Methods 162–163 (2019) 3–11

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Probing RNA structure and interaction dynamics at the single molecule level

Adrien Chauvier[1], Javier Cabello-Villegas[1], Nils G. Walter*

Single Molecule Analysis Group, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Keywords:
Single-molecule fluorescence microscopy
Conformational change
Kinetic analysis
Riboswitch
RNA structure

ABSTRACT

RNA structures and their dynamic fluctuations lie at the heart of understanding key biological process such as transcription, splicing, translation and RNA decay. While conventional bulk assays have proven to identify and characterize key pathway intermediates, the generally dynamic nature of RNA structures renders the information obtained from time and ensemble averaging techniques necessarily lacking in critical details. Here we detail Single-Molecule Kinetic Analysis of RNA Transient Structure (SiM-KARTS), a method that readily monitors structural fluctuations of single RNA molecules through the repetitive interaction of fluorescent probes with an unlabeled, surface-immobilized RNA target of virtually any length and in any biological context. In addition, we demonstrate the broad applicability of SiM-KARTS by kinetically fingerprinting the binding of cognate tRNA ligand to single immobilized T-box riboswitch molecules. SiM-KARTS represents a valuable tool for probing biologically relevant structure and interaction features of potentially many diverse RNA metabolic pathways.

1. Introduction

1.1. RNAs are structurally polymorphic in service of their function

In all living organisms, a plethora of recently discovered non-coding RNAs plays a major role in modulating the expression of protein-coding messenger (m)RNAs by regulating the processes involved in RNA metabolism from transcription to processing, translation and degradation [1]. As versatile information molecules that can both recognize specific sequences by complementary binding and fold into complex structures with advanced functions such as catalysis, RNAs often adopt distinct, yet dynamic conformations and three-dimensional architectures in service of their diverse functions [2,3]. The relationships between particular RNA structures and their functions are rapidly being unraveled for increasing numbers of RNA molecules, including those implicated in disease, such as transfer RNAs (tRNAs), long non-coding RNAs (lncRNAs) and RNAs involved in translational control [4,5].

The critical nature and practical relevance of this structural polymorphy of RNA is exemplified by riboswitches, structural elements embedded near the 5′ends of bacterial mRNAs that regulate the mRNA's transcription and/or protein synthesis upon binding of a specific metabolite or ion ligand. Riboswitches typically undergo a structural rearrangement in response to ligand binding that modulates either the relative likelihood of forming a terminator stem that helps dislodge RNA polymerase (RNAP) during transcription or the accessibility of the Shine-Dalgarno (SD) sequence that helps recruit the ribosome for mRNA translation. Thus, riboswitches are intriguing model systems for establishing the link between RNA structure, dynamics and function.

To regulate gene expression tightly, riboswitches often rely on kinetic partitioning, where the local binding of a small ligand shifts the interdependent conformational dynamics in a way that affects a global conformational change, following a dynamic, cascading structural linchpin mechanism [6]. Thus, quantitative tools for probing the time dependence of both local and global RNA structure become essential for establishing how RNAs exert their profound functions in gene expression.

1.2. Advantages of RNA analysis at the single molecule level

For dynamic structural motifs such as riboswitches, single-molecule (SM) techniques are often essential for a quantitative analysis of key structures involved in the modulation of gene expression. For example, the kinetic and structural differences of riboswitch folding in the absence and presence of ligand relative to the kinetics and architectural features of the cellular gene expression machinery are key determinants of whether the downstream gene(s) are expressed. Consequently, their ability to uniquely detect and quantify rare molecular events, transiently visited states and diverse kinetic behaviors have led to a rapid adoption of SM methods for probing riboswitch mechanisms.

As one example, single molecule fluorescence (or Förster) resonance

* Corresponding author.
E-mail address: nwalter@umich.edu (N.G. Walter).
[1] A.C. and J.C-V. contributed equally to this work.

https://doi.org/10.1016/j.ymeth.2019.04.002
Received 27 January 2019; Received in revised form 28 March 2019; Accepted 1 April 2019
Available online 03 April 2019

energy transfer (smFRET) allows for kinetic and conformational analyses of RNA structures by employing a pair of fluorophores strategically placed into the RNA target sequence to monitor specific intra- or intermolecular distances over time [7,8]. In other types of experiments such as those utilizing optical tweezers, the RNA of interest is subjected to a particular force vector that induces unfolding and reveals aspects of its folding free energy landscape [9].

Alternatively, RNA conformations can be monitored by probing the RNA molecule with short complementary DNA oligonucleotides. Past examples of this strategy include measurements of the transition kinetics between alternate conformations of an RNA by monitoring absorption [10] or fluorescence changes [11] upon stopped-flow mixing, or through vulnerability of the resulting RNA:DNA hybrid to RNase H degradation [12]. Additionally, stable hybridization to DNA oligonucleotide microarrays has been used to probe larger RNA structures [13]. In these methods, non-equilibrium measurements can be used to assess the sum of RNA folding and unfolding rate constants. More recently, RNase H sensitivity after binding of DNA oligonucleotides has been employed to probe the conformational changes of riboswitches in response to their cognate ligands [14–16]. As ensemble-averaging bulk approaches, these assays cannot provide detailed kinetic information on the dynamic changes in local RNA structure, which obscures the conformational rearrangements at the heart of riboswitch function. In addition, the readout by RNase H cleavage destroys the RNA target and thus does not lend itself to studying the long-term equilibrium dynamics between alternate conformations.

Here, we describe a quantitative method for probing the structural dynamics at the SM level of virtually any RNA without the need for any direct labeling, which can be applied under practically any buffer and biological matrix conditions, and over any timescale. This method, termed SiM-KARTS (for Single Molecule Kinetic Analysis of RNA Transient Structure), interrogates local RNA structure formation based on its accessibility toward the site-specific transient binding of fluorescent probes (FPs) [17], which is read out by total internal reflection fluorescence (TIRF) microscopy observation of the immobilized target (Fig. 1). The FPs bear the fluorescent readout tag and are easily designed to have a melting temperature close to the interrogation temperature and thus permit the repeated binding to and dissociation from their target RNA. Unlike using an RNase H-based readout, reversible transitions between distinct RNA conformations can be monitored at the SM level over long periods of time using the near-inexhaustible FP reservoir in solution. Since this vast pool of fluorophore labeled probe molecules resides outside of the field of illumination, photobleaching does not alter notably the effective FP concentration, allowing experiments to be run for hours if desired for observing slow conformational changes. In addition, the responses of each single RNA molecule to multiple buffer exchanges can be recorded. The kinetic pattern, or temporal fingerprint, of the binding events obtained by SiM-KARTS allows one to decipher multiple states adopted by an RNA over time, and reveal heterogeneous (“non-ergodic”) behaviors between molecules. A concrete example of SiM-KARTS for surveying ligand-dependent riboswitch dynamics is given (Fig. 1), but the technology can be applied to almost any type of RNA from bacteria to eukaryotes and be adapted to study other biological processes, including those involving protein and/or RNA binding, for which an example is also given.

