM343459_2B.csfasta.tmp

[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343459_2A.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% rm GSM343459_2B.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343459_2B.csfasta.tmp
GSM343459_2B.csfasta
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% mv GSM343459_2A.csfasta.tmp
GSM343459_2A.csfasta
```

Combine A & B files into single csfasta and qual files

```
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cat GSM343461_6A_QV.qual
GSM343461_6B_QV.qual >GSM343461.qual
[nwalter-imac:~/Desktop/anthraxSeq/SOLiDreads] palund% cat GSM343461_6A.csfasta
GSM343461_6B.csfasta >GSM343461.csfasta
```

Build colorspace index for SOLiD read

```
bowtie-build -C -f genome.fasta,plasmid.fasta refIndexName
[nwalter-imac:~/Desktop/anthraxSeq/refGenome] palund% bowtie-build -C -f
AE017225.1.fasta,AE017336.2.fasta ../sterneIndex/AE017225_AE017336
```

Inspect index once built

```
bowtie-inspect -s refIndexName
[nwalter-imac:~/Desktop/anthraxSeq/sterneIndex] palund% bowtie-inspect -s
AE017225_AE017336
```

Colorspace alignment

```
bowtie refIndexName -C --best -t -f short_read.csfasta.txt -Q short_read.qual -S
output.sam
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343460_5.csfasta.txt -Q ../SOLiDreads/GSM343460_5.qual -S
GSM343460.sam

Time loading reference: 00:00:00
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:00
Seeded quality full-index search: 01:27:38
# reads processed: 34513145
# reads with at least one reported alignment: 15410066 (44.65%)
# reads that failed to align: 19103079 (55.35%)
Reported 15410066 alignments to 1 output stream(s)
Time searching: 01:27:38
Overall time: 01:27:38
```

want to see if number of aligned reads is significantly increased by using the try hard option (-y) in aligner. "Try as hard as possible to find valid alignments when they exist"

```
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -f -
y ../SOLiDreads/GSM343460_5.csfasta.txt -Q ../SOLiDreads/GSM343460_5.qual -S
GSM343460_tryHard.sam

Time loading reference: 00:00:00
```

```
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:00
Seeded quality full-index search: 01:33:12
# reads processed: 34513145
# reads with at least one reported alignment: 15410066 (44.65%)
# reads that failed to align: 19103079 (55.35%)
Reported 15410066 alignments to 1 output stream(s)
Time searching: 01:33:12
Overall time: 01:33:12
```

-->Looks like tryhard option doesn't improve the % of aligned reads

```
[nwalter-imac:~/Desktop/anthraxSeq/ELA-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343458_1.csfasta.txt -Q ../SOLiDreads/GSM343458_1.qual -S
GSM343458.sam

Time loading reference: 00:00:00
Time loading forward index: 00:00:01
Time loading mirror index: 00:00:00
Seeded quality full-index search: 01:30:34
# reads processed: 34099100
# reads with at least one reported alignment: 14418519 (42.28%)
# reads that failed to align: 19680581 (57.72%)
Reported 14418519 alignments to 1 output stream(s)
Time searching: 01:30:36
Overall time: 01:30:36

[nwalter-imac:~/Desktop/anthraxSeq/LSA-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343462_7.csfasta -Q ../SOLiDreads/GSM343462_7.qual -S GSM343462.sam
> bowtieStats.out &
[1] 15448
[nwalter-imac:~/Desktop/anthraxSeq/LSA-S] palund% Time loading reference: 00:00:01
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:00
Seeded quality full-index search: 03:07:38
# reads processed: 45061700
# reads with at least one reported alignment: 13880515 (30.80%)
# reads that failed to align: 31181185 (69.20%)
Reported 13880515 alignments to 1 output stream(s)
Time searching: 03:07:39
Overall time: 03:07:39

[nwalter-imac:~/Desktop/anthraxSeq/LLC-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343461.csfasta  -Q ../SOLiDreads/GSM343461.qual  -S GSM343461.sam
Time loading reference: 00:00:01
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:00
Seeded quality full-index search: 03:03:53
# reads processed: 52412661
# reads with at least one reported alignment: 32226945 (61.49%)
# reads that failed to align: 20185716 (38.51%)
Reported 32226945 alignments to 1 output stream(s)
Time searching: 03:03:54
Overall time: 03:03:54

[nwalter-imac:~/Desktop/anthraxSeq/MLC-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343459.csfasta -Q ../SOLiDreads/GSM343459.qual -S GSM343459.sam
Time loading reference: 00:00:00
Time loading forward index: 00:00:01
```

```
Time loading mirror index: 00:00:00
Seeded quality full-index search: 01:41:31
# reads processed: 28664981
# reads with at least one reported alignment: 18204563 (63.51%)
# reads that failed to align: 10460418 (36.49%)
Reported 18204563 alignments to 1 output stream(s)
Time searching: 01:41:33
Overall time: 01:41:33

[nwalter-imac:~/Desktop/anthraxSeq/LSC-S] palund%
bowtie ../sterneIndex/AE017225_AE017336 -C --best -t -
f ../SOLiDreads/GSM343463.csfasta -Q ../SOLiDreads/GSM343463.qual -S GSM343463.sam >
bowtieStats.out &
[1] 15701
[nwalter-imac:~/Desktop/anthraxSeq/LSC-S] palund% Time loading reference: 00:00:00
Time loading forward index: 00:00:00
Time loading mirror index: 00:00:00
Too few quality values for read: 449_1398_299_F3
       are you sure this is a FASTQ-int file?
Seeded quality full-index search: 02:10:19
Time searching: 02:10:19
Overall time: 02:10:19
```

Convert SAM to BAM

```
samtools view -S -b -o my.bam my.sam
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343460.bam
GSM343460.sam

[nwalter-imac:~/Desktop/anthraxSeq/ELA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343458.bam
GSM343458.sam

[nwalter-imac:~/Desktop/anthraxSeq/MLC-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343459.bam
GSM343459.sam &

[nwalter-imac:~/Desktop/anthraxSeq/LLC-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343461.bam
GSM343461.sam

[nwalter-imac:~/Desktop/anthraxSeq/LSA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343462.bam
GSM343462.sam

[nwalter-imac:~/Desktop/anthraxSeq/LSC-S] palund%  ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools view -S -b -o GSM343463.bam
GSM343463.sam
```

Sort and create index for BAM file

```
samtools sort my.bam my.sorted
samtools index my.sorted.bam
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools sort GSM343460.bam GSM343460.sorted
[bam_sort_core] merging from 8 files...

[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools index GSM343460.sorted.bam

[nwalter-imac:~/Desktop/anthraxSeq/ELA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools sort GSM343458.bam GSM343458.sorted
```

```
[bam_sort_core] merging from 8 files...

[nwalter-imac:~/Desktop/anthraxSeq/ELA-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools index GSM343458.sorted.bam

[nwalter-imac:~/Desktop/anthraxSeq/MLC-S] palund% ~/Desktop/samtools-
0.1.18_0.darwin_11.x86_64/opt/local/bin/samtools sort GSM343459.bam  GSM343459.sorted
&
```

Convert BAM to BED file

```
bamToBed -i in.bam > out.bed
[nwalter-imac:~/Desktop/anthraxSeq/ELA-S] palund% bamToBed -i
GSM343458.sorted.bam>GSM343458.sorted.bed
```

Search for a particular read within a file

```
grep serachTerm fileName
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% grep 966_462_538_F3 GSM343460.sam
966_462_538_F3      0      gi|49176966|gb|AE017225.1| 117    255    33M    *      0
       0      TCATTGCTATAGCTACTTTTTTTTGATATTATA TCKN=:J[MIRRSTRKQXI39O[7!,DG*!'<8 XA:i:1
       MD:Z:3NM:i:0  CM:i:3

[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% grep 966_462_538_F3
GSM343460.sorted.bed
gi|49176966|gb|AE017225.1| 116    149    966_462_538_F3      255    +
```

When writing the bed file from the SAM/BAM file, all starting coordinates are adjusted from the bam i.e., a read that aligned to nt 117-149 of the chromosome is reported as 116 -149 in the bed file (see above). In bed files, features start at coordinate+1.

Use BEDtools to calculate coverage at each nt

