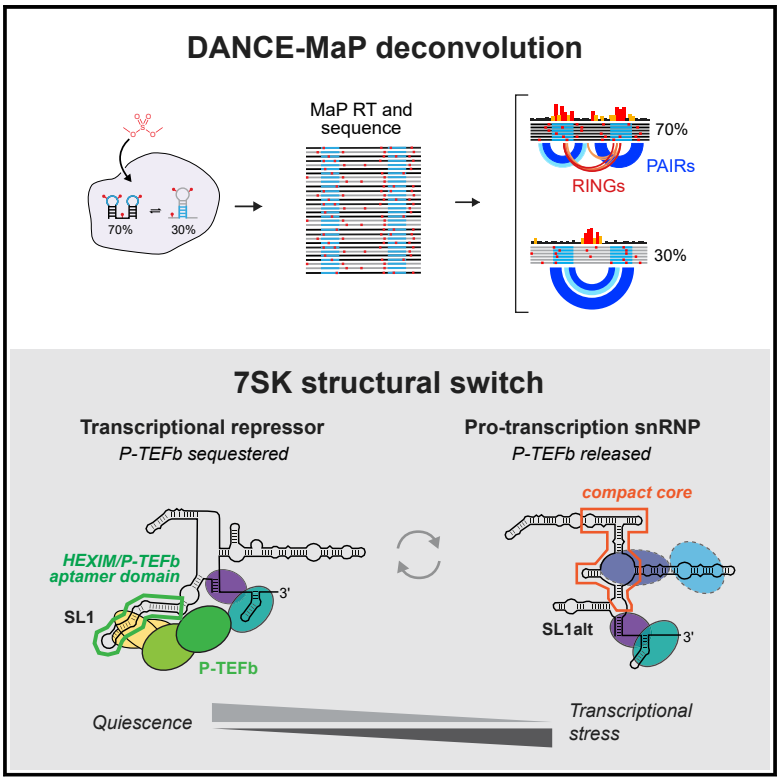
Molecular Cell

Article

Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP

Graphical abstract



Authors

Samuel W. Olson, Anne-Marie W. Turner, J. Winston Arney, ..., David M. Margolis, Kevin M. Weeks, Anthony M. Mustoe

Correspondence

weeks@unc.edu (K.M.W.), anthony.mustoe@bcm.edu (A.M.M.)

In brief

Olson et al. present a new RNA chemical probing strategy, DANCE-MaP, for determining dynamic RNA structural ensembles in living cells. DANCE-MaP reveals that the human 7SK noncoding RNA encodes a structural switch that regulates transcription in response to cell state.

Highlights

- DANCE-MaP enables complete characterization of RNA structural ensembles in cells
- The 7SK RNA encodes a large-scale structural switch that modulates P-TEFb binding
- 7SK structure switches dynamically in response to cell growth and stress
- The 7SK switch is a potential therapeutic target for modulating transcription

Olson et al., 2022, Molecular Cell *82*, 1708–1723
May 5, 2022
https://doi.org/10.1016/j.molcel.2022.02.009

CellPress

CellPress

Molecular Cell

Article

Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP

Samuel W. Olson,[1] Anne-Marie W. Turner,[2,3] J. Winston Arney,[1] Irfana Saleem,[4] Chase A. Weidmann,[1] David M. Margolis,[2,3] Kevin M. Weeks,[1,6,*] and Anthony M. Mustoe[4,5,*]

[1]Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290, USA
[2]Department of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA
[3]UNC HIV Cure Center, University of North Carolina, Chapel Hill, NC 27599, USA
[4]Verna and Marrs McClean Department of Biochemistry and Molecular Biology, Therapeutic Innovation Center (THINC), Baylor College of Medicine, Houston, TX 77030, USA
[5]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[6]Lead contact
*Correspondence: weeks@unc.edu (K.M.W.), anthony.mustoe@bcm.edu (A.M.M.)
https://doi.org/10.1016/j.molcel.2022.02.009

SUMMARY

7SK is a conserved noncoding RNA that regulates transcription by sequestering the transcription factor P-TEFb. 7SK function entails complex changes in RNA structure, but characterizing RNA dynamics in cells remains an unsolved challenge. We developed a single-molecule chemical probing strategy, DANCE-MaP (deconvolution and annotation of ribonucleic conformational ensembles), that defines per-nucleotide reactivity, direct base pairing interactions, tertiary interactions, and thermodynamic populations for each state in RNA structural ensembles from a single experiment. DANCE-MaP reveals that 7SK RNA encodes a large-scale structural switch that couples dissolution of the P-TEFb binding site to structural remodeling at distal release factor binding sites. The 7SK structural equilibrium shifts in response to cell growth and stress and can be targeted to modulate expression of P-TEFbresponsive genes. Our study reveals that RNA structural dynamics underlie 7SK function as an integrator of diverse cellular signals to control transcription and establishes the power of DANCE-MaP to define RNA dynamics in cells.

INTRODUCTION

RNA molecules fold into complex secondary and tertiary structures that direct broad regulatory functions (Sharp, 2009; Cech and Steitz, 2014). Most RNAs can fold into more than one structure, which enable RNAs to function as regulatory switches that change conformation in response to ligand or protein binding (Dethoff et al., 2012). Diverse riboswitches have been identified in mRNAs that regulate transcription, splicing, and translation in response to metabolites (Breaker, 2012) and proteins (Ray et al., 2009; Fu et al., 2013). Large-scale RNA structural dynamics are also critical to the function of ribonucleoprotein (RNP) complexes such as the ribosome (Rodnina et al., 2017; Sengupta et al., 2019) and the spliceosome (Wilkinson et al., 2020). Despite their broad importance, RNA switches remain exceedingly difficult to identify, quantify in terms of their structure and in-cell equilibria, or link to functional outcomes.

The 7SK RNA is the 332-nucleotide noncoding RNA component of the 7SK small noncoding RNA-protein complex (snRNP) (Peterlin et al., 2012; Quaresma et al., 2016). 7SK is principally thought to function by sequestering Cdk9/Cyclin T1 (termed positive transcription elongation factor b, P-TEFb), a key transcription factor that phosphorylates RNA polymerase II (RNA Pol II) paused at promoter-proximal regions to facilitate productive elongation (Peterlin et al., 2012; Quaresma et al., 2016). The 7SK snRNP also regulates splicing (Barboric et al., 2009; Egloff et al., 2017; Ji et al., 2021) and chromatin remodeling (Eilebrecht et al., 2011; Flynn et al., 2016). The functions of the 7SK snRNP are driven by coordinated changes in its bound protein components. The core snRNP is stabilized by the proteins MePCE—which methylates and remains bound to the 5′ phosphate of 7SK RNA (Yang et al., 2019)—and LARP7—which binds to the RNA 3′ end (Krueger et al., 2008; Eichhorn et al., 2018). P-TEFb is sequestered through interactions with the accessory protein dimer HEXIM1/2, which binds to the 7SK RNA at a high-affinity stem-loop structure, SL1 (Peterlin and Price, 2006; Czudnochowski et al., 2010; Martinez-Zapien et al., 2016). Upon stimulation of transcription, various factors act on 7SK to liberate P-TEFb and HEXIM1/2 (Peterlin et al., 2012; Quaresma et al., 2016). The P-TEFb-free 7SK is subsequently bound by distinct proteins and likely has a heterogenous composition (Peterlin et al., 2012; Quaresma et al., 2016). An understanding of 7SK regulatory mechanisms will both illuminate fundamental aspects of transcriptional control and also inform ongoing efforts to inhibit transcription in disease settings, especially cancer (Olson et al., 2018) and, conversely, to reverse transcriptional latency of

HIV as part of a cure strategy (Cary et al., 2016; Margolis et al., 2020).