2. Material and methods

2.1. RNA target synthesis

SiM-KARTS monitors the binding and dissociation of oligonucleotides to and from RNA molecules that are, in one typical modality of the experiment, attached to a quartz microscope slide. In contrast to smFRET experiments, which require the site-specific, proximal, (near-) covalent incorporation of two fluorophores into the target RNA, the target to be studied by SiM-KARTS can be easily generated at low cost using, for example, *in vitro* transcription by T7 RNAP, thus enabling the analysis of very long RNA sequences that would be difficult to chemically synthesize for fluorophore incorporation due to low yield and high cost. Two modifications of the RNA under investigation may still be necessary. First, immobilization of the target RNA on the quartz slide is achieved by hybridization to a thermodynamically stably associated capture probe (CP) that itself is (near-)covalently coupled to the slide surface, typically by an incorporated biotin moiety that binds the streptavidin coated microscope slide. In a long RNA, a position distal to the RNA sequence to be probed by the FP can be chosen easily within the native sequence, or a relatively unstructured anti-capture sequence can be appended; or one of the RNA termini can be chemically or enzymatically coupled to a biotin moiety [18]. In addition, while not strictly required it is often desirable to mark the exact position of each target RNA on the slide by further specifying or appending a unique unstructured RNA sequence that can stably bind a fluorescent DNA localization probe (LP, typically labeled with a distinguishable color).

2.2. Probe design

2.2.1. Capture probe (CP) and localization probe (LP)

For specific immobilization of the RNA target onto the streptavidin-coated microscope slide, we typically use a CP that is terminally labeled with a biotin moiety. This approach avoids the need to directly attach the biotin to the RNA molecule, while retaining binding efficiency to the slide. The CP is specifically designed to hybridize to the target RNA and allow the capture of the complex onto the microscope slide via the biotin moiety. Usually, we employ a DNA-based CP, but in principle it can include any type of nucleic acid modification such as RNA, peptide or locked nucleic acids (LNA), depending on the system under study and its specific requirements. A second probe, the LP labeled with a Cy3 (green) fluorophore instead of biotin to mark the target RNA position, is designed in parallel using similar criteria. In the present method, we describe the use of two separate probes to attach to and localize the RNA molecule for enhanced specificity, however, it is possible to design a doubly-modified DNA oligonucleotide with both biotin and Cy3 that would ensure both criteria.

For a thermodynamically stable interaction between the CP/LP and the RNA target, a high predicted melting temperature (Tm) is preferable. The design is made semi-empirically using basic online Tm and self-structure prediction tools such as Oligoanalyzer from Integrated DNA Technologies Inc. (IDT). The main goal is to obtain a CP/LP design with a predicted Tm (under default conditions) of $> 60\,^{\circ}C$ and a very low self-structure score (below the temperature in which the experiment will be performed). Ideally, the CP and LP will bind to opposite ends of the target RNA to only produce surface-immobilized Cy3 labels from fully intact RNA targets. The CP and LP are then conveniently ordered from a commercial supplier such as IDT with terminal biotin and Cy3 labels, respectively, and HPLC purification.

When studying RNA structures it is important to aim to not alter the RNA’s architecture. In fact, hybridizing a DNA segment to the RNA sequence for labeling could affect its natural folding and function. To avoid this challenge, it may be advisable to append arbitrary hybridization sequences to the 5′ and/or 3′ends of the target RNA for attachment of the CP and LP, chosen by computational folding predictions to not interfere with folding of the native RNA. These hybridization sequences should be specifically optimized to minimize self-structures that would prevent CP/LP binding as well as complementarities with the target RNA. Such design optimization can be easily performed using online folding prediction tools such as Mfold [19] or RNAStructure [20] and the efficiency of binding be confirmed using non-denaturing gel electrophoresis.

2.2.2. Fluorescent probe (FP)

To enable the observation of dynamic RNA structures, readily reversible binding and dissociation of the FP on the same RNA molecule is

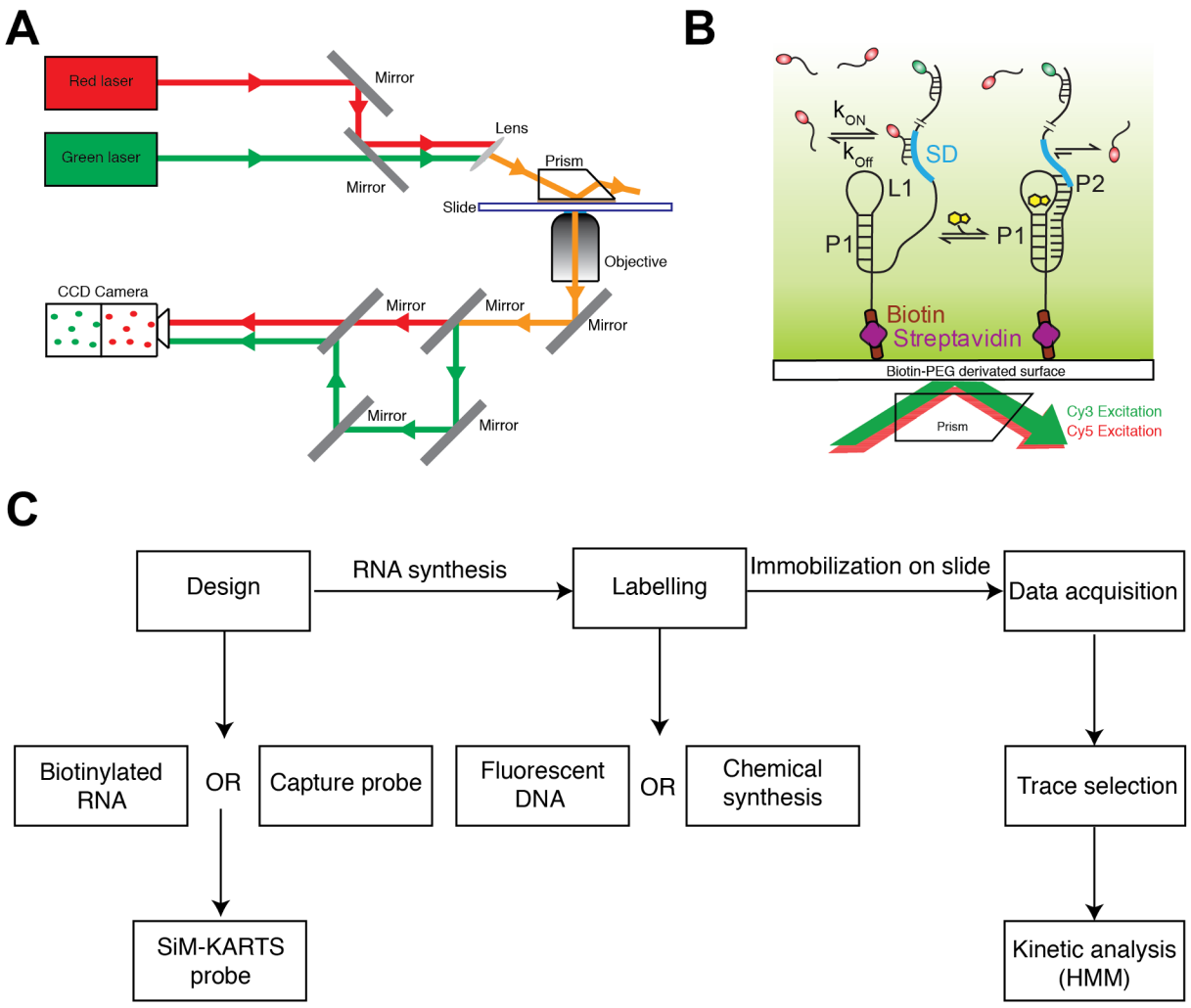

Fig. 1. Overview of the SiM-KARTS method. (A) Prism-type TIRF microscope. (B) Schematic illustrating the experimental principles of SiM-KARTS with the example of the accessibility of the Shine-Dalgarno (SD) sequence of the preQ$_1$ riboswitch described in detail in Rinaldi et al. 2016 [17]. The mRNA harboring the riboswitch is labeled through hybridization with a Cy3 DNA oligonucleotide towards the 3′ end, and SD sequence accessibility is surveyed with a Cy5-labeled anti-SD RNA probe. (C) General experimental design pipeline used for SiM-KARTS.