```
coverageBed -a alignedReads.sorted.bed -b areaOfInterest.bed -s -d>
areaOfInterest_coverage.bed
[nwalter-imac:~/Desktop/anthraxSeq/MLA-S] palund% coverageBed -a GSM343460.sorted.bed
-b pXO1_minus.bed -s -d> pXO1_GSM343460coverage.bed
```

C.4 Sequence comparisons for S1 homologs from selected bacterial species.

Alignment

```
Learn how to print this alignment in color   Repeat number: 1          UniProt ID

1    --MRMNDFL-EGYVFRPIKAGDIVKGKVVKVLEDGVVAD-IGYKGEAFVPKEELSLNPEF   56   Q8RA75    Q8RA75_THETN
1    MTESFAQLFEESLKEIETRPGSIVRGVVVAIDKDVVLVD-AGLKSESAIPAEQFKNAQGE   59   P0AG67    RS1_ECOLI
1    MTEEMN---Q--IDVQVPEVGDVVKGIVTKVEDKHVDVEIINVKQSGIIPISELSSLHVE   55   P38494    RS1H_BACSU
1    MVEKMN---EEVMDSKELQVGDVVTGSVTKVEEKQVLVN-VGYKTDGVIPISELANVHIE   56   Q81SX5    Q81SX5_BACAN
         :             . *.:* * *. : .. * .:    * .. :* .::

57   DVLSNFHEGEELDLYIVMVENEEGSVVASKIKAEEKLGEEKIEKAYKNREIVKGEVVKVV  116   Q8RA75    Q8RA75_THETN
60   LE---IQVGDEVDVALDAVEDGFGETLLSREKAKRHEAWITLEKAYEDAETVTGVINGKV  116   P0AG67    RS1_ECOLI
56   KASDVVKVDDELDLKVTKVE--DDALILSKRAVDADRAWEDLEKKFETKEVFEAEVKDVV  113   P38494    RS1H_BACSU
57   KASDVVELDQTLELKVIKLE--ENDLVLSKRAVDAEKAWVELQEKFNSGHVFDVTVKDIV  114   Q81SX5    Q81SX5_BACAN
         ..  : ::: :  :*      : *:  .. . .   ::: ::  . .   :   *
2
117  KGGVIAYSLGKRLFIPASHLSTKRVENLEEYLGKILELLIIEYVPG-KKIIASRKEVIKL  175   Q8RA75    Q8RA75_THETN
117  KGGFTVELNGIRAFLPGSLVDVRPVRDTLHLEGKELEFKVIKLDQKRNNVVVSRRAVIES  176   P0AG67    RS1_ECOLI
114  KGGLVVDI-GVRGFIPASLVEAHFVEDFTDYKGKTLSLLVVELDRDKNRVILSHRAVVES  172   P38494    RS1H_BACSU
115  NGGLVVDL-GVRGFIPASLVEVHYVEDFTDYKGKTLAVKIVELDREKNRVILSHKAVVEL  173   Q81SX5    Q81SX5_BACAN
     :**. .   * * *:*.* :..: *.:  .  ** * . :::     :.:: *:: *::

176  EEEKMKAELLERLEEGDVIEGKVSRIIPKGAFVNLGGMDGFIPISELGHGRIKSPEEVIK  235   Q8RA75    Q8RA75_THETN
177  ENSAERDQLLENLQEGMEVKGIVKNLTDYGAFVDLGGVDGLLHITDMAWKRVKHPSEIVN  236   P0AG67    RS1_ECOLI
173  EQANKKQELLQSLEVGSVLDGKVQRLTDFGAFVDIGGIDGLVHISQLSHSHVEKPSDVVE  232   P38494    RS1H_BACSU
174  ELDSKKKEAISSLKEGDVVEGTVQRLTDFGAFVNVGGVDGLVHISQISHERVEQPSEVLE  233   Q81SX5    Q81SX5_BACAN
     *    : : :. *: *  :.* *..:   ****::**:**:: *:::.  ::: *.::::
3                                                4
236  PGHKVSAYVLKVDKKEGKITLSLKKLLPNPWENVKEKYHEGDILKAKITRIMPFGVFAEL  295   Q8RA75    Q8RA75_THETN
237  VGDEITVKVLKFDRERTRVSLGLKQLGEDPWVAIAKRYPEGTKLTGRVTNLTDYGCFVEI  296   P0AG67    RS1_ECOLI
233  EGQEVKVKVLSVDRDNERISLSIKDTLPGPWNQIGEKVKPGDVLEGTVQRLVSFGAFVEI  292   P38494    RS1H_BACSU
234  OGOKVKVKVLSVDADTORISLSIKAAOPGPWENVGGEVKAGDIREGVVKRLVTFGAFVEI  293   Q81SX5    Q81SX5_BACAN
      *.::.. **..* .  ::*.:*     **  :  .   *    . : .:  :* *.*:

296  EPGVEGLVHKNNLEKSIKSYSPEDII------TVEVLEVS--SDRISLKEV----PPFEG  343   Q8RA75    Q8RA75_THETN
297  EEGVEGLVHVSEMDWTNKNIHPSKVVNVGDVVEVMVLDIDEERRRISLGLKQCKANPW--  354   P0AG67    RS1_ECOLI
293  LPGVEGLVHISQISNKH-IGTPHEVLEEGQTVKVKVLDVNENEERISLSMRELEETPKAD  351   P38494    RS1H_BACSU
294  LPGVEGLVHVSQIANRH-VKNPNEVLEMGQEVKVKVLEVHVAEKRISLSIKEALEEN-NV  351   Q81SX5    Q81SX5_BACAN
       ******* .::        * .::       * **::     ****
                                                                5
344  EDDIKQT---EYQELTINLGEILKKKF-------------------------------  367   Q8RA75    Q8RA75_THETN
355  ----QQFAETH------NKGDRVEGKIKSITDFGIFIGLDGGIDGLVHLSDISWNVAGEE  404   P0AG67    RS1_ECOLI
352  QEDYRQYQAKEETSTGFQLGDLIGDKLNKLK---------------------------  382   P38494    RS1H_BACSU
352  TEDYSQYEPNA-DSATFQLSDIIGEQLKKLKK--------------------------  382   Q81SX5    Q81SX5_BACAN
         *           : .: :  ::
                                                  6
368  ------------------------------------------------------------  367   Q8RA75    Q8RA75_THETN
405  AVREYKKGDEIAAVVLQVDAERERISLGVKQLAEDPFNNWVALNKKGAIVTGKVTAVDAK  464   P0AG67    RS1_ECOLI
383  ------------------------------------------------------------  382   P38494    RS1H_BACSU
383  ------------------------------------------------------------  382   Q81SX5    Q81SX5_BACAN

368  ------------------------------------------------------------  367   Q8RA75    Q8RA75_THETN
465  GATVELADGVEGYLRASEASRDRVEDATLVLSVGDEVEAKFTGVDRKNRAISLSVRAKDE  524   P0AG67    RS1_ECOLI
383  ------------------------------------------------------------  382   P38494    RS1H_BACSU
383  ------------------------------------------------------------  382   Q81SX5    Q81SX5_BACAN