Multiple studies have indicated that P-TEFb binding and release involves remodeling of the 7SK RNA structure (Krueger et al., 2010; Flynn et al., 2016; Brogie and Price, 2017). However, 7SK structure and its dynamics remain poorly defined. 7SK sequences are extraordinarily conserved, limiting informative evolutionary covariation analysis (Rivas et al., 2017; Kalvari et al., 2021). Several structure models for 7SK have been proposed based on chemical probing experiments (Wassarman and Steitz, 1991; Marz et al., 2009; Brogie and Price, 2017; Luo et al., 2021), but the heterogeneity of 7SK challenges interpretation of these ensemble-average measurements.

Single-molecule chemical probing is emerging as a transformative approach for characterizing RNA structure and dynamics in living cells. These technologies are made possible by mutational profiling (MaP) reverse transcription, whereby a polymerase reads through multiple chemical adducts per RNA molecule, recording them as mutations in complementary DNA (Homan et al., 2014). Massively parallel sequencing is then used to measure correlated modification events across thousands to millions of molecules, yielding information on RNA structural ensemble composition (Homan et al., 2014; Tomezsko et al., 2020; Morandi et al., 2021) and through-space secondary (Krokhotin et al., 2017; Cheng et al., 2017; Mustoe et al., 2019) and tertiary (Homan et al., 2014; Dethoff et al., 2018; Sengupta et al., 2019) interactions. Although powerful, existing single-molecule analysis frameworks only extract one type of information at a time: Ensemble deconvolution strategies can measure multiple coexisting per-nucleotide reactivity profiles but do not directly measure base pairs. Conversely, the coexistence of multiple RNA structures makes it challenging to disambiguate direct base pair and tertiary interaction measurements.

To overcome these issues, we developed DANCE-MaP (deconvolution and annotation of ribonucleic conformational ensembles), a maximum likelihood (ML) strategy that extracts and annotates a large fraction of the total information from single-molecule chemical probing experiments. DANCE-MaP directly visualizes complex RNA ensembles based on MaP probing data at nucleotide resolution, including detection of base pairs and tertiary interactions, in a single experiment. We extensively benchmarked this strategy by analyzing the adenine riboswitch and uncovered significant complexity, even within this well-defined RNA structural ensemble. We then applied DANCE-MaP to discover a large-scale allosteric switch in the 7SK RNA that links P-TEFb release and cell state-specific transcriptional regulation to concerted remodeling of the 7SK structure. Our work establishes DANCE-MaP as a powerful framework for resolving complex RNA structural ensembles in cells and for linking these structures to function.

RESULTS

Deconvolution of complex RNA ensembles with thermodynamic accuracy

In single-molecule chemical probing experiments, each MaP read represents a structural snapshot of an individual RNA molecule. For RNAs that fold into multiple structures, each structure will generate distinct groups of reactive and unreactive nucleotides that are comodified or unmodified in a correlated manner (Homan et al., 2014). We implemented a ML framework that uses a modified Bernoulli mixture model to fit single-molecule reads to multiple reactivity profiles, sequentially increasing the number of fitted states until the optimal solution is identified (Figures 1A, 1B, and S1). Our ML strategy, developed independently, shares an architecture used previously (Tomezsko et al., 2020) but includes adjustments to handle missing data, improve robustness, and capture information from all four RNA nucleotides. Under idealized scenarios, our ML framework can deconvolute ensembles consisting of up to five states with populations $\geq$5% (Data S1). Deconvolution accuracy breaks down as ensemble heterogeneity increases to four or more highly divergent states, meaning that individual deconvoluted states may contain residual structural heterogeneity. Overall, reactivity profiles and populations are resolved with mean errors of <1% for ensembles consisting of distinctive states.

We validated our ML framework using the *Vibrio vulnificus add* adenine riboswitch, which folds into a two-state ensemble consisting of a translation OFF state that occludes the Shine-Dalgarno sequence and a translation ON state that binds adenine (Reining et al., 2013). Structure probing experiments were performed on an *in vitro* transcribed RNA using dimethyl sulfate (DMS) under conditions where the reagent reacts with all four nucleotides (Mustoe et al., 2019). When analyzed by a conventional averaged reactivity profile, the riboswitch appears to fold into a single well-defined structure (Figure 1C). By comparison, deconvolution with our ML framework yields structures matching those of the translation ON and OFF states at a ratio of 40:60 (Figures 1C–1E), in agreement with independent measurements (Reining et al., 2013). We adapted algorithms developed for SHAPE probing analysis (Siegfried et al., 2014; Smola et al., 2015a; Weeks, 2021) to compute pairing probabilities guided by DMS reactivities (Mustoe et al., 2019). Pairing probability analysis revealed notable heterogeneity in the P2alt region in the OFF state (Figure 1E), consistent with prior observations (Reining et al., 2013; Tomezsko et al., 2020), which would be challenging to identify using standard minimum free-energy approaches. We also observe the potential for an extended P1 pairing (P1ext) in the ON state (Figure 1E).

We next tested the ability of our ML framework to measure *add* riboswitch switching in response to adenine binding. The population of the ON state increases exactly as expected following the adenine binding isotherm, enabling us to compute a $K_d \approx 2\ \mu M$ in agreement with literature values (Figure 1F) (Reining et al., 2013). Consistent with prior studies (Warhaut et al., 2017; Tomezsko et al., 2020), both OFF and ON conformations are populated even at saturating adenine ligand (Figures 1F and S2). The OFF state reactivity profiles are essentially identical at all adenine concentrations (Figure 1G, bottom). By contrast, reactivity changes within the ligand-binding pocket of the ON state indicate adenine-induced stabilization of the aptamer domain tertiary structure (Figure 1G, top). The ability to resolve ligand-binding thermodynamics and subtle state-specific structural changes is unique to our study and validates the precision of our ML deconvolution implementation.