desired to occur for many cycle throughout the observation window. Typically, this means that the Tm between the target and the FP has to be close to the temperature at which the experiment will be performed (usually at room temperature, or 20–25 °C). In addition, the lifetime of the bound state has to be at least 2–3 times longer than the camera exposure time (100 ms in our case), yet not too long so as to limited the number of binding/dissociation cycles per RNA molecule over a practical window of observation (typically tens of minutes). For this consideration, the length of the FP is a major optimization parameter. Given that the dissociation rate constant of a DNA oligonucleotide under standard temperature and ionic strength (especially for shorter oligonucleotides) is exponentially dependent on its length, the optimal length of the FP probe with a 50% GC content, for example, will be around 7–8 nucleotides complementary to the target. To assess RNA conformational changes at the timescale at which they occur, the binding/unbinding of the FP should be faster than the RNA conformational change dynamic. To this end, a fast association/dissociation rate constant of the FP might be necessary and can be controlled by shortening the length of the FP or introducing non-complementary mutations to prevent too stable binding of the probe. Additionally, to further modulate the binding and dissociation times of the FP to and from its target, the buffer salt concentration can be modulated. For example, we often use a high monovalent salt concentration (e.g., 600 mM KCl in the present method) to promote binding, but other ion concentrations such as magnesium can equally be adjusted as needed to study the RNA conformational dynamics of interest. Alternatively, the design of the FP could take into account the lower salt concentration found *in vivo* (around 300 mM in *E. coli*) to be able to survey RNA conformational dynamics under near-physiological conditions. As for the CP and LP, the FP is ordered HPLC-purified from a commercial supplier, but instead with a typically terminal Cy5 (red) fluorophore attached.

In the case of a large and/or highly structured RNA being probed, it can be useful to first test and optimize the binding of the FP and LP to a shorter construct, containing just the segment to which the probes are meant to bind. Such a construct can be designed similarly to the full-length sequence, but just containing the binding regions for the biotinylated CP and Cy3-labeled LP flanking the sequence being targeted by the Cy5-labeled FP. The ionic conditions, particularly the magnesium concentration, can then easily be varied for optimal association and dissociation of the FP probe to the simplified target construct, thus providing an approximate calibration for the expected efficiency of FP binding to its full-length RNA target. The optimal salt concentration, including particularly Mg^{2+}, should first be evaluated theoretically for its effect on the T_m, using calculators provided by vendors such as IDT, since an increase in Mg^{2+} concentration by 10 mM typically raises the T_m of FPs by around 10 °C. Given that the dissociation rate constant does not correlate with the T_m [34] since oligonucleotides will have

kinetics dependent on other variables, such as pH and toehold interactions that easily "zip up to form" a longer double helix, any prediction may not be accurate, and the final optimization has to occur experimentally. A screening of Mg^{2+} and K^{+} in concentrations ranging from 1 to 50 and 50–600 mM, respectively, may be informative to test whether a specific FP's dynamics are adequate.

If an increase or decrease in the dissociation rate constant is desired, and changing the ionic strength does not provide the desired effect or is not an option for other reasons, instead the oligonucleotide length can be changed, as both the ON and OFF rate constants are known to increase with a shortening of oligonucleotide length [34]. For example, the optimization of oligonucleotides designed to bind an immobilized miR-141 RNA with target sequence GGUAAAGAUCG showed that two 10 residue FPs with sequence /5Cy5/CATCTTTACC and /5Cy5/CCAT CTTTAC exhibited fewer than 10 binding/dissociation events over 10 min. Decreasing the FP size to the 9 residues /5Cy5/CCATCTTTA, yielded an average of 42 such events, with bound and unbound life times in the order of 7.5 and 12 s, respectively, which was a desirable solution for that system [34]. To generally assess specificity of binding of the designed FP, control experiments are advisable wherein the FP targetsite is blocked by hybridizing a long antisense DNA oligonucleotide across the site that is expected to significantly reduce the FP binding frequency [17].

2.3. Slide preparation

Single molecules are imaged under the TIRF microscope by attaching them to quartz microscope slides. The slides are prepared according to established techniques [21,22]. Briefly, quartz slides are coated with polyethylene glycol (PEG) to prevent non-specific binding, mixed with biotin-PEG, where biotin is used as a tight ligand for tetrameric streptavidin, which in turn can bind additional biotin labeled molecules, in our case the CP, thus affixing it to the slide. Quartz microscope slides can be reused but need to be thoroughly cleaned by scrubbing with detergent, followed by sonication, rubbing with methanol and sonication in 1 M potassium hydroxide (KOH), flaming under a propane torch, and finally incubation in a solution of 14.3% v/v ammonium hydroxide (NH_4OH, 28–30 wt. %) and 14.3% v/v hydrogen peroxide (H_2O_2, 30–35 wt. %), activated as a dilute alkaline piranha (or "RCA standard clean 1" [23]) solution by heating to 60–70 °C. In order to bind the PEG and biotin-PEG to the slide, derivatives containing N-hydroxysuccinimide (NHS) ester, namely methoxy-PEG-succinimidyl valerate and biotin-PEG-succinimidyl valerate (m-PEG and biotin-PEG, respectively, from Laysan Bio, Inc.) are attached to amino groups placed on the slide. These amino groups are installed on the slide by incubation in the presence of 2% (v/v) 3-aminopropyl triethoxysilane (APTES) in acetone for 20 min at room temperature. Once the PEG reaction is complete (2 h to overnight incubation in 0.1 M sodium bicarbonate), any remaining free silane-amines are scavenged with disulfo-succinimidyl tartrate (DST) (43 mM in 1 M sodium bicarbonate) for 30 min. The slides then are prepared as described previously [21,22], making a channel with a coverslip through which solutions can be flowed in. The prepared slides can be stored at room temperature in a dark and dry environment for two weeks.

2.4. Sample preparation and imaging

All data shown in the following were collected using a prism-type TIRF microscope based on an Olympus IX-71 frame equipped with a 60× water-immersion objective (Olympus Uplanapo, 1.2NA) with an intensified CCD camera (I-Pentamax, Princeton Instruments) or a more recent Hamamatsu C13440-20CU scientific CMOS camera. Cy5 excitation was provided by a 640-nm red laser (Coherent CUBE 640-100C, 100mW) and Cy3 excitation by a 532-nm green laser (CrystaLaser CL532-150mW-L) (Fig. 1A). Typical laser powers used were 25 mW for 532 nm and 40 mW for 640 nM.

First, slides were flowed with the SiM-KARTS buffer of choice to check for the integrity of the sealing. Next, for streptavidin-coated slides we flow a solution containing 0.2 mg/mL of streptavidin in SiM-KARTS buffer onto the slide, then wait for 5 min to allow the streptavidin to bind to the biotinylated surface. Excess streptavidin is then washed away by flowing again the same buffer without streptavidin.

When using a CP, incubation of a 1:1 ratio of CP and target RNA is performed prior to flowing the complex onto the slide. To this end, CP and RNA can be heat-annealed at 70 °C in SiM-KARTS buffer and allowed to cool to room temperature over 20 min, although lower temperature annealing protocol can also be employed if denaturation of the target RNA is to be avoided. Finally the CP-RNA mixture is diluted to the desired concentration (between 50 and 100 pM) in SiM-KARTS buffer containing an additional 10–12-fold excess of CP to ensure that the complex would stay intact during dilution. The complex is then flowed onto the slide and incubated for 5 min to allow the capture interaction between biotin and streptavidin to occur. Alternatively, the target RNA can be directly captured onto a slide coated with biotinylated CP.