368  --------------------------------  367   Q8RA75    Q8RA75_THETN
525  ADEKDAIATVNKQEDANFSNNAMAEAFKAAKGE  557   P0AG67    RS1_ECOLI
383  --------------------------------  382   P38494    RS1H_BACSU
383  --------------------------------  382   Q81SX5    Q81SX5_BACAN
```

Figure C.4-1 Alignment of S1 homologs.

Comparison of homologous S1 proteins from *T. tengcongensis* (THETN), *E. coli* (ECOLI), *B. subtilis* (BACSU) and *B. anthracis* (BACAN). Protein alignment was performed using the Clustal Omega program at the UniProt website (http://www.uniprot.org/align/). OB-fold domain repeats are color coded (6 in *E. coli* S; 4 in S1 of the other species shown).

Appendix Table C.4-1 DNA primers for generating *in vitro* transcription templates: *Tte* mRNA truncations.

For reverse primers, nucleotides in lower case are complementary to the riboswitch aptamer region; capital letters indicate nucleotides that are complementary to the ORF. Extinction coefficients are provided by Life Technologies.

Name	ε_{260} (M^{-1} cm^{-1})	Sequence (5′ to 3′)	PCR annealing temperature (°C)
Forward:			
pUC19_Univ_01F	181 500	TTTCCCAGTCACGACGTT	--
Reverse:			
Tte1564_+30_01R	273 600	AGCTAAATCTTTTATTCTTTTTTTGG	60
Tte1564_+22_01R	216 500	CTTTTATTCTTTTTTTGGGCAC	60
Tte1564_+6_01R	192 100	GGGCACaaaattacctc	55
Tte1564_+3_01R	201 900	CACaaaattacctcccttg	57
Tte1564_-1_01R	220 800	aaaattacctcccttgttttg	59
Tte1564_-11_01R	191 100	cccttgttttgttaactgg	57

Appendix Table C.4-2 Complete sequences for *Tte* mRNA truncations and DNA capture strand.

Extinction coefficients are calculated using OligoCalc[140] unless otherwise noted. Nucleotides in italics are part of the capture strand-RNA hybrid; the sequence of the aptamer construct used in Suddala and Rinaldi *et al.*[83] is underlined; capital letters indicate nucleotides in the ORF.

Name	ε_{260} (M^{-1} cm^{-1})	Sequence (5′ to 3′)
Capture strand	227 700[a]	*cattttgttgctcactgccc*atata-biotin
Tte^{+30}	1 277 139	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg agguaauuuuGUGCCCAAAAAAAGAAUAAAAGAUUUAGCU
Tte^{+22}	1 183 432	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg agguaauuuuGUGCCCAAAAAAAGAAUAAAAG
Tte^{+6}	945 180	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg agguaauuuuGUGCCC
Tte^{+3}	918 274	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg agguaauuuuGUG
Tte^{-1}	881 057	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg agguaauuuu
Tte^{-11}	757 576	*gggcagugagcaacaaaaug*cucaccugggucgcaguaaccccaguuaacaaaacaaggg

[a] Extinction coefficient provided by IDT.

C.5 Example protocol for 70S initiation complex formation assay

Date	**5/8/2013**									
ProjExp #	**CB8**									
Description:	**Initiation complex formation assay w/ truncated Tte transcripts**									
Initiation Complex Formation										
	3.1	**rxns**								component Mg2+
Order of Add'n	mMix for 3.1	Vol (uL)	[Stock]	[Final]	**40**	**uL final vol**	*Order of Add'n*		mM	nmol
1	12.4	4.00	10	1	x Polymix buffer (w/o PO4)			*4*	50	200
9	6.2	2.00	100	5	mM KH2PO4/K2HPO4, pH 7.5			*4*		
6	1.24	0.40	100	1	mM DTT			*4*		
7	0.124	0.04	1000	1	mM MgCl2			*4*	1000	40
5	2.48	0.80	100	2	mM GTP		**Proj/Expt**	*4*		
2	2.011	0.65	185	3	uM IF1		AA27	*4*	0	0
3	2.163	0.70	172	3	uM IF2		AA26	*4*	10	6.98
4	3.739	1.21	99.5	3	uM IF3		AA22	*4*	0	0
10	47.634	**15.366**	7.81	3	**uM tRNA fMet**		AC30	*4*		
8	23.846	**7.69**	7.80	**1.5**	uM 70S		AC17	*4*	7	53.8
--	--	**1.000**	80.00	2	**uM Capture strand_01R**			*2*		
--	--			**1**	**uM mRNA**		Various	*3*		
		6.15			H2O			*1*		
		40	uL Total						[Mg2+] final:	7.52 mM
						32.851	uL of mMix per rxn			
	#1	**#2**	**#3**							
Polymix	4.00	4.00	4.00							
K-Phosphate	2.00	2.00	2.00			3.190442	% glycerol			
DTT	0.40	0.40	0.40							
MgCl2	0.04	0.04	0.04							
GTP	0.80	0.80	0.80							
IF1	0.65	0.65	0.65							
IF2	0.70	0.70	0.70							
IF3	1.21	1.21	1.21			In mMix				
tRNA-fMet	15.37	15.37	15.37			Added to tube				
70S	7.69	7.69	7.69							
Capture Strand	**1.000**	**1.000**	**1.000**							
[mRNA]	*15*	*15*	*9.65*			**Rxn**	*mRNA*			
mRNA	**2.667**	**2.667**	**4.147**			**#1**	*CC22 Tte1564_+30-Cy3*			
mQH2O	**3.48**	**3.48**	**2.00**			**#2**	*CC22 Tte1564_-11-Cy3*			
Total	40.00	40.00	40.00			**#3**	*CC31_Tte1564-HindIII*			
	35S-Met			**Sucrose Cushion**						
Half-life	87.4		*Add'n order*	1	X Buffered Sucrose		C1	V1	V2	
Orig Assay Date	2/26/2013		*3*	1	X Polymix w/o PO4		10	900	9000	
Today	5/8/2013		*4*	5	mM KH2PO4/K2HPO4		100	450		
			5	1	mM DTT		1000	9		
Decay Factor, f	0.569450767		*1*	1.1	M Sucrose		1.375	7200		
			6	15	mM addt'l MgCl2*		1000	135		
2.5	35S Met		*7*	0.5	mM EDTA		500	9		
20	Met							8703		
40	uM tRNA		*2*		final Mg2+= 20mM total			297	mQH2O	
0.205045888	Charging efficiency							9000		
8.201835529	uM fMet-tRNA-fMet									
0.91	uM 35fMet-tRNA-fMet, initial				Account for charging efficiency when calculating					
0.52	uM 35fMet-tRNA-fMet x decay factor				effective tRNA-fMet concentration					
7.81	**uM fMet-tRNA-fMet Today** (AC33)									
					400uL charging rxn w/ 10uM tRNA -> assume that all					
					tRNAs pelleted and are resuspended in 100uL					
					30mM KOAc+1mM DTT					

Protocol:

0) Thaw RNAs in water on wet ice. Vortex, spin down and store on wet ice. Thaw capture strands at RT.

1) Aliquot mRNA, water and Capture strand to halfsie tubes. Heat in **90C Cu bath 45sec** then let cool at **RT for 15 min**. Transfer to wet ice. Use capture strand only for no mRNA control.

2) While mRNA is cooling, prepare sucrose cushion in 15 mL falcon tube, store on wet ice.

3) Thaw aliquot of ribosomes from -80C in water on wet ice. Prepare mMix in order noted. Thaw tRNAs on wet ice, and add last on hot bench.

4) Pipet to mix and aliquot to rxn tubes containing annealed mRNA. Place in **37C water bath** for **45 min** (actual time = **52 min**).

5) During rxn, pipet 1.3mL of sucrose cushion into tubes and store in rotor in cold room. Pre-cool ultracentrifuge to 4C.

6) Transfer rxn tubes to wet ice and cout 1uL. Dilute IC rxn to 200uL with dilution buffer.

7) Carefully layer rxn onto cushion using P200: touch tip to surface then pull back to side wall and pipet slowly.

8) Spin at **69k x rpm at 4C for 2 hrs** to pellet ICs. (actual time = 2.5hr)

9) After spin, carefully pipet off cushion using P1000 and P200 being carfeul not to mix layers. Do not reuse tips.

10) Carefully resuspend pellet in Resuspension buffer using resuspension buffer, transfer to halfise tube on wet ice. Count 1uL.

11) Aliquot to tubes and store on dry ice in rad freezer.

			Dilution buffer				
		Add'n order	1	X Buffered Sucrose	C1	V1	V2
1	**X Polymix Buffer w/o PO4**	*2*	1	X Polymix w/o PO4	10	100	1000
5	mM MgOAc	*3*	5	mM KH2PO4/K2HPO4	100	50	
0.5	mM CaCl2	*5*	1	mM DTT	1000	1	
8	mM Putrescine	*4*	15	mM addt'l MgCl2*	1000	15	
1	mM Spermidine					166	
5	*mM KH2PO4/K2HPO4, pH 7.6*	*1*		final Mg2+= 20mM total		834	mQH2O
95	mM KCl					1000	
5	mM NH4Cl						
1	*mM DTT*		**Pellet Resuspension Buffer**				
Omitted when making 10X buffer mix		*Add'n order*	1	X Buffer	C1	V1	V2
Adjust pH to 7.5 w/ HCl, KOH then mQH2O QS to final		*2*	1	X Polymix w/o PO4	10	50	500
		3	5	mM KH2PO4/K2HPO4,	100	25	
		4	1	mM DTT	1000	0.5	
						75.5	
		1		[Mg2+] = 5mM		424.5	mQH2O