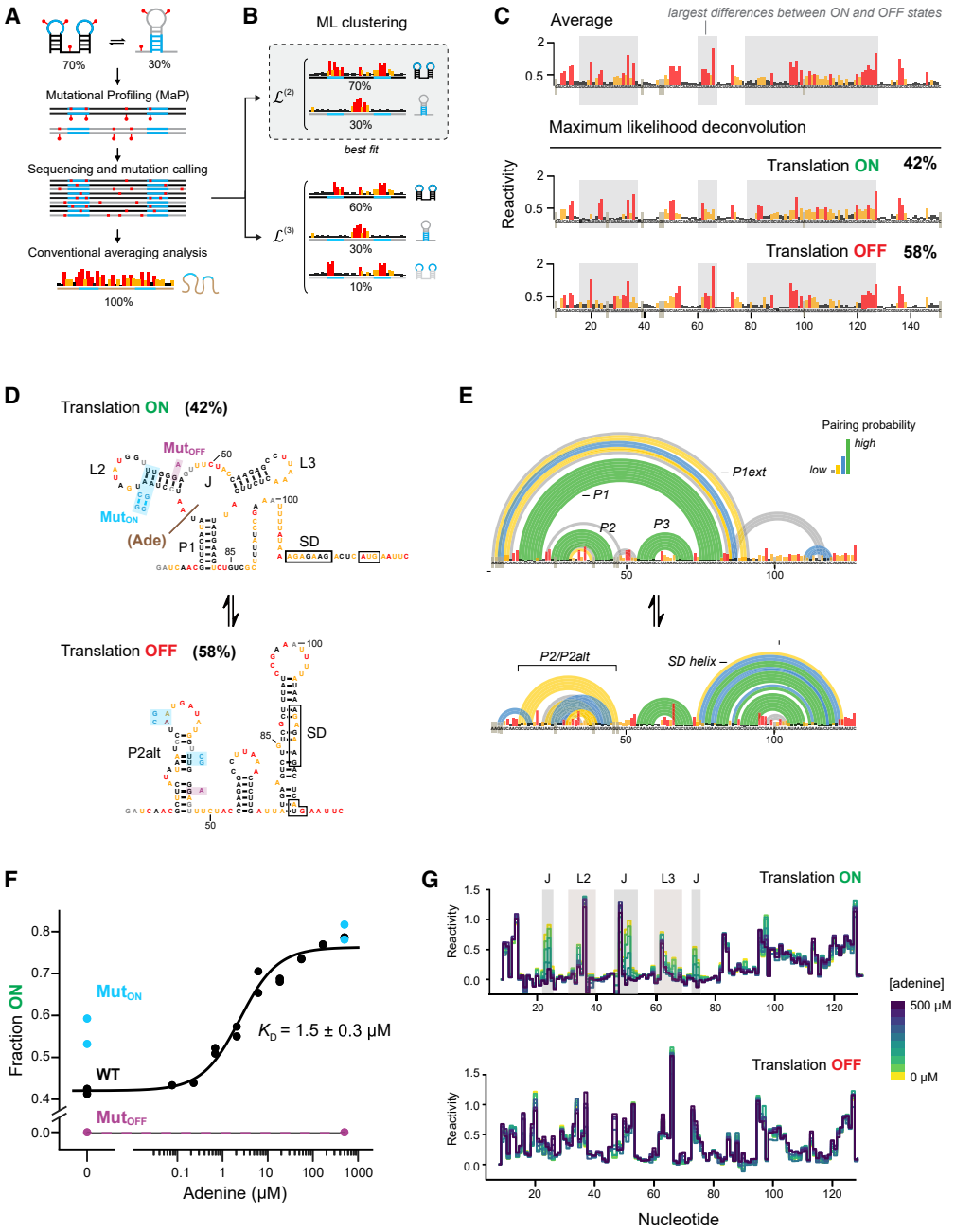


Figure 1. Maximum likelihood deconvolution enables thermodynamically rigorous analysis of RNA conformational ensembles

(A and B) Schematic of single-molecule chemical probing and ML ensemble deconvolution. RNAs with multiple structures generate distinctive chemical modification patterns. (A) Conventional analysis yields a single, average per-nucleotide reactivity profile that may not be representative of any underlying structure state. (B) ML analysis reveals reactivity profiles and populations of each state.

(C) Averaged and deconvoluted DMS-MaP data for the ON and OFF states of the *add* adenine riboswitch in the absence of ligand. High, medium, and low DMS reactivities are shown in red, yellow, and black, respectively. Major reactivity differences are shaded.

(D) Deconvoluted MaP data for the *add* riboswitch in the absence of ligand, superimposed on the NMR-defined (Reining et al., 2013) ON and OFF states. Mut_{ON} (Reining et al., 2013) and Mut_{OFF} mutants are also illustrated. The *add* RNA is numbered per prior convention (Reining et al., 2013).

(E) State-specific pairing probabilities for the ON and OFF states computed from deconvoluted reactivity profiles.

(F) Population of the ON state as a function of adenine concentration for native and mutant sequences. Reported K_d represents the mean ± SD from fits to two independent datasets.

(G) Reactivities for the ON and OFF states as a function of adenine concentration. Regions that undergo adenine-dependent protection are emphasized with gray shading.

See also Figures S1 and S2 and Data S1.

We further validated our approach using mutants that perturb the OFF-ON equilibrium. Consistent with other studies (Reining et al., 2013; Warhaut et al., 2017; Tian et al., 2018), destabilizing the P2alt helix via the Mut_{ON} mutation only modestly shifted the equilibrium toward the ON state (Figure 1F); we discovered that this incomplete shift is due, in part, to the creation of a misfolded third state (Figure S2). Destabilizing P2 via our Mut_{OFF} mutation switched the equilibrium to 65:35 between OFF and a state that is not adenine responsive (Figures 1F and S2). Collectively, these data validate the robustness and thermodynamic accuracy of our ML deconvolution framework.

Direct measurement of base pairs and tertiary interactions for individual RNA states

ML deconvolution provides critical insights into ensemble composition and populations; however, the set of RNA base pairs must be inferred via structure modeling. Modeling is often ambiguous, with multiple pairings compatible with the same data. Bound proteins can further complicate data interpretation (Smola et al., 2015b). Potential tertiary interactions are also not observed. Single-molecule probing data can address these challenges and contain characteristic correlations that directly report on through-space base pairing (PAIRs) and tertiary interactions (RINGs) (Figure 2A) (Homan et al., 2014; Mustoe et al., 2019). We developed a read assignment strategy that, in combination with our ML deconvolution approach, enables simultaneous measurement of per-nucleotide reactivity profiles and state-specific base pairing and tertiary interactions in complex structural ensembles. We term this approach deconvolution and annotation of ribonucleic conformational ensembles measured by mutational profiling (DANCE-MaP). Analysis of simulated MaP datasets confirmed that DANCE-MaP enables accurate measurement of state-specific RING and PAIR correlations (STAR Methods; Figure S1; Data S1).

Without ML deconvolution, PAIR and RING analyses of *add* riboswitch DMS-MaP data yielded a dense, ambiguous meshwork of correlations (Figure 2B). In contrast, DANCE-MaP revealed a specific and near-complete network of direct PAIR interactions that clearly define the ON and OFF state secondary structures (Figure 2C). Equally striking, RINGs identify the L2-L3 loop-loop tertiary interaction in the ON state (Figure 2C), which forms even in the absence of ligand (Warhaut et al., 2017). These data were reproducible across a comprehensive adenine titration and in the Mut_{ON} mutant (Figure S2). In sum, DANCE-MaP enables complete multistate structural analysis within a single, concise chemical probing experiment.

Native 7SK RNA exists as a multistate structural ensemble

To evaluate the 7SK RNA structural ensemble and its role in regulating transcription, we performed DMS-MaP experiments on living human Jurkat cells and obtained single-molecule DMS probing data (>300,000× coverage) for the 7SK RNA. Conventional averaging analysis yielded per-nucleotide reactivity profiles generally compatible with previously proposed SL1, SL3, and SL4 structures (Wassarman and Steitz, 1991; Marz et al., 2009) (Figure 3A). However, as observed previously (Wassarman and Steitz, 1991; Krueger et al., 2010; Brogie and Price, 2017; Wang et al., 2019), many nucleotides exhibit intermediate reactivities, consistent with significant structural heterogeneity.

ML single-molecule analysis indicated that 7SK structural heterogeneity resolves into three states: A (40% ± 3 of the population), B (47% ± 2), and H (13% ± 2) (Figure 3A). In-cell state populations and reactivity profiles were highly reproducible over 10 biological replicates performed years apart ($R > 0.97$; Table S1). The minority H state is characterized by high reactivity across the RNA, indicating that this state is heterogenous. The majority A and B states have regions of punctate high and low nucleotide reactivity, consistent with well-defined structural states. Nucleotides throughout the SL1 region, including U28, U30, U66, and U68, are unreactive in state A but reactive in states B and H. This region was previously shown to change conformation upon P-TEFb release (Krueger et al., 2010; Brogie and Price, 2017). Differences between the A and B states were also observed elsewhere in the RNA, such as in the regions flanking SL3.

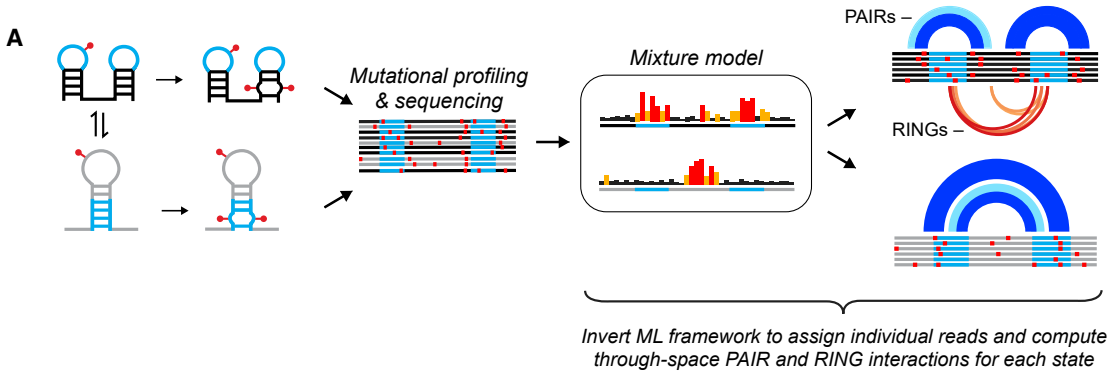
Experiments performed on extracted and refolded RNA from Jurkat cells (referred to as cell free) indicated that 7SK also populates three states in the absence of proteins: A (43% ± 8), B (29% ± 6), and a cell-free-specific mixed M state (28% ± 1) (Table S1). Cell-free states A and B are the same as observed in cells and exhibit only diffuse reactivity protections and enhancements (Figure 3B), implying that these states are dynamically rather than stably bound by proteins in cells. The M state is well structured and appears to be an intermediate between A and B (Figure S3). The lack of the M state in cells suggests that bound cellular factors specifically favor the A and B states. In sum, the 7SK RNA encodes two primary states that are energetically balanced and which behave similarly with or without bound proteins.