To reduce photobleaching and increase the longevity of the fluorescence signal, the FP is prepared at the desired concentration (typically 20–100 nM) in SiM-KARTS buffer containing 5 mM protocatechuic acid (PCA), 50 nM protocatechuate-3,4-dioxygenase (PCD), and 2 mM Trolox (oxygen scavenger system, OSS) and flowed onto the slide with RNA molecules already bound on it. Alternatively, an OSS composed of 44 mM glucose, 165 U/mL glucose oxidase from *Aspergillus niger*, 2170 U/mL catalase from *Corynebacterium glutamicum*, and 5 mM Trolox can be used [24]. Of note, commercial enzyme preparations may be contaminated with RNases and should be tested for the same by incubating with RNA and subsequent gel electrophoretic analysis of RNA integrity. If evidence for RNase activity is found, switching enzyme sources typically solves the problem [25]. Usually around 5 min is necessary to allow the OSS to sufficiently deplete oxygen on the slide. The imaging buffer further should contain the desired ion concentrations (monovalent and magnesium ions) to perform the SiM-KARTS experiment. Data acquisition is performed with both lasers running (to excite both Cy3 and Cy5) (Fig. 1A, B) in a darkened room at an environmentally controlled temperature of 20 ± 3 °C.

The intensity of the laser detecting the FP in particular has to be optimized to achieve a signal-to-noise ratio sufficient to facilitate the analysis process (see Section 2.5). As the FP is present at relatively high concentration in the microfluidic channel, significant background signal is expected due to the free diffusion of excess FP into the evanescent illumination field generated by TIRF (~100 nm deep into the solution). To not overwhelm the signal, both FP concentration and laser power need to be optimized in parallel to clearly distinguish FP binding events from free FP background fluorescence. Under our conditions, the upper limit of the FP concentration to be flowed onto the slide is around 50 nM; however, depending of the imaging system used and camera sensitivity, this upper limit can be changed and needs to be determined empirically to find the optimal signal-to-noise ratio that permits an easy analysis process (see Section 2.5).

In cases where a ligand other than a designed oligonucleotide FP is used and the ligand bound times are on the order of tens of seconds, it is advisable to determine the photobleaching rate of the labeling fluorophore under the imaging conditions used by immobilizing it directly onto the slide. Once this rate constant is known, it can be used to correct the observed dissociation rate constant as described [26]. Complementarily, if photobleaching is limiting the measurements of slow dissociation rate constants, the TIRF illumination intensity can be decreased to decrease photobleaching, for example, using polarizable or neutral-density filters. The resulting lower fluorescence intensity can be compensated by extending the camera integration time, thus increasing signal intensity at the expense of the time resolution achieved. For very long ligand residence times (minutes or even hours), illumination can be performed in on/off cycles, e.g., 200 ms on, 19.8 s off, either

synchronizing the camera with the duty cycle or discarding the dark periods after acquisition.

2.5. Data analysis

2.5.1. Trace selection

Each movie consists of the field of view separated into the two emission colors that are detected side-by-side onto two separate regions (halves) of the camera chip (Fig. 1A). To find the corresponding spots for a same location on the slide, a correlation map for the microscope setup is calibrated using a fluorescent beads containing slide that is visible in both channels. This slide contains ~70 µl of a 1:2500 dilution of FluoSpheres (Invitrogen 0.2 µm, yellow-green fluorescent (505/515)) sealed under a coverslip, which is excited with a green laser at less than 1 mW power for acquisition of a 10 s movie at 10 frames per second. Matlab scripts are used to map the corresponding positions of the beads, separately viewed at the two emission colors of the experiment. This map is subsequently applied to the data containing movies to identify spots exhibiting Cy3 intensity during the first 10 frames recorded, followed by extraction of signal from the same slide location in the Cy5 channel. Typically we set our script to search for intensities above an empirically defined threshold in both channels, within a 9-by-9 pixel window with a center on the Cy3 peak intensity. The total intensities within those windows for each channel are plotted together versus time to generate individual molecule traces from each movie.

After this initial trace generation, the molecules are selected based on several required criteria (Fig. 2):

1) The presence of a single Cy3 dye as verified by the observation of a single photobleaching step of the fluorophore corresponding to the detection of a single labeled RNA molecule.
2) At least two Cy5 co-localization signals per trajectory corresponding to FP binding events with a signal-to-noise ratio of at least 3:1.

Based on these parameters, traces are selected for further kinetic analysis using the Hidden Markov modeling (HMM) described in the next section. The number of traces obtained from the initial analysis will depend on the sensitivity threshold defined for the trace selection. This threshold can be used as an initial screening to discard traces that have a low signal-to-noise ratio. After this initial filter, the user should further screen the traces that displayed both Cy3 and Cy5 fluorescence. Depending on the threshold set, the user may find that many traces do not meet the above criteria. Those traces are not further analyzed as they may originate, for instance, from uneven illumination of the slide, more than one RNA target molecule in the same area, or non-specific binding of FP to the slide. For the traces selected, the fluorescence intensities of the RNA target and FP are extracted, and the FP intensities are analyzed as follows.

2.5.2. Hidden Markov modeling and rate determination

HMM is a statistical algorithm that has been used for various applications such as sequence alignment [27] and smFRET analysis [7]. A HMM has three main parameter sets—the probability matrices of transition, emission, and initiation—that are optimized through an iterative process to find those parameters that best describe the data.

Prior to idealization using a two-state HMM, SiM-KARTS time traces are 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 of the whole dataset.

Next, all selected traces from a single experimental condition are stitched together using a custom Matlab script allowing to separate the Cy3 and Cy5 signals in two files formatted for QuB [28]. The QuB input format is a *.txt file with data columns for time and fluorescent signal. To perform HMM analysis, it is necessary that the traces are formatted in a manner amenable for manipulation with the desired analysis program. It should be noted that we are presenting only an analysis method based on QuB, but other programs are freely available such as HaMMy [29] and vb-FRET [30].

After importing the data into QuB, we create an initial HMM in the modeling window with the two possible states (unbound and bound to the FP probe) and arbitrarily defined rates between them (ideally on the order of the experimentally observed rates). The amplitudes and standard deviation of the two states are then estimated by using the 'Amps' function. This will initiate the model, and then the 'Idealize' function is used to build the idealization trace. After this procedure, it is necessary to check for the accuracy of the idealization of each trace and – as needed – manually refine the idealization result toward accurate fitting of the actual data.

While SiM-KARTS data have the advantage of yielding simple two-state (on/off) signals that often can be readily idealized, depending on the illumination conditions sometimes the signal-to-noise ratio may be less than ideal. In the QuB software, it is then possible to set different parameters that allow denoising to better distinguish between noise and signal. In particular, a low band-pass filter or, alternatively, Gaussian smoothing can be applied to suppress high frequency (one frame duration or shorter) intensity spikes that may be noise (or are below the time resolution of the experiment). For example, in experiments where the exposure time is 10 frames per second, during preprocessing we sometimes apply a 1 Hz QuB deconvolution "Filter" to suppress very short spikes.