						500	

Yield Calculations:

t-half (d)	87.4				
Rxn - before centrifugation		**#1**	**#2**	**#3**	**mMix**
	Measured Counts	699127	837423	763287	849381
	Vol Counted (uL)	1	1	1	1
	Dilution Factor	1	1	1	1
	Counts/uL	699127	837423	763287	849381
	Vol left after counting (uL)	39	393	39	
	Counts carried forward	27265953	329107239	29768193	
	Date	5/8/2013	5/8/2013	5/8/2013	
	% of Expected Counts	*100%*	*120%*	*109%*	27902693.63
					Expected counts in rxn
Rxn - after centrifugation		**#1**	**#2**	**#3**	
	Measured Counts	33932	6441	29810	
	Vol Counted (uL)	1	1	1	
	Dilution Factor	1	1	1	
	Counts/uL	33932	6441	29810	
	Vol** before counting (uL)	40	40	40	
	Counts	1357280	257640	1192400	
	Date	5/8/2013	5/8/2013	5/8/2013	
	Decay Factor	1.000	1.000	1.000	
	Counts (Decay Adjusted)	1357280	257640	1192400	
					Normalization factor
	Incorporation	***15%***	***0%***	***12%***	***33%***

Rxn	mRNA
#1	*CC22 Tte1564_+30-Cy3*
#2	*CC22 Tte1564_-11-Cy3*
#3	*CC31_Tte1564-HindIII*

counts in pellet/ counts before spinning/ norm factor = % incorporation

** added 40uL of resuspension buffer to tube and wash walls, actual recovery is between 42-45uL

Maximum possible recovery is based on limiting reagent and stoichiometry, in this case **mRNA** is limiting: (mRNA/tRNA) = normalization factor.

References

1. Laursen, B.S., Sørensen, H.P., Mortensen, K.K. & Sperling-Petersen, H.U. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* **69**, 101-23 (2005).
2. Milón, P. & Rodnina, M.V. Kinetic control of translation initiation in bacteria. *Crit. Rev. Biochem. Mol. Biol.* **47**, 334-48 (2012).
3. Kudla, G., Murray, A.W., Tollervey, D. & Plotkin, J.B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255-58 (2009).
4. Scharff, L.B., Childs, L., Walther, D. & Bock, R. Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genet.* **7**, e1002155 (2011).
5. Schmeing, T.M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-42 (2009).
6. Gualerzi, C.O., Fabbretti, A., Brandi, L., Milón, P. & Pon, C.L. Role of the initiation factors in mRNA start site selection and fMet-tRNA recruitment by bacterial ribosomes. *Isr. J. Chem.* **50**, 80-94 (2010).
7. Marshall, R.A., Aitken, C.E. & Puglisi, J.D. GTP hydrolysis by IF2 guides progression of the ribosome into elongation. *Mol. Cell* **35**, 37-47 (2009).
8. Santangelo, T.J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* **9**, 319-29 (2011).
9. Calogero, R.A., Pon, C.L., Canonaco, M.A. & Gualerzi, C.O. Selection of the mRNA translation initiation region by *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6427-31 (1988).
10. Skorski, P., Leroy, P., Fayet, O., Dreyfus, M. & Hermann-Le Denmat, S. The highly efficient translation initiation region from the *Escherichia coli rpsA* gene lacks a Shine-Dalgarno element. *J. Bacteriol.* **188**, 6277-85 (2006).
11. Fuglsang, A. Analysis of 5 ' UTR composition and gene expression: Canonical versus non-canonical start codons. *Biochem. Biophys. Res. Commun.* **335**, 71-75 (2005).
12. Moll, I., Hirokawa, G., Kiel, M.C., Kaji, A. & Bläsi, U. Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic Acids Res.* **32**, 3354-63 (2004).
13. Shine, J. & Dalgarno, L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342-6 (1974).
14. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34-38 (1975).
15. Hui, A. & de Boer, H.A. Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4762-6 (1987).
16. Ringquist, S. *et al.* Translation initiation in *Escherichia coli* - Sequences within the ribosome-binding site. *Mol. Microbiol.* **6**, 1219-29 (1992).

17. Ma, J., Campbell, A. & Karlin, S. Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J. Bacteriol.* **184**, 5733-45 (2002).
18. Chen, H., Bjerknes, M., Kumar, R. & Jay, E. Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* **22**, 4953-7 (1994).
19. Vellanoweth, R.L. & Rabinowitz, J.C. The influence of Ribosome-Binding Site elements on translational efficiency in *Bacillus-subtilis* and *Escherichia-coli in vivo*. *Mol. Microbiol.* **6**, 1105-14 (1992).
20. Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475**, 118-21 (2011).
21. Tuller, T., Waldman, Y.Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.* **107**, 3645-50 (2010).
22. Takyar, S., Hickerson, R.P. & Noller, H.F. mRNA helicase activity of the ribosome. *Cell* **120**, 49-58 (2005).
23. Sørensen, M.A., Fricke, J. & Pedersen, S. Ribosomal protein S1 is required for translation of most, if not all, natural mRNAs in *Escherichia coli in vivo*. *J. Mol. Biol.* **280**, 561-9 (1998).
24. Van Dieijen, G., Van Der Laken, C.J., Van Knippenberg, P.H. & Van Duin, J. Function of *Escherichia coli* ribosomal protein S1 in translation of natural and synthetic messenger RNA. *J. Mol. Biol.* **93**, 351-66 (1975).
25. Tedin, K., Resch, A. & Bläsi, U. Requirements for ribosomal protein S1 for translation initiation of mRNAs with and without a 5´ leader sequence. *Mol. Microbiol.* **25**, 189-99 (1997).
26. Farwell, M.A., Roberts, M.W. & Rabinowitz, J.C. The effect of ribosomal protein S1 from *Escherichia coli* and *Micrococcus luteus* on protein synthesis *in vitro* by *E. coli* and *Bacillus subtilis*. *Mol. Microbiol.* **6**, 3375-83 (1992).
27. Moll, I., Resch, A. & Bläsi, U. Discrimination of 5´-terminal start codons by translation initiation factor 3 is mediated by ribosomal protein S1. *FEBS Lett.* **436**, 213-17 (1998).
28. Subramanian, A.-R. Structure and functions of ribosomal protein S1. *Prog. Nucleic Acid Res. Mol. Biol.* **28**, 101-42 (1983).
29. Dahlberg, A.E. Two forms of the 30 S ribosomal subunit of *Escherichia coli*. *J. Biol. Chem.* **249**, 7673-78 (1974).
30. Duval, M. *et al. Escherichia coli* ribosomal protein S1 unfolds structured mRNAs onto the ribosome for active translation initiation. *PLoS Biol.* **11**, e1001731 (2013).
31. Lauber, M.A., Rappsilber, J. & Reilly, J.P. Dynamics of ribosomal protein S1 on a bacterial ribosome with cross-linking and mass spectrometry. *Mol. Cell. Proteomics* **11**, 1965-76 (2012).
32. Sengupta, J., Agrawal, R.K. & Frank, J. Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11991-6 (2001).
33. Boni, I.V., Isaeva, D.M., Musychenko, M.L. & Tzareva, N.V. Ribosome-messenger recognition - messenger-RNA target sites for ribosomal-protein S1. *Nucleic Acids Res.* **19**, 155-62 (1991).

34. Komarova, A.V., Tchufistova, L.S., Dreyfus, M. & Boni, I.V. AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *J. Bacteriol.* **187**, 1344-9 (2005).
35. Ringquist, S. *et al.* High-affinity RNA ligands to *Escherichia coli* ribosomes and ribosomal protein S1: Comparison of natural and unnatural binding sites. *Biochemistry* **34**, 3640-48 (1995).
36. de Smit, M.H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proc. Natl. Acad. Sci.* **87**, 7668-72 (1990).
37. de Smit, M.H. & van Duin, J. Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J. Mol. Biol.* **235**, 173-84 (1994).
38. de Smit, M.H. & van Duin, J. Translational standby sites: How ribosomes may deal with the rapid folding kinetics of mRNA. *J. Mol. Biol.* **331**, 737-43 (2003).
39. Unoson, C. & Wagner, E.G.H. Dealing with stable structures at ribosome binding sites - Bacterial translation and ribosome standby. *RNA Biol.* **4**, 113-17 (2007).