Direct base pair mapping and structure modeling reveal 7SK architecture

We next obtained higher depth sequencing datasets (>3 million reads) that provide power sufficient to detect through-space PAIRs and RINGs across the 7SK RNA. These data revealed numerous PAIR signals that directly report base-paired structural elements distinctive to each state (Figure 3C) and that were reproducible between in-cell and cell-free environments and across consolidated replicates (Figure S4). We used PAIR data in combination with per-nucleotide reactivity profiles to build detailed secondary structure models for the A and B states (Figures 3D and S3). Residual heterogeneity indicates that states A and B should be interpreted as class averages rather than pure states, but each state possesses defining structural features that support a link between 7SK conformational dynamics and P-TEFb binding and release.

State A is the P-TEFb binding-competent state with a dynamic SL0 stem

State A blends features predicted by early probing studies (Wassarman and Steitz, 1991) and more recent evolutionary analyses (Marz et al., 2009) (Figure S5). The SL1 helix is the defining structural feature of state A and is supported by PAIRs detected both in cells and in-cell-free RNA (Figure 3C). SL1 has been extensively validated as the recognition site for HEXIM1/2 and P-TEFb based on *in vitro* binding assays (Lebars et al., 2010;



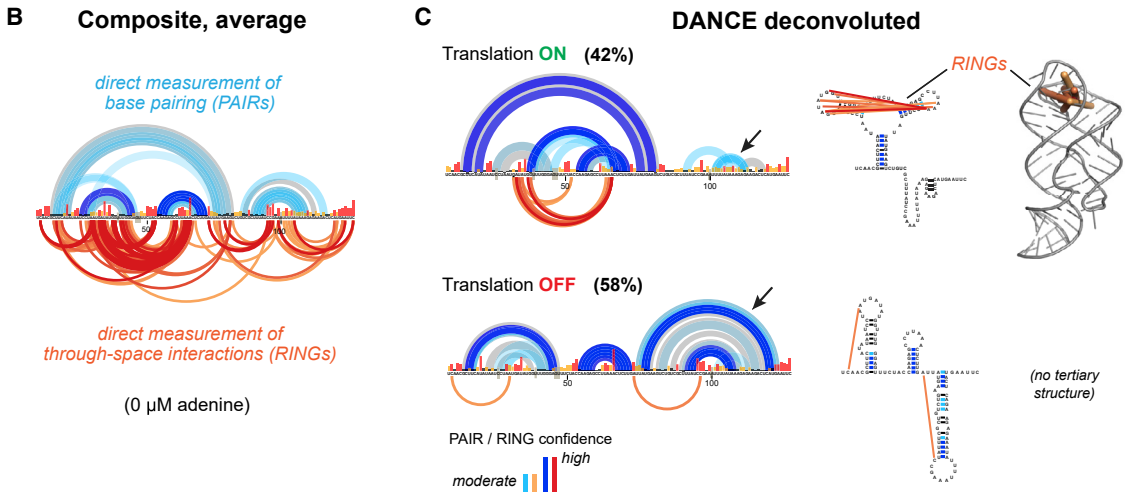


Figure 2. DANCE-MaP detection of state-specific RNA base pairs and tertiary interactions

(A) DMS induces correlated disruptions of base pairing and tertiary interactions, measurable by MaP. Following ML deconvolution of per-nucleotide reactivity profiles, individual reads are assigned to distinct states. PAIR and RING correlation analyses are then used to directly detect base pairing and through-space tertiary interactions, respectively.

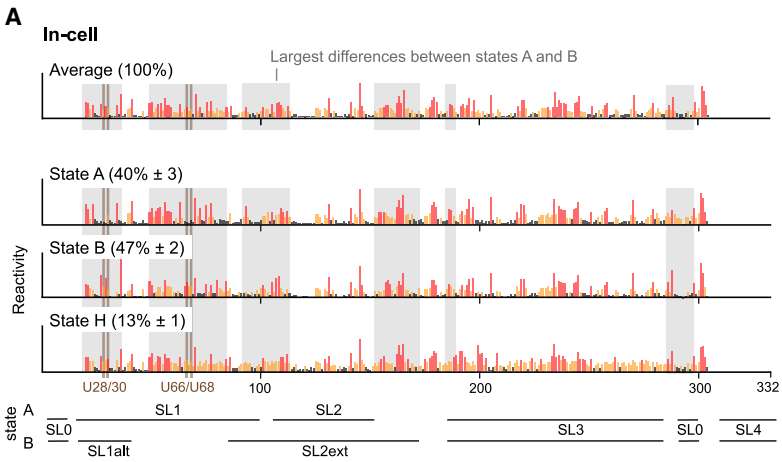
(B) PAIR and RING analyses of composite (nondeconvoluted) *add* riboswitch data measured in the absence of adenine ligand. PAIRs are shown at top, superimposed on the modeled secondary structure state (light gray). High and moderate confidence PAIRs (called principal and minor previously; Mustoe et al., 2019) are dark and light blue, respectively. RINGs are shown at bottom, colored according to statistical significance.

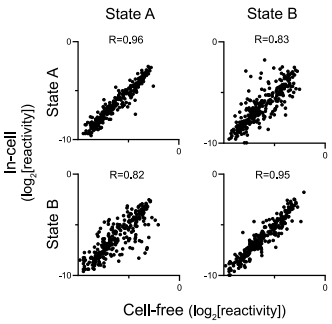
(C) Left: DANCE-MaP deconvolution of *add* riboswitch data. Arrows highlight state-specific PAIRs only observed upon deconvolution. Middle: PAIR and RING correlations superimposed on standard secondary structure diagrams. Right: ON state RINGs superimposed on the crystal structure of the aptamer domain (PDB: 4tzx; Zhang and Ferré-D'Amaré, 2014).

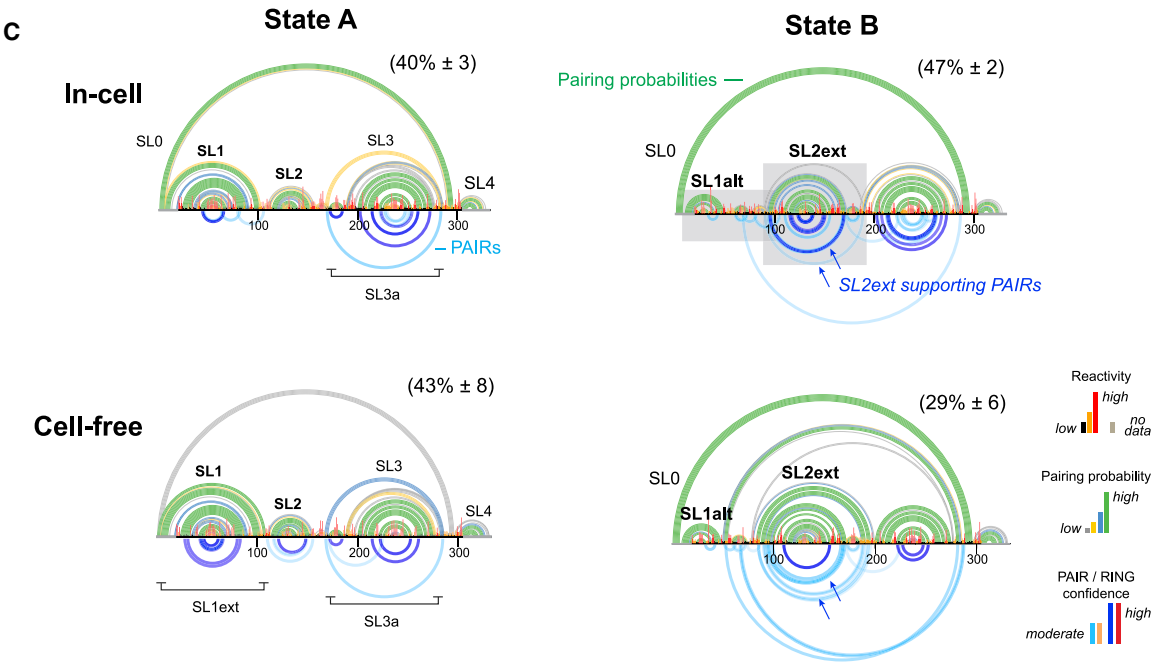
See also Figures S1 and S2 and Data S1.