The idealized time traces are analyzed as follows to obtain the binding and dissociation rate constants. For the bimolecular association, the bound dwell times are inversely related to the dissociation rate constant, while the unbound dwell times are inversely related to the association rate constants. Given that each bound (or unbound) state is expected – at least in first approximation – to have the same probability to end over time, collectively their dwells should follow an exponential decay frequency distribution, with short dwells more frequent than long dwells. This distribution is fit with an exponential decay function to obtain the corresponding rate constant. Specifically, once time traces of individual molecules have been extracted from a recorded movie, the dwells for the bound and unbound conditions are extracted for each trace of interest through idealization of those states in the software QuB or other such trace idealization packages (Fig. 2). The bound and unbound dwells for all traces are separated into corresponding lists. In a data analysis and graphing software such as Origin Pro, these dwell times are sorted by increasing length, clustered in bins of a constant size, and plotted as a histogram by bin frequency so that a curve resembling an exponential decay is obtained. This curve is fit with an exponential decay equation. Alternatively, the dwell times can be plotted as a cumulative frequency distribution, which gives a smoother curve resembling an exponential increase function, and fit with the following equation to obtain the average lifetime (t): $y = y0 + A1*(1 - e^{(-x/t1)})$, where the reciprocal of t is the corresponding rate constant, which in the case of the association rate constant is simply divided by the concentration of the probe to obtain the bimolecular rate constant. If a poor fit with a single-exponential function is observed, two different kinetic processes could be the cause, and instead a double-exponential equation of the form $(y = y0 + A1*(1 - e^{(-x/t1)}) + A2*(1 - e^{(-x/t2)}))$ is used, wherein t1 and t2 are the two lifetimes. A1 and A2 are the amplitudes for each rate constant, and an average rate constant r_{ave} is obtained by calculating: $r_{ave} = 1/(a1*t1 + a2*t2)$, where $a1 = A1/(A1 + A2)$ and $a2 = A2/(A1 + A2)$).

2.5.3. Distinguishing accessible and non-accessible conformations

One unique advantage of SM studies is their ability to discern among conformational states that individual molecules adopt, provided that the target conformations have lifetimes longer than those of an FP binding and dissociation cycle. To provide this condition, optimization

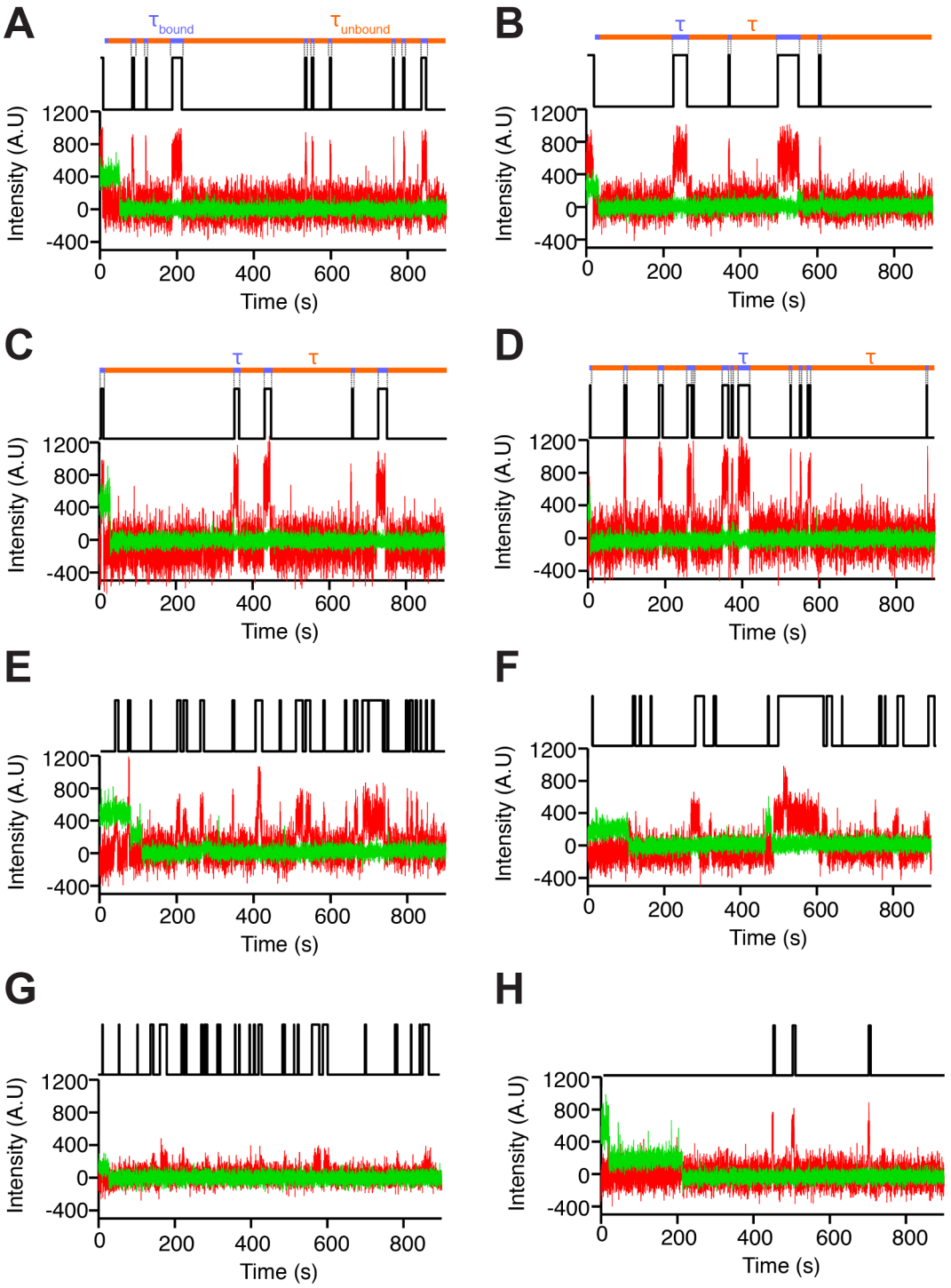

Fig. 2. Trace selection criteria. (A–D) Examples of traces that are selected based on the criteria described in the text. HMM of the binding events is depicted above each trace and dwell time distributions are shown with τ_{bound} (blue) and $\tau_{unbound}$ (orange) indicated. (E–H) Examples of traces that are rejected from further analysis based on the presence multiple Cy5 intensity levels (E,F), low signal-to-noise ratio (G), or a double photo-bleaching step of Cy3 (H). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the FP to yield sufficiently fast association/dissociation rate constants may be needed, and obtained by shortening its length or by introducing non-complementary mutations to accelerate particularly dissociation from the target. The exploration of more than one probe binding site may help in empirically finding a probe with the desired residence times suited for detecting the conformational transitions of interest.

Ideally, a target molecule will switch between FP binding and non-binding conformations, as would be the case for transitions between an "open" and a "closed" state, respectively. In this situation, there will be two (or more) pairs of association and dissociation rate constants of an FP to the same RNA target molecule. Such behavior is observed directly through the SiM-KARTS experiment as periods of high FP accessibility (i.e., "bursts" of binding events) showing high association rates of the target, and conversely few binding events ("non-bursts") when the target is inaccessible [17]. These non-random temporal fluctuations can

be identified through the use of the Fano factor, which for each SM trace is defined as the ratio of the variance of the number of spikes over the mean number of spikes within a time frame [31]. The average Fano factor and deviation can then be calculated for all molecules in a single experimental condition. If the spike distribution is random and results from a Poisson distribution of binding events due to a single pair of association and dissociation rates, this ratio should be equal to 1 and will be independent of the time window used. Instead, a Fano factor increasing with the time window presents evidence that the spike distribution is non-random and ought to be investigated further with a detailed spike train analysis as described by Rinaldi *et al.* [17]. In the latter work, SiM-KARTS using an RNA FP containing the anti-SD sequence of the corresponding 16S ribosomal RNA was deployed to reveal ligand-dependent accessibility changes of the SD sequence of a bacterial mRNA hosting a 7-aminomethyl-7-deazaguanine ($preQ_1$) sensing riboswitch (Fig. 1B). The observation of alternating periods of bursts and non-bursts underscored that single RNA molecules fluctuate between a more open and a more closed SD sequence, respectively, an equilibrium that is shifted towards closing (and thus less ribosomal translation initiation) upon binding of the $preQ_1$ ligand [17]. SiM-KARTS was also used for non-equilibrium ligand-jump experiments [17] and thus generally shown to complement prior smFRET studies of ligand-dependent two-state folding of this translational $preQ_1$ riboswitch [32], while providing unique insights into the parameters affecting downstream gene regulation.