40. Qu, X., Lancaster, L., Noller, H.F., Bustamante, C. & Tinoco, I. Ribosomal protein S1 unwinds double-stranded RNA in multiple steps. *Proc. Natl. Acad. Sci.* **109**, 14458-63 (2012).
41. Skouv, J., Schnier, J., Rasmussen, M.D., Subramanian, A.R. & Pedersen, S. Ribosomal protein S1 of *Escherichia coli* is the effector for the regulation of its own synthesis. *J. Biol. Chem.* **265**, 17044-9 (1990).
42. Tchufistova, L.S., Komarova, A.V. & Boni, I.V. A key role for the mRNA leader structure in translational control of ribosomal protein S1 synthesis in gamma-proteobacteria. *Nucleic Acids Res.* **31**, 6996-7002 (2003).
43. Skorski, P., Proux, F., Cheraiti, C., Dreyfus, M. & Hermann-Le Denmat, S. The deleterious effect of an insertion sequence removing the last twenty percent of the essential *Escherichia coli rpsA* gene is due to mRNA destabilization, not protein truncation. *J. Bacteriol.* **189**, 6205-12 (2007).
44. Hajnsdorf, E. & Boni, I.V. Multiple activities of RNA-binding proteins S1 and Hfq. *Biochimie* **94**, 1544-53 (2012).
45. Saguy, M., Gillet, R., Skorski, P., Hermann-Le Denmat, S. & Felden, B. Ribosomal protein S1 influences trans-translation *in vitro* and *in vivo*. *Nucleic Acids Res.* **35**, 2368-76 (2007).
46. Tomita, K. Structures and functions of Qbeta replicase: translation factors beyond protein synthesis. *Int. J. Mol. Sci.* **15**, 15552-70 (2014).
47. Brown, D. & Gold, L. Template recognition by an RNA-dependent RNA polymerase: identification and characterization of two RNA binding sites on Q beta replicase. *Biochemistry* **34**, 14765-74 (1995).
48. Takeshita, D., Yamashita, S. & Tomita, K. Molecular insights into replication initiation by Qβ replicase using ribosomal protein S1. *Nucleic Acids Res.* **42**, 10809-22 (2014).
49. Vasilyev, N.N., Kutlubaeva, Z.S., Ugarov, V.I., Chetverina, H.V. & Chetverin, A.B. Ribosomal protein S1 functions as a termination factor in RNA synthesis by Qβ phage replicase. *Nat Commun* **4**, 1781 (2013).
50. Sukhodolets, M.V., Garges, S. & Adhya, S. Ribosomal protein S1 promotes transcriptional cycling. *RNA* **12**, 1505-13 (2006).

51. Lebars, I., Hu, R.M., Lallemand, J.Y., Uzan, M. & Bontems, F. Role of the substrate conformation and of the S1 protein in the cleavage efficiency of the T4 endoribonuclease RegB. *J. Biol. Chem.* **276**, 13264-72 (2001).
52. Dethoff, E.A., Chugh, J., Mustoe, A.M. & Al-Hashimi, H.M. Functional complexity and regulation through RNA dynamics. *Nature* **482**, 322-30 (2012).
53. Smith, A.M., Fuchs, R.T., Grundy, F.J. & Henkin, T.M. Riboswitch RNAs: Regulation of gene expression by direct monitoring of a physiological signal. *RNA Biol.* **7**, 104-10 (2010).
54. Marzi, S., Fechter, P., Chevalier, C., Romby, P. & Geissmann, T. RNA switches regulate initiation of translation in bacteria. *Biol. Chem.* **389**, 585-98 (2008).
55. Kaberdin, V.R. & Bläsi, U. Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol. Rev.* **30**, 967-79 (2006).
56. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24 (2013).
57. Henkin, T.M. Preface — Riboswitches. *Biochim. Biophys. Acta-Gene Regul. Mech.* **1839**, 899 (2014).
58. Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451-63 (2004).
59. Roth, A. *et al.* A riboswitch selective for the queuosine precursor preQ$_1$ contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14**, 308-17 (2007).
60. Meyer, M.M., Roth, A., Chervin, S.M., Garcia, G.A. & Breaker, R.R. Confirmation of a second natural preQ$_1$ aptamer class in *Streptococcaceae* bacteria. *RNA* **14**, 685-95 (2008).
61. Liberman, J.A., Salim, M., Krucinska, J. & Wedekind, J.E. Structure of a class II preQ$_1$ riboswitch reveals ligand recognition by a new fold. *Nat. Chem. Biol.* **9**, 353-5 (2013).
62. Liberman, J.A. *et al.* Structural analysis of a class III preQ$_1$ riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E3485-94 (2015).
63. Jenkins, J.L., Krucinska, J., McCarty, R.M., Bandarian, V. & Wedekind, J.E. Comparison of a preQ$_1$ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* **286**, 24626-37 (2011).
64. Spitale, R.C., Torelli, A.T., Krucinska, J., Bandarian, V. & Wedekind, J.E. The structural basis for recognition of the preQ$_0$ metabolite by an unusually small riboswitch aptamer domain. *J. Biol. Chem.* **284**, 11012-16 (2009).
65. Klein, D.J., Edwards, T.E. & Ferre-D'Amare, A.R. Cocrystal structure of a class I preQ$_1$ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **16**, 343-44 (2009).
66. Peselis, A. & Serganov, A. Structure and function of pseudoknots involved in gene expression control. *WIREs RNA* **5**, 803-22 (2014).
67. Breaker, R.R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43**, 867-79 (2011).
68. Nudler, E. & Mironov, A.S. The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29**, 11-7 (2004).
69. Winkler, W.C. & Breaker, R.R. Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* **59**, 487-517 (2005).
70. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-6 (2004).

71. Batey, R.T., Gilbert, S.D. & Montange, R.K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411-5 (2004).
72. Mandal, M. & Breaker, R.R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **11**, 29-35 (2004).
73. Blount, K.F., Wang, J.X., Lim, J., Sudarsan, N. & Breaker, R.R. Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.* **3**, 44-9 (2007).
74. Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275-9 (2004).
75. Barrick, J.E. & Breaker, R.R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8**, R239 (2007).
76. Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300-4 (2006).
77. Dann, C.E., 3rd *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130**, 878-92 (2007).
78. Regulski, E.E. *et al.* A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.* **68**, 918-32 (2008).
79. Shi, Y., Zhao, G. & Kong, W. Genetic analysis of riboswitch-mediated transcriptional regulation responding to Mn^{2+} in *Salmonella*. *J. Biol. Chem.* **289**, 11353-66 (2014).
80. Anthony, P.C., Perez, C.F., Garcia-Garcia, C. & Block, S.M. Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proc. Natl. Acad. Sci. USA* **109**, 1485-9 (2012).
81. Haller, A., Altman, R.B., Soulière, M.F., Blanchard, S.C. & Micura, R. Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc. Natl. Acad. Sci. USA* **110**, 4188-93 (2013).
82. Haller, A., Rieder, U., Aigner, M., Blanchard, S.C. & Micura, R. Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.* **7**, 393-400 (2011).
83. Suddala, K.C. *et al.* Single transcriptional and translational $preQ_1$ riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res.* **41**, 10462-75 (2013).
84. Wilson, R.C. *et al.* Tuning riboswitch regulation through conformational selection. *J. Mol. Biol.* **405**, 926-38 (2011).
85. Eichhorn, C.D., Kang, M.J. & Feigon, J. Structure and function of $preQ_1$ riboswitches. *Biochim. Biophys. Acta-Gene Regul. Mech.* **1839**, 939-50 (2014).
86. Leontis, N.B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499-512 (2001).
87. Johnson-Buck, A. *et al.* Kinetic fingerprinting to identify and count single nucleic acids. *Nat. Biotechnol.* **33**, 730-2 (2015).