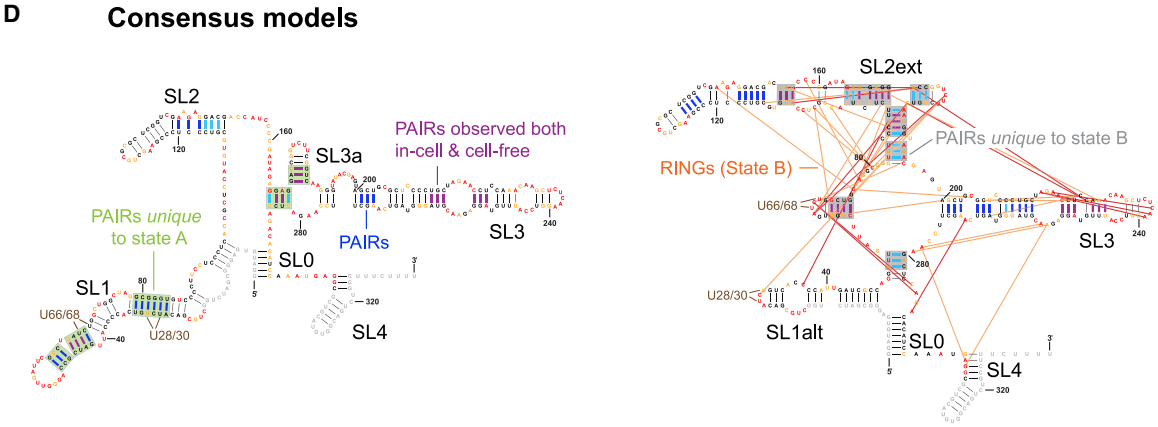
Martinez-Zapien et al., 2016), analysis of P-TEFb-bound 7SK fractions from cells (Brogie and Price, 2017), and in-cell functional assays (Egloff et al., 2006; Fujinaga et al., 2014). The population of state A in cells, 40%, is also consistent with the estimated fraction of 7SK bound by P-TEFb (Nguyen et al., 2001; Yang et al., 2001). Thus, we assign state A as the P-TEFb-bound state.

Structure modeling indicates that this P-TEFb-bound state contains a long-range pairing interaction, SL0, between the 5′ and 3′ ends that "circularizes" the RNA in cells (Marz et al., 2009). We lack data for the 5′ strand of SL0 due to overlap with the primer binding site, but the 3′ strand of SL0 is lowly to moderately reactive, consistent with formation of a dynamic, partially stable stem (Figures 3 and S3). By comparison, cell-free 7SK RNA features an alternative extended form of SL1, incompatible with SL0 (SL1ext; Figure 3C). The increased stability of SL0 in cells is likely due to favorable interactions between MePCE and LARP7 proteins, which are bound to the 5′ and 3′ ends of the 7SK RNA (Muniz et al., 2013; Eichhorn et al., 2018; Yang et al., 2019); indeed, SL0 pairing facilitates MePCE-LARP7 interactions *in vitro* (Brogie and Price, 2017). Given that SL0 is clearly present in state B (Figures 3 and S3), our data









(legend on next page)

thus support that the 7SK RNP primarily exists in a ''circular'' form in cells.

Additional structural features include the SL2, SL3, and SL4 stems, proposed in prior studies (Wassarman and Steitz, 1991; Marz et al., 2009; Brogie and Price, 2017). PAIR signals provide direct validation of the SL2 and SL3 stems (Figures 3C and 3D). PAIR analysis further reveals a long-range SL3a interaction (Figure 3C). We also performed RING analysis to search for potential tertiary interactions but did not observe compelling RING signals (Figure S4).

State B is the P-TEFb-released state with remodeled SL1 and a compact central core

State B has a dramatically remodeled structure without close literature precedent (Figures 3C and S5). Most notably, SL1 is absent. Instead, this region folds into the previously postulated SL1alt stem (Krueger et al., 2010; Brogie and Price, 2017). Although overlap with the primer binding site precludes measurement of SL1alt-specific PAIRs, per-nucleotide DMS reactivities and pairing probabilities clearly support SL1alt. P-TEFb does not bind to the isolated SL1alt hairpin (Czudnochowski et al., 2010; Fujinaga et al., 2014), and P-TEFb binding to full-length 7SK converts SL1alt to SL1 *in vitro* (Brogie and Price, 2017). Conversely, the release of P-TEFb induces conversion of SL1 to SL1alt (Krueger et al., 2010; Brogie and Price, 2017). Thus, we conclude that state B constitutes the P-TEFb-released state.

State B additionally features a major extension of SL2, which we term SL2ext, not predicted by prior analyses (Figures 3C and S5). SL2ext is directly supported by PAIRs in both in-cell and cell-free RNAs. PAIR analysis was essential for resolving these interactions: SL2ext is not predicted when structure is modeled only using per-nucleotide reactivities (Figure S3). Moderate DMS reactivities indicate that SL2ext is dynamic, and these dynamics are enhanced in cells, consistent with binding of this region by proteins (Van Herreweghe et al., 2007; Ji et al., 2013; Flynn et al., 2016). The consistency between cell-free and in-cell PAIRs leads us to conclude that SL2ext is a consensus feature of state B (Figure 3D).

RING analysis revealed a dense network of correlations in state B (Figures 3D and S4). Some of these RINGs are likely indirect and reflect unresolved minor states. Nonetheless, the consistency and density of observed RINGs suggest that state B contains a compact central core stabilized by dynamic tertiary interactions.

Combined, these data reveal that SL1alt is coupled to the formation of a distinctive central domain structure, featuring SL2ext and a compact core. This model rationalizes prior observations that in full-length 7SK, SL1alt is stabilized by high salt concentrations (Brogie and Price, 2017). Importantly, this model also implies that formation (or dissolution) of SL1 and the ability to bind to P-TEFb is allosterically coupled to the central domain structure, explaining why HEXIM1/2-P-TEFb binding induces structural changes throughout 7SK (Brogie and Price, 2017). As we discuss further below, the central (RING containing) domain is one of the primary sites bound by P-TEFb release factors, consistent with 7SK allostery playing a key role in P-TEFb regulation.

State H is a heterogenous P-TEFb-released state

State H features SL0, SL1alt, and SL2 stems but is otherwise unstructured (Figures 3A and S3). The presence of SL1alt implies that, like state B, state H does not bind P-TEFb. Based on our analysis of simulated data (Data S1), this state is likely a composite of diverse, lowly populated protein-bound structures.