3. Expanding applications

In addition to oligonucleotide binding for secondary structure probing at the single molecule level, the same conceptual approach can be applied to any molecule that reversibly binds to a target, as long as both can be fluorescently labeled. In the following example, we adapted SiM-KARTS to study the binding kinetics of $tRNA^{Gly}$ to a T-box riboswitch, serving as the probe and target, respectively (Fig. 3-B) [33]. The T-box is found in the 5′ untranslated region of Gram-positive bacteria where it controls the expression of genes responsible for maintaining aminoacylated tRNA pools, for example, the GlyQS operon that codes for glycyl-tRNA synthetase in *Bacillus subtilis*. The riboswitch consists of a Stem-I, which binds the tRNA's anticodon and elbow regions, and a downstream antiterminator, which is stabilized by base pairing with the 3′ terminal residues of the non-aminoacylated (or uncharged) tRNA (Fig. 3A). Once the tRNA becomes charged, the antiterminator can no longer be stabilized and a mutually exclusive terminator structure forms instead, thus stopping transcription when aminoacylated tRNAs dominate. Using SiM-KARTS, we studied the binding of the tRNA to T-box intermediates of transcription, including one that can form the antiterminator, in order to determine the kinetic parameters of the binding event that triggers anti-termination before the full terminator sequence even appears from the RNA exit tunnel of bacterial RNAP [33].

3.1. General experimental strategy

To observe transient tRNA:T-box interactions, the tRNA ligand is labeled with Cy5, while the riboswitch is attached to the slide as a "bait" through a biotin molecule incorporated into its 5' end during T7 polymerase transcription with a 5'-biotin-GMP initiator and fluorescently labeled with a Cy3 labeled DNA oligonucleotide hybridizing to its 3′ end [33] (Fig. 3B). Upon analysis, single riboswitch molecules are localized in a field of view by their Cy3 fluorescence and the corresponding Cy5 signal intensity is monitored (Fig. 3C). The time-dependent Cy5-tRNA intensity changes at individual T-box locations are plotted over time and visually inspected. The presence of Cy3 signal undergoing single-step photobleaching identifies traces corresponding to single T-box molecules. Those with sharp Cy5 intensity changes are analyzed as containing binding events: High Cy5 intensity corresponds to tRNA bound to the riboswitch, whereas low Cy5 intensity reflects the absence of tRNA (Fig. 3C). The durations of the intensity changes are fit with a simple two-state HMM and the resulting bound and unbound dwells time for all traces plotted cumulatively and fit with a single- or double-exponential increase function to obtain the dissociation and association rate constants, respectively, of the tRNA ligand to the T-box riboswitch (Fig. 3C).

3.2. Concrete example for sample preparation and data collection

The T-box bait is a 211-residue RNA spanning a section containing Stem I and the antiterminator of the glycine T-box riboswitch from *Bacillus subtilis*, with a biotin moiety covalently attached to its 5′ end and on its 3′ end the added sequence 5′-GAGCACCACGAACAA-3′ to bind stably a 14-nucleotide LP 5′-TGTTCGTGGTGCTC/Cy3/-3′ synthesized by IDT. Neither the LP nor its target are predicted by the RNA structure software [20] to form intramolecular secondary structures or interfere with riboswitch folding. The T_m of this oligonucleotide is predicted as 58.8 °C, thus at room temperature it remains stably bound to the RNA target.

The tRNA ligand is used as the SiM-KARTS probe and synthesized *in vitro* by enzymatically ligating two chemically synthesized 35- and 40-nucleotide long RNA fragments to be able to install an internal Cy5 label on U46 [33]. An RNA mix containing 100 nM of the T-box riboswitch RNA and 2 μM of the Cy3-labeled DNA are heat-annealed at 70 °C in SiM-KARTS buffer (10 mM MES, pH 6.1, 50 mM KCl, 10 mM $MgCl_2$) then allowed to cool to room temperature over 20 min. A 60 pM dilution of the complex is prepared in 200 μl of 10 mM MES, pH 6.1, 50 mM KCl and 10 mM $MgCl_2$, flowed onto the streptavidin-coated slide, and incubated for 5 min at room temperature. Next, 12.5 nM Cy5-labeled $tRNA^{Gly}$ in the same buffer supplemented with the PCA/PCD/Trolox oxygen scavenger is injected. After incubating for 5 to 15 min to allow the oxygen to be scavenged, 15-min movies are collected at a rate of 10 frames per second with both green (532 nm) and red (637 nm) lasers illuminating the sample on the prism-type TIRF microscope. Optionally, 2-by-2 pixel binning can be used on the Hamamatsu scientific CMOS camera to increase the signal-to-noise ratio while reducing the spatial resolution to 216 nm. Time traces are extracted and inspected visually to select those fulfilling the criteria of signal-to-noise threshold, single-step photobleaching of Cy3, and observation of at least two Cy5 binding events. For selected traces, the Cy5 intensities are recorded in QuB format and fit with a two-state HMM that is initialized with first-pass guesses for the ON and OFF rates of 0.1 and 0.01 s^{-1}, respectively. After applying an automatic amplitude estimation and a 1 Hz filter in QuB, the HMM idealization is visually inspected to discard traces that show a poor signal-to-noise ratio and consequent idealization artifacts (Fig. 2). The idealized traces are saved in QuB dwt format, which serially stores the dwell times of state 1 (unbound) and state 2 (bound). Using Matlab scripts created in-house, the data are processed to separate the bound and unbound dwell times, which then are sorted by magnitude and fitted with either single- or double-exponential functions (depending on residuals and goodness of fit) in Origin Pro software (Fig. 3D). For the data presented in Fig. 3D, the reciprocal of the unbound lifetime over the tRNA concentration yields a binding rate constant of 1.01 $\mu M^{-1} s^{-1}$, while the reciprocal of the two tRNA-bound lifetimes yields dissociation rate constants of 0.25 s^{-1} and 0.08 s^{-1}, corresponding to a dissociation equilibrium constant (K_D) of ~168 nM. These values are consistent with those from previous studies [33,34] and support a model wherein the tRNA binding and dissociation rates are fine-tuned to help coordinate ligand binding with the rate of transcription and transcriptional pausing for an efficient readout of the tRNA aminoacylation state. Specifically, $tRNA^{Gly}$ is captured quickly by the initially transcribed 5′ Stem I segment of the riboswitch, remains bound for an average time of only ~4 s based on its relatively fast dissociation rate constant (0.25 s^{-1}), and so gets the chance to quickly interact with the antiterminator to, only if uncharged, form an