88. Jungmann, R. *et al.* Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami. *Nano Lett.* **10**, 4756-61 (2010).
89. Dupuis, N.F., Holmstrom, E.D. & Nesbitt, D.J. Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophys. J.* **105**, 756-66 (2013).

90. Eimer, W., Williamson, J.R., Boxer, S.G. & Pecora, R. Characterization of the overall and internal dynamics of short oligonucleotides by depolarized dynamic light scattering and NMR relaxation measurements. *Biochemistry* **29**, 799-811 (1990).
91. Ouldridge, T.E., Sulc, P., Romano, F., Doye, J.P. & Louis, A.A. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Res.* **41**, 8886-95 (2013).
92. Schwille, P., Oehlenschlager, F. & Walter, N.G. Quantitative hybridization kinetics of DNA probes to RNA in solution followed by diffusional fluorescence correlation analysis. *Biochemistry* **35**, 10182-93 (1996).
93. Cisse, I.I., Kim, H. & Ha, T. A rule of seven in Watson-Crick base-pairing of mismatched sequences. *Nat. Struct. Mol. Biol.* **19**, 623-7 (2012).
94. Snoussi, K. & Leroy, J.L. Imino proton exchange and base-pair kinetics in RNA duplexes. *Biochemistry* **40**, 8898-904 (2001).
95. O'Toole, A.S., Miller, S. & Serra, M.J. Stability of 3' double nucleotide overhangs that model the 3' ends of siRNA. *RNA* **11**, 512-6 (2005).
96. Greenfeld, M., Solomatin, S.V. & Herschlag, D. Removal of covalent heterogeneity reveals simple folding behavior for P4-P6 RNA. *J. Biol. Chem.* **286**, 19872-9 (2011).
97. Hyeon, C., Lee, J., Yoon, J., Hohng, S. & Thirumalai, D. Hidden complexity in the isomerization dynamics of Holliday junctions. *Nat. Chem. Biol.* **4**, 907-14 (2012).
98. Okumus, B., Wilson, T.J., Lilley, D.M. & Ha, T. Vesicle encapsulation studies reveal that single molecule ribozyme heterogeneities are intrinsic. *Biophys. J.* **87**, 2798-806 (2004).
99. Tan, E. *et al.* A four-way junction accelerates hairpin ribozyme folding via a discrete intermediate. *Proc. Natl. Acad. Sci. USA* **100**, 9308-13 (2003).
100. de Silva, C. & Walter, N.G. Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *RNA* **15**, 76-84 (2009).
101. Gourévitch, B. & Eggermont, J.J. A nonparametric approach for detection of bursts in spike trains. *J. Neurosci. Methods* **160**, 349-58 (2007).
102. Cai, L., Friedman, N. & Xie, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358-62 (2006).
103. Yu, J., Xiao, J., Ren, X.J., Lao, K.Q. & Xie, X.S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600-03 (2006).
104. Eden, U.T. & Kramer, M.A. Drawing inferences from Fano factor calculations. *J. Neurosci. Methods* **190**, 149-52 (2010).
105. Li, G.W., Oh, E. & Weissman, J.S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538-41 (2012).
106. Lai, D., Proctor, J.R. & Meyer, I.M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461-73 (2013).
107. Poot, R.A., Tsareva, N.V., Boni, I.V. & van Duin, J. RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10110-5 (1997).
108. 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).
109. Laalami, S., Zig, L. & Putzer, H. Initiation of mRNA decay in bacteria. *Cellular and molecular life sciences : CMLS* **71**, 1799-828 (2014).
110. Picard, F. *et al.* Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC Genomics* **13**, 528 (2012).

111. Golding, I., Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025-36 (2005).
112. So, L.H. *et al.* General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* **43**, 554-60 (2011).
113. Balazsi, G., van Oudenaarden, A. & Collins, J.J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910-25 (2011).
114. Beaumont, H.J., Gallie, J., Kost, C., Ferguson, G.C. & Rainey, P.B. Experimental evolution of bet hedging. *Nature* **462**, 90-3 (2009).
115. Hebenstreit, D., Deonarine, A., Babu, M.M. & Teichmann, S.A. Duel of the fates: the role of transcriptional circuits and noise in CD4+ cells. *Curr. Opin. Cell Biol.* **24**, 350-8 (2012).
116. Tuller, T. & Zur, H. Multiple roles of the coding sequence 5' end in gene expression regulation. *Nucleic Acids Res.* **43**, 13-28 (2015).
117. Todd, G.C. & Walter, N.G. Secondary structure of bacteriophage T4 gene *60* mRNA: implications for translational bypassing. *RNA* **19**, 685-700 (2013).
118. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507-16 (2008).
119. Rueda, D. *et al.* Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance. *Proc. Natl. Acad. Sci. USA* **101**, 10066-71 (2004).
120. Rasnik, I., McKinney, S.A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nat. Methods* **3**, 891-3 (2006).
121. Blanco, M. & Walter, N.G. Analysis of complex single-molecule FRET time trajectories. *Methods Enzymol.* **472**, 153-78 (2010).
122. Abelson, J. *et al.* Conformational dynamics of single pre-mRNA molecules during *in vitro* splicing. *Nat. Struct. Mol. Biol.* **17**, 504-12 (2010).
123. Shaw, J.J. & Green, R. Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol. Cell* **28**, 458-67 (2007).
124. Schagger, H. & Vonjagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100-KDa. *Anal. Biochem.* **166**, 368-79 (1987).
125. Komarova, A.V., Tchufistova, L.S., Supina, E.V. & Boni, I.V. Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA* **8**, 1137-47 (2002).
126. Salah, P. *et al.* Probing the relationship between Gram-negative and Gram-positive S1 proteins by sequence analysis. *Nucleic Acids Res.* **37**, 5578-88 (2009).
127. Giorginis, S. & Subramanian, A.R. The major ribosome binding site of *Escherichia coli* ribosomal protein S1 is located in its N-terminal segment. *J. Mol. Biol.* **141**, 393-408 (1980).
128. McGinness, K.E. & Sauer, R.T. Ribosomal protein S1 binds mRNA and tmRNA similarly but plays distinct roles in translation of these molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13454-9 (2004).
129. Okada, T., Wower, I.K., Wower, J., Zwieb, C.W. & Kimura, M. Contribution of the second OB fold of ribosomal protein S1 from *Escherichia coli* to the recognition of tmRNA. *Biosci., Biotechnol., Biochem.* **68**, 2319-25 (2004).

130. Kolb, A., Hermoso, J.M., Thomas, J.O. & Szer, W. Nucleic acid helix-unwinding properties of ribosomal protein S1 and the role of S1 in mRNA binding to ribosomes. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2379-83 (1977).
131. Thomas, J.O. & Szer, W. RNA-helix-destabilizing proteins. *Prog. Nucleic Acid Res. Mol. Biol.* **27**, 157-87 (1982).
132. Studer, S.M. & Joseph, S. Unfolding of mRNA secondary structure by the bacterial translation initiation complex. *Mol. Cell* **22**, 105-15 (2006).
133. Blose, J.M., Proctor, D.J., Veeraraghavan, N., Misra, V.K. & Bevilacqua, P.C. Contribution of the closing base pair to exceptional stability in RNA tetraloops: roles for molecular mimicry and electrostatic factors. *J. Am. Chem. Soc.* **131**, 8474-84 (2009).
134. Abdelkafi, M. *et al.* Common structural features of UUCG and UACG tetraloops in very short hairpins determined by UV absorption, Raman, IR and NMR spectroscopies. *J. Biomol. Struct. Dyn.* **14**, 579-93 (1997).
135. Mergny, J.L. & Lacroix, L. Analysis of thermal melting curves. *Oligonucleotides* **13**, 515-37 (2003).