Mutational analysis validates the importance of central core in 7SK structural switching

Our data suggest that 7SK contains an allosteric switch that couples remodeling of SL1 and the HEXIM1/2-P-TEFb binding site to the formation of new structures within a central core. We validated this model using mutational analysis. As an initial control, we performed DANCE-MaP experiments on *in vitro* transcripts of 7SK. This RNA folds into a two-state ensemble (Figure S6) with A and B states present at 71% ± 4% and 29% ± 4%, respectively ($R > 0.95$ between reactivity profiles; $n = 3$). The most significant differences in the *in vitro* transcript compared with cell-free RNA are the further destabilization of SL0 in state A and the lack of an M state. Because the M state contains SL0 (Figure S3), the lack of M is also likely attributable to SL0 destabilization. This SL0 destabilization may reflect increased electrostatic repulsion from the 5′-triphosphate compared with the endogenous transcript, which contains a 5′-γ-methyl-triphosphate cap (Jeronimo et al., 2007; Yang et al., 2019). 7SK may also contain a pseudouridylation modification at U250 in the SL3 stem (Zhao et al., 2016), but there were minimal differences in this region when compared with cell-free RNA. These data again emphasize that the A and B states are intrinsic features of the 7SK RNA.

To validate the B state structure, we introduced three (destabilizing) mismatches in SL1 that left the SL1alt pairing intact (mutant M1; Figures 4A and 4B). DANCE-MaP experiments revealed that M1 exclusively forms B-like states (Figures 4C and S6B; Table S1). Rescue of the M1 mutation by restoring base pairing complementarity in SL1 (M1+M2) recovered the native A:B equilibrium (78% ± 2% and 22% ± 2% populations, respectively; Figures 4C and S6B; Table S1).

We next investigated the role of SL0 in 7SK switching. Our in-cell data suggest that SL0 forms in both states, although it is

Figure 3. 7SK RNA intrinsically codes for a large-scale structural switch

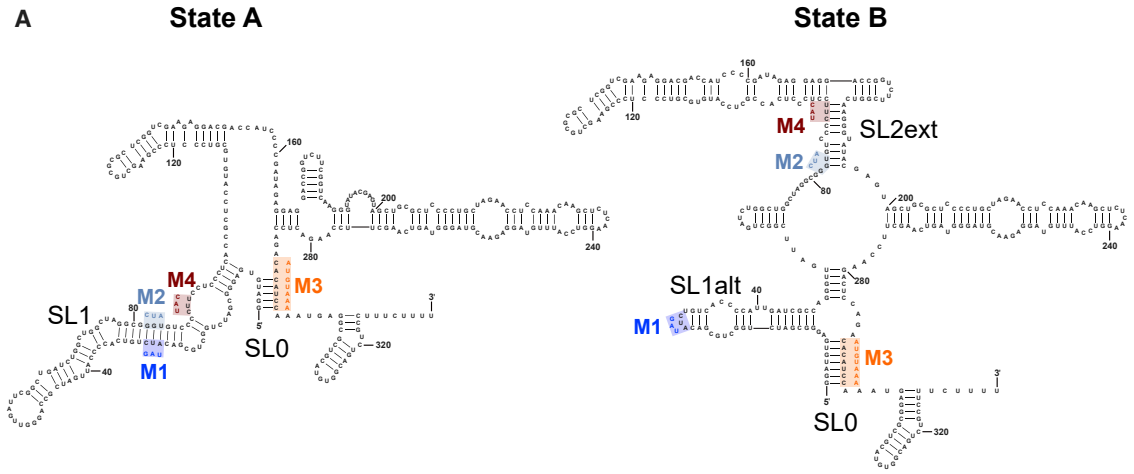
(A) Averaged and DANCE-deconvoluted reactivity profiles for 7SK RNA in cells. Major differences between states A and B are highlighted with gray shading. Population averages and standard deviations were computed over 10 replicates (Table S1).

(B) Comparison of state A and B per-nucleotide reactivities for in-cell and cell-free 7SK. Pearson's R is shown.

(C) In-cell and cell-free structural models for states A and B. Modeled base pairing probabilities (top) and directly measured PAIRs (bottom) are shown as arcs.

(D) Consensus secondary structure models for state A (left) and state B (right). Per-nucleotide reactivities are colored as per (A). RINGs observed for state B are shown with orange lines. PAIRs that uniquely support either state A or B are boxed in green and gray, respectively. PAIRs observed in both in-cell and cell-free datasets, or in only one dataset, are shown with magenta and blue lines, respectively.

See also Figures S3, S4, and S5.



B

Mutant	Structure change	Expected Result
M1	Disrupt SL1	Favor B
M1+M2	Restore SL1	Restore equilibrium
M3	Disrupt SL0	No impact
M4	Disrupt SL2alt	Favor A

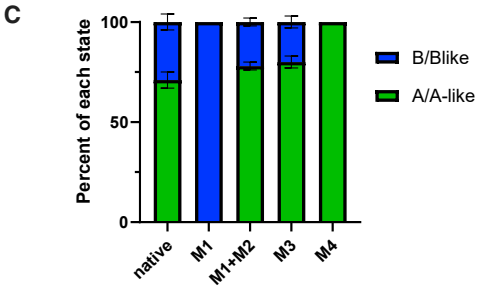


Figure 4. Mutational analyses validate 7SK states A and B

(A) Mutations shown superimposed on consensus state A and B structural models.

(B) Summary of designed structural impact for each mutant.

(C) Ensemble distribution observed for the native sequence RNA and each mutant, produced as *in vitro* transcripts. Data represent the mean and SD.

See also Figure S6.

more stable in state B. Others have proposed that SL0 dynamics drive SL1:SL1alt switching (Brogie and Price, 2017). Ablation of SL0 via the M3 mutation had minimal impact on the 7SK ensemble: 80% ± 3% of the RNA was in an A-like state and 20% ± 3% was in a B-like state that lacks SL0 (Figures 4C and S6B; Table S1). Mutation of only three of seven base pairs in SL0 gave similar results (not shown). Thus, SL0 does not drive the A:B equilibrium.

Finally, we examined the role of SL2ext in A:B switching. We designed the mutant M4 to disrupt the PAIR-supported three-helix junction at the base of SL2ext in state B without perturbing SL1 in state A (Figures 4A and 4B). Strikingly, this mutation fully shifted the ensemble to A-like states (Figures 4C and S6B; Table S1). Prior studies have also observed that mutations in the SL2ext region induce global remodeling of the 7SK structure (Brogie and Price, 2017; Luo et al., 2021), although the mechanistic basis was not resolved. Thus, even though SL2ext shows intermediate stability, this region is critical to 7SK A:B switching. These experiments validate the DANCE-resolved A and B states and establish that the central core is an energetically accessible platform for modulating 7SK structure and activity.

The 7SK structural equilibrium depends on cell state

As the 7SK-P-TEFb axis regulates transcription, we hypothesized that the 7SK ensemble may be sensitive to cell growth and transcriptional load. We used DANCE-MaP to probe the 7SK equilibrium in human RPE-1 cells, an untransformed epithelial cell line that undergoes quiescence upon contact inhibition (Leontieva et al., 2014). The same 7SK states observed in Jurkat cells are present in RPE-1 cells (Figures 5A and S7; Table S1), but in proliferating RPE-1 cells, the populations of states A and H increase by 25%. This shift in state populations in RPE-1 cells may reflect differences in growth rate or transcriptional regulation. Significantly, the population of state A, which sequesters P-TEFb, increased by 15% in quiescent RPE-1 cells relative to proliferating cells (from 47% ± 2% to 54% ± 1%; Figures 5B and S7; Table S1). Thus, quiescence, which is marked by reduced transcription and increased transcriptional pausing (Core and Adelman, 2019), is associated with a shift in the 7SK ensemble toward P-TEFb sequestration.