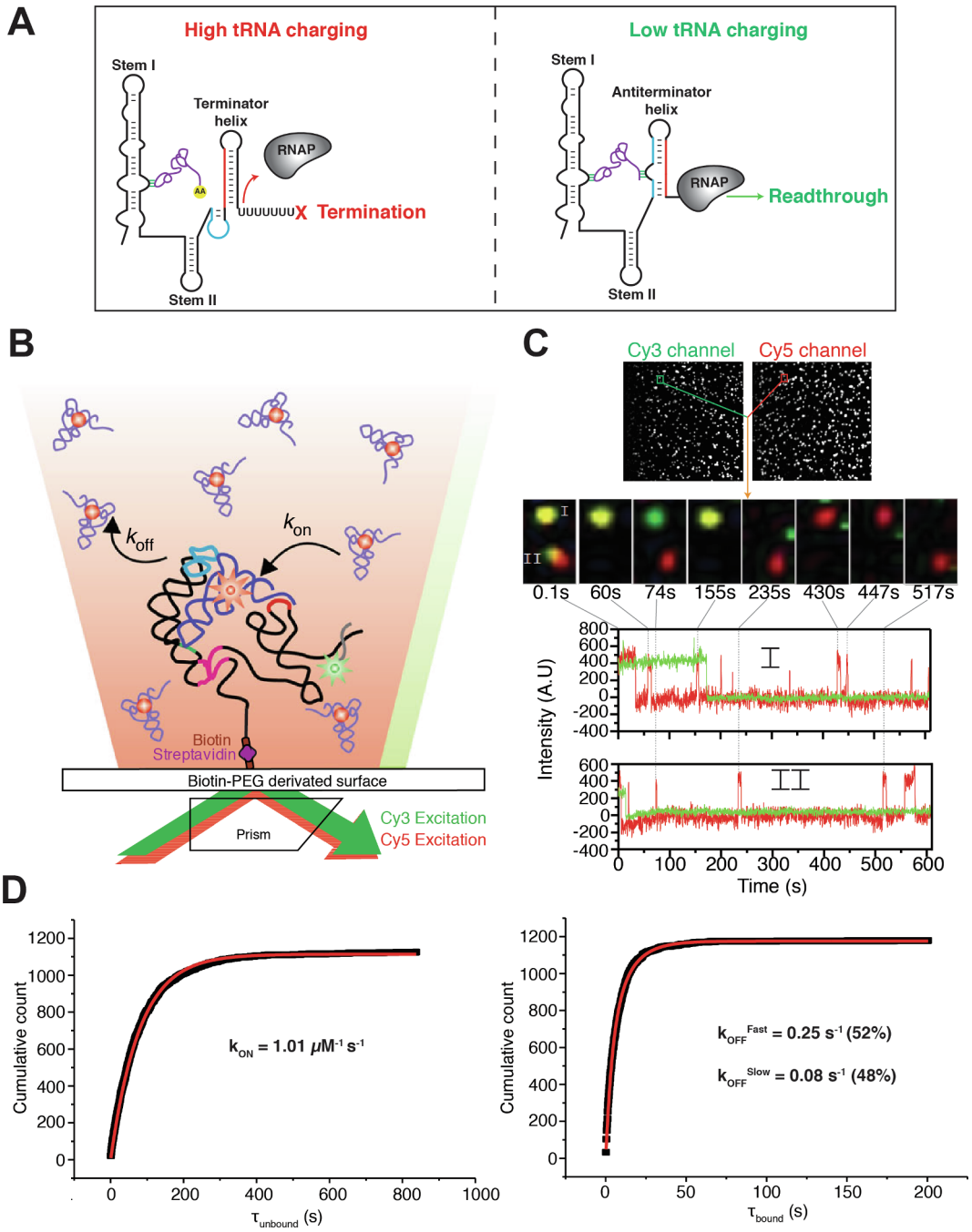

Fig. 3. Kinetic analysis of tRNA binding to the T-box riboswitch. (A) Overview of gene regulation induced by tRNA binding to the T-Box riboswitch. When the tRNA is aminoacylated, weak interaction with the riboswitch induces the formation of a terminator stem-loop, followed by a poly-uridine stretch that together dislodge the RNA polymerase (RNAP) and thus terminate transcription. When the tRNA is uncharged, additional interactions promote the formation of an anti-terminator stem-loop that allows the transcription to progress downstream (B) Schematic illustrating the principle of SiM-KARTS assays using a fluorescently labeled tRNA. (C) The same field of view on a microscope slide is recorded with both green and red fluorophore emitting at the same time (upper panel). The fluorescence signal intensities in both channels are recorded over time and shown overlaid (middle panel), with green indicating Cy3 emission (labeled T-box riboswitch), red representing Cy5 emission (labeled t-RNA), and yellow indicating co-localized molecules as tRNA binds to immobilized T-box. The two resulting single molecule traces, labeled I and II, are shown in the lower panel. (D) $\tau_{unbound}$ (left panel) and τ_{bound} (right panel) were fitted with single- and double-exponential increase functions, respectively, to extract the kinetic parameters of tRNA binding. Amplitudes for each of the double exponential terms are shown in parenthesis (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ultrastable "snap-locked" complex characterized by much slower dissociation [33]. This extremely slow dissociation rate constant of the tRNA is observed only if the antiterminator sequence is present and the tRNA is not aminoacylated, decreasing the dissociation dynamics of the complex greatly while locking in place the antiterminator and preventing the formation of the competing terminator stem.

4. Conclusions and prospects

Here we have described a set of methodologies that uses dual-color illumination of two differentially fluorophore labeled species to study their interaction and either probe the site-specific secondary structure or broader binding partner association of RNA molecules at the single

molecule level. The approach can be easily adopted to include smFRET to confirm the proximity of the two coinciding labels. It is also conceptually similar to single-molecule recognition through equilibrium Poisson sampling (SiMREPS), used to detect nucleic acids with ultra-high sequence specificity [35–37], but emphasizes structure and interaction kinetics over sequence identity.

SiM-KARTS allows for specific probing of local RNA structure and can be applied to very long RNAs that cannot be easily probed by other methods. As a structure probing technique, in principle it can be converted to a high-throughput format with FPs tiled across the entire target RNA sequence. Not only secondary, but also tertiary structure changes can likely be detected in real-time with a time resolution primarily limited by the camera frame rate needed to obtain a sufficient signal-to-noise ratio. Using two probes hybridizing to different locations of the same target, distinct regions of the same RNA can be interrogated for (potentially mutually exclusive) RNA conformational changes. Given its versatility, SiM-KARTS can be adapted easily to other types of biological questions concerning RNA or protein recruitment in *trans*. The latter would be conceptually similar to the co-localization single-molecule microscopy (COSMOS) described by Gelles and coworkers [38–40]. More generally, any dynamic and equilibrium or non-equilibrium process can be monitored for which specific FPs can be designed, for example, the co-transcriptional folding of an RNA emerging from the exit tunnel of RNAP. The future appears bright for SiM-KARTS and related single molecule fluorescence microscopy tools designed to probe the dynamic nature of molecular biology.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Dr. Edward Nikonowicz for initiating the T-Box riboswitch project and providing the nucleic acids used in this study, Dr. Paul Lund and Dr. Sujay Ray for technical expertise and all members of the Walter laboratory for valuable feedback.

Funding

This work was supported by NIH grants GM062357, GM118524, and a sub-award on GM115857 to N.G.W.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2019.04.002.