136. McCann, J.J., Choi, U.B., Zheng, L., Weninger, K. & Bowen, M.E. Optimizing methods to recover absolute FRET efficiency from immobilized single molecules. *Biophys. J.* **99**, 961-70 (2010).
137. Blanchard, S.C., Kim, H.D., Gonzalez, R.L., Puglisi, J.D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12893-98 (2004).
138. Davoll, J. Pyrrolo[2,3-d]pyrimidines. *J. Chem. Soc.*, 131-38 (1960).
139. Sawano, A. & Miyawaki, A. Directed evolution of green fluorescent protein by a new versatile PCR strategy for site-directed and semi-random mutagenesis. *Nucleic Acids Res.* **28**, e78 (2000).
140. Kibbe, W.A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43-6 (2007).
141. He, B. *et al.* Rapid mutagenesis and purification of phage RNA polymerases. *Protein Expr. Purif.* **9**, 142-51 (1997).
142. Gurevich, V.V., Pokrovskaya, I.D., Obukhova, T.A. & Zozulya, S.A. Preparative *in vitro* mRNA synthesis using SP6 and T7 RNA polymerases. *Anal. Biochem.* **195**, 207-13 (1991).
143. Cunningham, P.R. & Ofengand, J. Use of inorganic pyrophosphatase to improve the yield of *in vitro* transcription reactions catalyzed by T7 RNA polymerase. *Biotechniques* **9**, 713-4 (1990).
144. Willkomm, D.K. & Hartmann, R.K. 3´-Terminal attachment of fluorescent dyes and biotin. in *Handbook of RNA Biochemistry* 86-94 (Wiley-VCH Verlag GmbH, 2008).
145. Kitagawa, M. *et al.* Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291-9 (2005).
146. Kapust, R.B. *et al.* Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993-1000 (2001).
147. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* **28**, 292 (2000).
148. Lancaster, L. & Noller, H.F. Involvement of 16S rRNA nucleotides G1338 and A1339 in discrimination of initiator tRNA. *Mol. Cell* **20**, 623-32 (2005).

149. Personal communication from Prof. Ruben L. Gonzalez, Jr., Columbia University, New York, NY.
150. Bevilacqua, P.C. & Cech, T.R. Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry* **35**, 9983-94 (1996).
151. Puglisi, J.D. & Tinoco Jr, I. Absorbance melting curves of RNA. *Methods Enzymol.* **180**, 304-25 (1989).
152. Owczarzy, R. Melting temperatures of nucleic acids: discrepancies in analysis. *Biophys. Chem.* **117**, 207-15 (2005).
153. Marky, L.A. & Breslauer, K.J. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**, 1601-20 (1987).
154. Joo, C. & Ha, T. Single-molecule FRET with total internal reflection microscopy. in *Single-molecule techniques: a laboratory manual* (eds. Selvin, P.R. & Ha, T.) (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2008).
155. Aitken, C.E., Marshall, R.A. & Pulglisi, J.D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.* **94**, 1826-35 (2008).
156. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809-19 (2007).
157. Peselis, A. & Serganov, A. Themes and variations in riboswitch structure and function. *Biochim. Biophys. Acta* **1839**, 908-18 (2014).
158. Liberman, J.A. & Wedekind, J.E. Riboswitch structure in the ligand-free state. *WIREs RNA* **3**, 369-84 (2012).
159. Savinov, A., Perez, C.F. & Block, S.M. Single-molecule studies of riboswitch folding. *Biochim. Biophys. Acta-Gene Regul. Mech.* **1839**, 1030-45 (2014).
160. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729-41 (2004).
161. Dambach, M. *et al.* The ubiquitous *yybP-ykoY* riboswitch is a manganese-responsive regulatory element. *Mol. Cell* **57**, 1099-109 (2015).
162. Marcano-Velazquez, J.G. & Batey, R.T. Structure-guided mutational analysis of gene regulation by the *Bacillus subtilis pbuE* adenine-responsive riboswitch in a cellular context. *J. Biol. Chem.* **290**, 4464-75 (2015).
163. Rieder, U., Kreutz, C. & Micura, R. Folding of a transcriptionally acting preQ$_1$ riboswitch. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10804-9 (2010).
164. Fuchs, R.T., Grundy, F.J. & Henkin, T.M. The S_{MK} box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* **13**, 226-33 (2006).
165. Fuchs, R.T., Grundy, F.J. & Henkin, T.M. S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the S_{MK} box translational riboswitch RNA. *Proc. Natl. Acad. Sci.* **104**, 4876-80 (2007).
166. Nou, X. & Kadner, R.J. Coupled changes in translation and transcription during cobalamin-dependent regulation of *btuB* expression in *Escherichia coli*. *J. Bacteriol.* **180**, 6719-28 (1998).
167. Nou, X. & Kadner, R.J. Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7190-5 (2000).
168. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043 (2002).

169. Lu, C. *et al.* Crystal structures of the SAM-III/S_{MK} riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat. Struct. Mol. Biol.* **15**, 1076-83 (2008).
170. Dalgarno, P.A. *et al.* Single-molecule chemical denaturation of riboswitches. *Nucleic Acids Res.* **41**, 4253-65 (2013).
171. Caban, K. & Gonzalez, R.L., Jr. The emerging role of rectified thermal fluctuations in initiator aa-tRNA- and start codon selection during translation initiation. *Biochimie* **114**, 30-8 (2015).
172. Aitken, C.E., Petrov, A. & Puglisi, J.D. Single ribosome dynamics and the mechanism of translation. *Annu. Rev. Biophys.* **39**, 491-513 (2010).
173. Blanchard, S.C. Single-molecule observations of ribosome function. *Curr. Opin. Struct. Biol.* **19**, 103-09 (2009).
174. Korostelev, A., Ermolenko, D.N. & Noller, H.F. Structural dynamics of the ribosome. *Curr. Opin. Chem. Biol.* **12**, 674-83 (2008).
175. Uemura, S. *et al.* Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012-7 (2010).
176. Uemura, S. *et al.* Peptide bond formation destabilizes Shine-Dalgarno interaction on the ribosome. *Nature* **446**, 454-57 (2007).
177. Masuda, T. *et al.* Initiation factor 2, tRNA, and 50S subunits cooperatively stabilize mRNAs on the ribosome during initiation. *Proc. Natl. Acad. Sci.* **109**, 4881-85 (2012).
178. Wang, J., Caban, K. & Gonzalez, R.L., Jr. Ribosomal initiation complex-driven changes in the stability and dynamics of initiation factor 2 regulate the fidelity of translation initiation. *J. Mol. Biol.* **427**, 1819-34 (2015).
179. MacDougall, D.D. & Gonzalez, R.L., Jr. Translation initiation factor 3 regulates switching between different modes of ribosomal subunit joining. *J. Mol. Biol.* **427**, 1801-18 (2015).
180. Tsai, A. *et al.* Heterogeneous pathways and timing of factor departure during translation initiation. *Nature* **487**, 390-3 (2012).
181. Milón, P., Konevega, A.L., Gualerzi, C.O. & Rodnina, M.V. Kinetic checkpoint at a late step in translation initiation. *Mol. Cell* **30**, 712-20 (2008).
182. Milón, P. *et al.* Transient kinetics, fluorescence, and fret in studies of initiation of translation in bacteria. *Methods Enzymol.* **430**, 1-30 (2007).
183. Bao, Q. *et al.* A complete sequence of the *T. tengcongensis* genome. *Genome Res.* **12**, 689-700 (2002).
184. Nawrocki, E.P. *et al.* Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43**, D130-7 (2015).
185. Burge, S.W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, D226-D32 (2013).
186. Passalacqua, K.D. *et al.* Structure and complexity of a bacterial transcriptome. *J. Bacteriol.* **191**, 3203-11 (2009).
187. Suryanarayana, T. & Subramanian, A.R. An essential function of ribosomal protein S1 in mRNA translation. *Biochemistry* **22**, 2715-19 (1983).