P-TEFb interaction with 7SK is also regulated dynamically in response to transcriptional stress (Peterlin et al., 2012; Quaresma et al., 2016). Flavopiridol is a pan-CDK inhibitor that

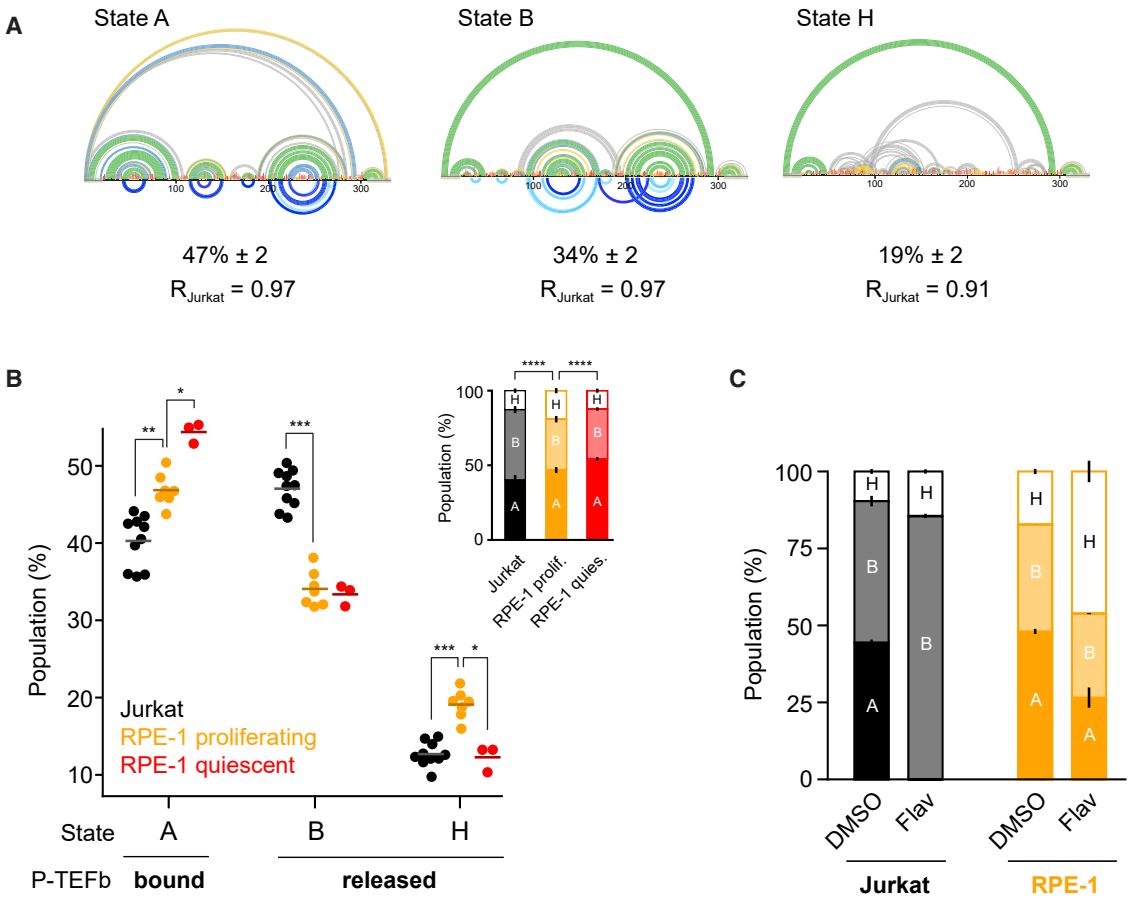


Figure 5. The 7SK equilibrium is regulated by cell state

(A) 7SK structural ensemble in proliferating RPE-1 cells resolved by DANCE-MaP. Structure models are shown as in Figure 3C. Pearson's R, comparing RPE-1 and Jurkat in-cell reactivities, are shown.

(B) 7SK ensemble populations for Jurkat cells and proliferating and quiescent RPE-1 cells. Comparisons between individual state populations evaluated using a two-sided Mann-Whitney U test. Inset, population data shown in complete distribution format. Comparisons between complete ensembles were performed using a Dirichlet likelihood ratio test (Shaw et al., 2019). n = 10, 7, and 3 for Jurkat, proliferating RPE-1, and quiescent RPE-1 cells, respectively. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$.

(C) Population data in Jurkat and RPE-1 cells treated with vehicle (DMSO, 0.01%) or 1 μM flavopiridol for 1 h. Data represent the mean and SD (n = 2).

See also Figure S7 and Table S1.

suppresses transcription by inhibiting CDK9 activity (Chao et al., 2000). To compensate for reduced CDK9 activity, cells release P-TEFb from 7SK (Biglione et al., 2007), which conventional probing experiments have indicated results in structural changes in 7SK RNA (Krueger et al., 2010; Brogie and Price, 2017). We directly visualized these structural changes by performing DANCE-MaP experiments in Jurkat and RPE-1 cells after 1 h treatment with 1 μM (saturating) flavopiridol (Biglione et al., 2007). Flavopiridol treatment induces dramatic remodeling of the 7SK structural ensemble in both cell types, with conversion of state A into states B and H (Figures 5C and S7). These data establish that the 7SK conformational equilibrium is rapidly remodeled, coincident with P-TEFb release.

Stabilization of state B induces transcription in cells

To explore the potential of targeting the 7SK ensemble as a strategy for controlling transcription, we developed an antisense oligonucleotide to selectively disrupt state A without impacting the major helices unique to state B (ASO-B; Figure 6A; see also Figure S8). ASO-B is backbone and sugar modified and thus does not mediate RNase H cleavage. Binding of ASO-B shifted the 7SK structural ensemble to exclusively B-like states, whereas a control ASO (MM-B) with five central mismatches had minimal impact (Figures 6B and 6C).

We tested the ability of ASO-B to modulate transcription in cells by monitoring the induction of *HEXIM1* mRNA expression, which specifically increases upon P-TEFb release from 7SK and constitutes a well-validated readout of P-TEFb-mediated transcription activation (He et al., 2006; Castelo-Branco et al., 2013; Liu et al., 2014; Bugai et al., 2019). As an initial control, we confirmed using HEK293T cells that *HEXIM1* mRNA increases in response to the small molecule JQ1, a bromodomain inhibitor that induces P-TEFb release via a 7SK-independent mechanism (Bartholomeeusen et al., 2012; Fujinaga et al.,