References

[1] T.R. Cech, J.A. Steitz, The noncoding RNA revolution—trashing old rules to forge new ones, Cell 157 (2014) 77–94.
[2] H.M. Al-Hashimi, N.G. Walter, RNA dynamics: it is about time, Curr. Opin. Struct. Biol. 18 (2008) 321–329.
[3] E.J. Strobel, A.M. Yu, J.B. Lucks, High-throughput determination of RNA structures, Nat. Rev. Genet. 19 (2018) 615–634.
[4] T.R. Mercer, M.E. Dinger, J.S. Mattick, Long non-coding RNAs: insights into functions, Nat. Rev. Genet. 10 (2009) 155–159.
[5] S.A. Mortimer, M.A. Kidwell, J.A. Doudna, Insights into RNA structure and function from genome-wide studies, Nat. Rev Genet. (2014).
[6] S. Ray, A. Chauvier, N.G. Walter, Kinetics coming into focus: single-molecule microscopy of riboswitch dynamics, RNA Biol. 1–9 (2018).
[7] M. Blanco, N.G. Walter, Analysis of complex single-molecule FRET time trajectories, Methods Enzymol. 472 (2010) 153–178.
[8] S. Ray, J.R. Widom, N.G. Walter, Life under the microscope: single-molecule fluorescence highlights the RNA world, Chem. Rev. 118 (2018) 4120–4155.
[9] W.J. Greenleaf, K.L. Frieda, D.A.N. Foster, M.T. Woodside, S.M. Block, Direct observation of hierarchical folding in single riboswitch aptamers, Science 319 (2008) 630–633.
[10] K.A. LeCuyer, D.M. Crothers, Kinetics of an RNA conformational switch, PNAS 91 (1994) 3373–3377.
[11] J.A. Hoerter, M.N. Lambert, M.J. Pereira, N.G. Walter, Dynamics inherent in helix 27 from *Escherichia coli* 16S ribosomal RNA, Biochemistry 43 (2004) 14624–14636.
[12] D.K. Treiber, J.R. Williamson, Kinetic oligonucleotide hybridization for monitoring kinetic folding of large RNAs, Methods Enzymol. 317 (2000) 330–353.
[13] M. Sohail, S. Akhtar, E.M. Southern, The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides, RNA 5 (1999) 646–655.
[14] G.A. Perdrizet 2nd, I. Artsimovitch, R. Furman, T.R. Sosnick, T. Pan, Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch, PNAS 109 (2012) 3323–3328.
[15] A. Lussier, L. Bastet, A. Chauvier, D.A. Lafontaine, A kissing loop is important for btuB riboswitch ligand sensing and regulatory control, J. Biol. Chem. 290 (2015) 26739–26751.
[16] A. Chauvier, F. Picard-Jean, J.-C. Berger-Dancause, L. Bastet, M.R. Naghdi, A. Dubé, P. Turcotte, J. Perreault, D.A. Lafontaine, Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation, Nat. Commun. 8 (2017) 13892.
[17] A.J. Rinaldi, P.E. Lund, M.R. Blanco, N.G. Walter, The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts, Nat. Commun. 7 (2016) 8976.
[18] A.J. Rinaldi, K.C. Suddala, N.G. Walter, Native purification and labeling of RNA for single molecule fluorescence studies, Methods Mol. Biol. 1240 (2015) 63–95.
[19] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res. 31 (2003) 3406–3415.
[20] J.S. Reuter, D.H. Mathews, RNA structure: software for RNA secondary structure prediction and analysis, BMC Bioinf. 11 (2010) 129.
[21] N. Michelotti, C. de Silva, A.E. Johnson-Buck, A.J. Manzo, N.G. Walter, A bird's eye view tracking slow nanometer-scale movements of single molecular nano-assemblies, Methods Enzymol. 475 (2010) 121–148.
[22] R. Lamichhane, A. Solem, W. Black, D. Rueda, Single-molecule FRET of protein-nucleic acid and protein-protein complexes: surface passivation and immobilization, Methods 52 (2010) 192–200.
[23] W. Kern, The evolution of silicon wafer cleaning technology, J. Electrochem. Soc. 137 (1990) 1887–1892.
[24] C.E. Aitken, R.A. Marshall, J.D. Puglisi, An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments, Biophys. J. 94 (2008) 1826–1835.
[25] G. Senavirathne, M.A. Lopez Jr., R. Messer, R. Fishel, K.E. Yoder, Expression and purification of nuclease-free protocatechuate 3,4-dioxygenase for prolonged single-molecule fluorescence imaging, Anal. Biochem. 556 (2018) 78–84.
[26] D. Rueda, G. Bokinsky, M.M. Rhodes, M.J. Rust, X. Zhuang, N.G. Walter, Single-molecule enzymology of RNA: essential functional groups impact catalysis from a distance, PNAS 101 (2004) 10066–10071.
[27] D.W. Mount, Using hidden Markov models to align multiple sequences, Cold Spring Harbor Protoc. (2009 (2009)) pdb.top41.
[28] F. Qin, L. Li, Model-based fitting of single-channel dwell-time distributions, Biophys. J. 87 (2004) 1657–1671.
[29] S.A. McKinney, C. Joo, T. Ha, Analysis of single-molecule FRET trajectories using hidden Markov modeling, Biophys. J . 91 (2006) 1941–1951.
[30] J.E. Bronson, J. Fei, J.M. Hofman, R.L. Gonzalez, C.H. Wiggins, Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data, Biophys. J. 97 (2009) 3196–3205.
[31] U.T. Eden, M.A. Kramer, Drawing inferences from Fano factor calculations, J. Neurosci. Methods 190 (2010) 149–152.
[32] K.C. Suddala, A.J. Rinaldi, J. Feng, A.M. Mustoe, C.D. Eichhorn, J.A. Liberman, J.E. Wedekind, H.M. Al-Hashimi, C.L. Brooks 3rd, N.G. Walter, Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure, Nucleic Acids Res. 41 (2013) 10462–10475.
[33] K.C. Suddala, J. Cabello-Villegas, M. Michnicka, C. Marshall, E.P. Nikonowicz, N.G. Walter, Hierarchical mechanism of amino acid sensing by the T-box riboswitch, Nat. Commun. 9 (2018) 1896.
[34] J. Zhang, B. Chetnani, E.D. Cormack, D. Alonso, W. Liu, A. Mondragon, J. Fei, Specific structural elements of the T-box riboswitch drive the two-step binding of the tRNA ligand, Elife 7 (2018).
[35] A. Johnson-Buck, X. Su, M.D. Giraldez, M. Zhao, M. Tewari, N.G. Walter, Kinetic fingerprinting to identify and count single nucleic acids, Nat. Biotechnol. 33 (2015) 730–732.
[36] S.L. Hayward, P.E. Lund, Q. Kang, A. Johnson-Buck, M. Tewari, N.G. Walter, Ultraspecific and amplification-free quantification of mutant DNA by single-molecule kinetic fingerprinting, J. Am. Chem. Soc. 140 (2018) 11755–11762.
[37] A. Johnson-Buck, J. Li, M. Tewari, N.G. Walter, A guide to nucleic acid detection by single-molecule kinetic fingerprinting, Methods 153 (2019) 3–12.
[38] L.J. Friedman, J. Gelles, Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation, Cell 148 (2012) 679–689.
[39] T.T. Harden, C.D. Wells, L.J. Friedman, R. Landick, A. Hochschild, J. Kondev, J. Gelles, Bacterial RNA polymerase can retain σ70 throughout transcription, PNAS 113 (2016) 602–607.
[40] L.E. Tetone, L.J. Friedman, M.L. Osborne, H. Ravi, S. Kyzer, S.K. Stumper, R.A. Mooney, R. Landick, J. Gelles, Dynamics of GreB-RNA polymerase interaction allow a proofreading accessory protein to patrol for transcription complexes needing rescue, PNAS 114 (2017) E1081–E1090.