188. Isono, K. & Isono, S. Lack of ribosomal protein S1 in *Bacillus stearothermophilus*. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 767-70 (1976).
189. Farwell, M.A. & Rabinowitz, J.C. Protein synthesis *in vitro* by *Micrococcus luteus*. *J. Bacteriol.* **173**, 3514-22 (1991).
190. Roberts, M.W. & Rabinowitz, J.C. The effect of *Escherichia coli* ribosomal protein S1 on the translational specificity of bacterial ribosomes. *J. Biol. Chem.* **264**, 2228-35 (1989).

191. Sorokin, A., Serror, P., Pujic, P., Azevedo, V. & Ehrlich, S.D. The *Bacillus subtilis* chromosome region encoding homologues of the *Escherichia coli mssA* and *rpsA* gene products. *Microbiology* **141**, 311-9 (1995).
192. Lauber, M.A., Running, W.E. & Reilly, J.P. *B. subtilis* ribosomal proteins: structural homology and post-translational modifications. *J. Proteome Res.* **8**, 4193-206 (2009).
193. Ohashi, Y. *et al.* Expression profiling of translation-associated genes in sporulating *Bacillus subtilis* and consequence of sporulation by gene inactivation. *Biosci. Biotechnol. Biochem.* **67**, 2245-53 (2003).
194. Milón, P., Maracci, C., Filonava, L., Gualerzi, C.O. & Rodnina, M.V. Real-time assembly landscape of bacterial 30S translation initiation complex. *Nat. Struct. Mol. Biol.* **19**, 609-15 (2012).
195. Saito, K. & Nomura, M. Post-transcriptional regulation of the *str* operon in *Escherichia coli*. Structural and mutational analysis of the target site for translational repressor S7. *J. Mol. Biol.* **235**, 125-39 (1994).
196. Dorywalska, M. *et al.* Site-specific labeling of the ribosome for single-molecule spectroscopy. *Nucleic Acids Res.* **33**, 182-89 (2005).
197. Culver, G.M. Meanderings of the mRNA through the ribosome. *Structure* **9**, 751-58 (2001).
198. Hickerson, R., Majumdar, Z.K., Baucom, A., Clegg, R.M. & Noller, H.F. Measurement of internal movements within the 30 S ribosomal subunit using Förster resonance energy transfer. *J. Mol. Biol.* **354**, 459-72 (2005).
199. Majumdar, Z.K., Hickerson, R., Noller, H.F. & Clegg, R.M. Measurements of internal distance changes of the 30 S ribosome using FRET with multiple donor-acceptor pairs: Quantitative spectroscopic methods. *J. Mol. Biol.* **351**, 1123-45 (2005).
200. Ermolenko, D.N. *et al.* Observation of intersubunit movement of the ribosome in solution using FRET. *J. Mol. Biol.* **370**, 530-40 (2007).
201. Lieberman, K.R. *et al.* The 23 S rRNA environment of ribosomal protein L9 in the 50 S ribosomal subunit. *J. Mol. Biol.* **297**, 1129-43 (2000).
202. Kaltschmidt, E. Ribosomal proteins. XIV: Isoelectric points of ribosomal proteins of *E. coli* as determined by two-dimensional polyacrylamide gel electrophoresis. *Anal. Biochem.* **43**, 25-31 (1971).
203. Julián, P. *et al.* The cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biol.* **9**, e1001095 (2011).
204. Simonetti, A. *et al.* A structural view of translation initiation in bacteria. *Cell. Mol. Life Sci.* **66**, 423-36 (2009).
205. Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-U73 (2008).
206. Eichhorn, C.D. Doctoral Thesis, University of Michigan, Horace H. Rackham School of Graduate Studies (2012).
207. Xu, D. & Côté, J.-C. Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3´ end 16S rDNA and 5´ end 16S – 23S ITS nucleotide sequences. *Int. J. Syst. Evol. Microbiol.* **53**, 695-704 (2003).
208. Britton, R.A. *et al.* Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol. Microbiol.* **63**, 127-38 (2007).
209. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006 0008 (2006).

210. Spedding, G. *Ribosomes and protein synthesis: a practical approach*, xxiv, 318 p. (Oxford University Press, Oxford ; New York, 1990).
211. Dahlberg, A.E., Dingman, C.W. & Peacock, A.C. Electrophoretic characterization of bacterial polyribosomes in agarose-acrylamide composite gels. *J. Mol. Biol.* **41**, 139-47 (1969).
212. Personal communication from Prof. Rachel Green's lab, John Hopkins University, Baltimore, MD.
213. Dubnoff, J.S. & Maitra, U. Isolation and properties of protein factors involved in polypeptide chain initiation in *Escherichia coli*. *Methods Enzymol.* **20**, 248-61 (1971).
214. Traub, P., Mizushima, S., Lowry, C.V. & Nomura, M. Reconstitution of ribosomes from subribosomal components. in *RNA and Protein Synthesis* (ed. Moldave, K.) p. 535 (Academic Press, New York, 1981).
215. Walker, S.E. & Fredrick, K. Preparation and evaluation of acylated tRNAs. *Methods* **44**, 81-86 (2008).
216. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat Biotech* **19**, 751-55 (2001).
217. Studer, S.M. & Joseph, S. Binding of mRNA to the bacterial translation initiation complex. *Methods Enzymol.* **430**, 31-44 (2007).
218. Wong, I. & Lohman, T.M. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5428-32 (1993).
219. Rio, D.C. Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harbor Protocols* **2012**, 1078-81 (2012).
220. Zaher, H.S., Shaw, J.J., Strobel, S.A. & Green, R. The 2'-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *EMBO J.* **30**, 2445-53 (2011).
221. Jelenc, P.C. & Kurland, C.G. Nucleoside triphosphate regeneration decreases the frequency of translation errors. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3174-8 (1979).
222. Jelenc, P.C. Rapid purification of highly active ribosomes from *Escherichia coli*. *Anal. Biochem.* **105**, 369-74 (1980).
223. Fei, J. *et al.* A highly purified, fluorescently labeled *in vitro* translation system for single-molecule studies of protein synthesis. *Methods Enzymol.* **472**, 221-59 (2010).
224. Culver, G.M. & Noller, H.F. Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5**, 832-43 (1999).
225. De Lay, N., Schu, D.J. & Gottesman, S. Bacterial small RNA-based negative regulation: Hfq and its accomplices. *J. Biol. Chem.* **288**, 7996-8003 (2013).
226. Al-Hashimi, H.M. & Walter, N.G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321-9 (2008).
227. Kang, M., Eichhorn, C.D. & Feigon, J. Structural determinants for ligand capture by a class II preQ$_1$ riboswitch. *Proc. Natl. Acad. Sci.* **111**, E663-E71 (2014).
228. 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).
229. Chen, B., Zuo, X., Wang, Y.X. & Dayie, T.K. Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy. *Nucleic Acids Res.* **40**, 3117-30 (2012).

230. Harris, D.A. & Walter, N.G. Terbium(III) footprinting as a probe of RNA structure and metal-binding sites. in *Handbook of RNA Biochemistry* 205-13 (Wiley-VCH Verlag GmbH, 2008).
231. Wang, J. *et al.* An analysis of the proteomic profile for *Thermoanaerobacter tengcongensis* under optimal culture conditions. *Proteomics* **4**, 136-50 (2004).
232. Wang, J. *et al.* The proteomic alterations of *Thermoanaerobacter tengcongensis* cultured at different temperatures. *Proteomics* **7**, 1409-19 (2007).
233. Wang, C. *et al.* Transcriptomic analysis of *Thermoanaerobacter tengcongensis* grown at different temperatures by RNA sequencing. *J. Genet. Genomics* **42**, 335-8 (2015).
234. Uemura, S. *et al.* Single-molecule imaging of full protein synthesis by immobilized ribosomes. *Nucleic Acids Res.* **36**, e70 (2008).