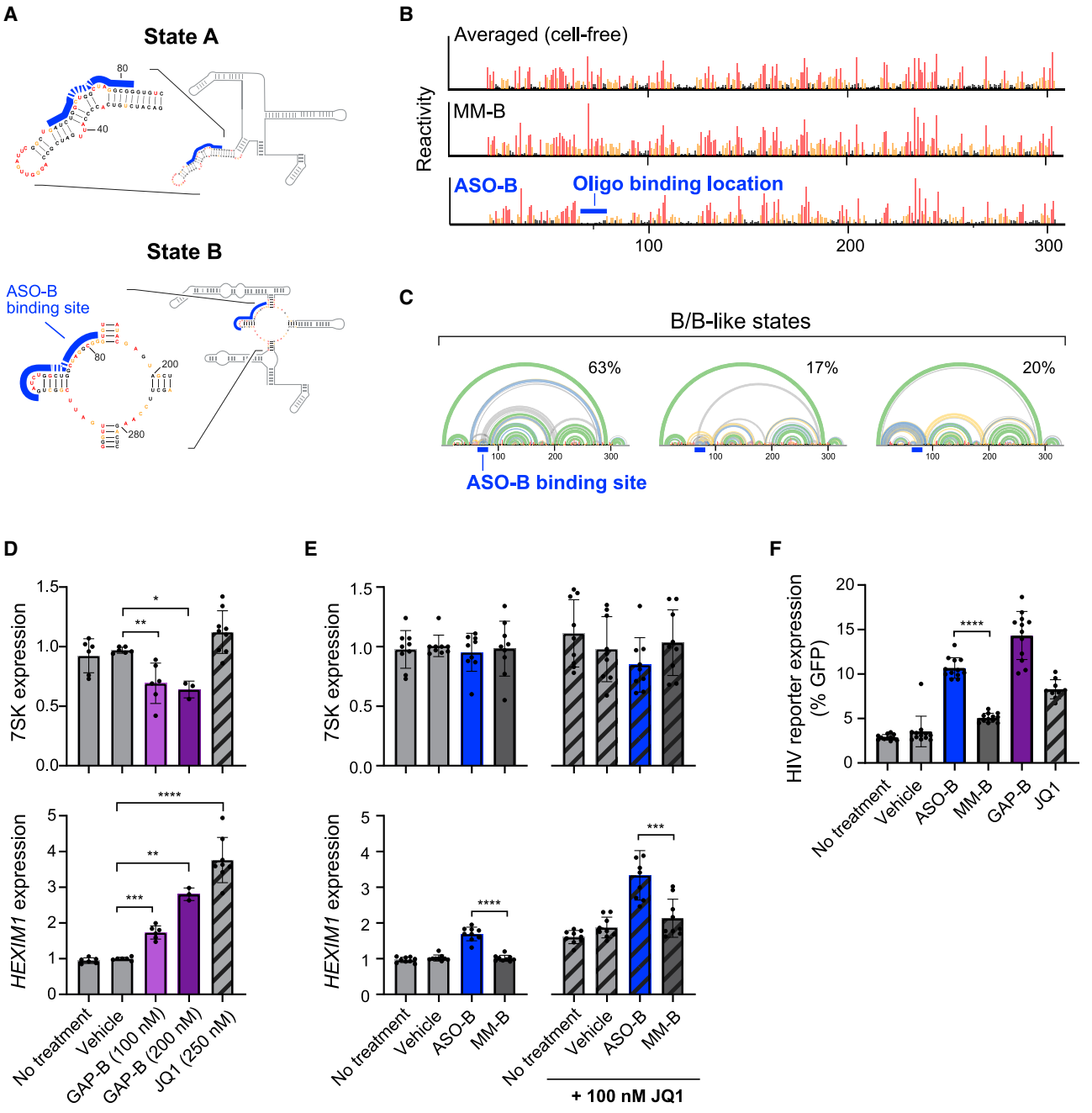


Figure 6. Stabilization of 7SK state B induces transcription

(A) ASO-B binding site shown superimposed on secondary structure models of 7SK states A and B. Mismatched region of the control MM-B oligo shown with dashed line.

(B) Per-nucleotide reactivities of cell-free 7SK RNA and cell-free RNA incubated with 100 nM MM-B or ASO-B.

(C) DANCE-MaP deconvolution of data from (B) indicates that ASO-B induces conversion of the 7SK ensemble to all B/B-like states.

(D and E) 7SK (top) and *HEXIM1* mRNA (bottom) levels measured by RT-qPCR from HEK293T cells treated for 24 h with (D) GAP-B or JQ1 or (E) 100 nM ASO-B, 100 nM MM-B, or ASO and JQ1. Data normalized to the control *TBP* mRNA. Data represent the mean and SD. n = 3 for 200 nM GAP-B, n = 9 for 250 nM JQ1, n = 6 for other measurements in (D), and n = 9 for all measurements in (E).

(F) Percent positive cells in HIV latency model 293LatF3 expressing internal GFP, treated for 24 h with 150 nM ASO-B/MM-B, 300 nM GAP-B, or 250 nM JQ1. The higher ASO-B and GAP-B concentrations give modest improvements in HIV reactivation over 100 nM ASO-B/MM-B and 200 nM GAP-B. The mismatch MM-B

(legend continued on next page)

2015) (Figure 6D). We validated ASO delivery using a gapmer ASO (GAP-B) that targets the same region of 7SK as ASO-B to induce RNase H cleavage (Figure 6D). GAP-B-mediated depletion of 7SK also releases P-TEFb and led to a 1.7-fold to 2.6-fold increase of *HEXIM1* mRNA expression. Significantly, treatment with the structure-shifting ASO-B yielded a 1.7-fold increase in *HEXIM1* expression without inducing degradation of 7SK (Figure 6E). Treatment with the mismatched MM-B control had no impact. We further examined whether ASO-B-mediated structure switching could complement transcriptional activation by JQ1. Coaddition of 100 nM JQ1 with ASO-B increased *HEXIM1* expression 3.3-fold, comparable with the upregulation observed upon treatment with 250 nM JQ1 alone (Figures 6D and 6E).

In addition to regulating transcription of endogenous genes, P-TEFb release from 7SK plays a pivotal role in HIV transcription (Herrmann and Rice, 1995). The virally encoded Tat protein recruits P-TEFb to the HIV promoter to circumvent cellular silencing mechanisms and drive efficient transcription (Mbonye and Karn, 2017). Compounds that can similarly activate HIV transcription in latently infected cells would be impactful as part of "kick-and-kill" cure strategies (Margolis et al., 2016). Latency reversal is typically studied in Jurkat-derived models, but these cells were not amenable to ASO transfection. Thus, we generated a HEK293T-derived cell line with a latent, integrated defective HIV provirus that expresses Tat and GFP upon reactivation of HIV transcription (Dobrowolski et al., 2019) (STAR Methods). The 293LatF3 reporter responds to traditional latency reversal agents, as expected (Figure S8). Treatment of 293LatF3 with ASO-B strongly activates GFP expression (2-fold relative to MM-B), comparing favorably to treatment with 250 nM JQ1 (Figure 6F). These data provide proof of principle for targeting the 7SK structural switch as a HIV latency reversal strategy. More broadly, the ability of ASO-B to induce transcription of diverse P-TEFbresponsive genes establishes a causal relationship between 7SK structural switching, P-TEFb release, and transcriptional activation.

DISCUSSION

DANCE-MaP enables complete analysis of RNA structural ensembles

Authoritatively defining RNA structural ensembles in cells and their responses to cellular stimuli has remained an unmet challenge. We developed DANCE-MaP, a single-molecule chemical probing technology that simultaneously measures reactivities at all four nucleotides, through-space base pairs (PAIRs) and tertiary interactions (RINGs) for coexisting RNA structural states in cells. DANCE-MaP further measures populations with thermodynamic precision, enabling measurement of ligand-binding affinity and of subtle but impactful structural differences between cell states.

Our studies of the adenine riboswitch and 7SK emphasize the complexity of RNA structural ensembles. The various states resolved by DANCE-MaP are distinguished by major structural landmarks but ultimately represent class averages rather than singular conformations. Within this context of underlying heterogeneity, the ability of DANCE-MaP to measure base pairing interactions directly and to estimate pairing probabilities within each state is a crucial advance. Direct PAIR measurements were essential for resolving the SL2ext structure unique to state B of 7SK. RINGs measured in the compact core of state B further emphasize that this state has a distinct higher-order structure. DANCE-MaP resolves otherwise invisible dynamic structures in cells and enables modeling of global RNA architectures with confidence.

DANCE-MaP provides many of the same measurements previously accessible only using the state-of-art NMR experiments, which have provided the ground-truth references for RNA ensembles (Liu et al., 2021). The tertiary RINGs measured in the ON state of the adenine riboswitch are of sufficient quality to guide accurate three-dimensional structure modeling (Homan et al., 2014; Li et al., 2020). The RINGs measured in 7SK state B are more challenging to interpret due to the residual dynamics of this state, but such dynamics would similarly challenge any biophysical technique. DANCE-MaP can be performed on endogenous RNAs in cells and requires modest experimental effort, paving the way for a new generation of biophysical studies in living systems.

Allostery couples 7SK P-TEFb binding domain to distal release factor binding sites

Regulated release of P-TEFb from the 7SK snRNP is a critical control point in transcription (Peterlin et al., 2012). We show that the 7SK RNA intrinsically encodes a large-scale structural switch that modulates its P-TEFb binding ability, unifying diverse prior observations of 7SK structure and regulatory function (Figure 7). Further, we show that the 7SK structural equilibrium is dynamically controlled in response to transcriptional demands. Treatment with a CDK9 inhibitor induced rapid (1 h) conversion of the P-TEFb-bound state A to states B and H, and cellular quiescence was associated with a shift of the 7SK equilibrium toward state A. These population changes correspond to the release or sequestration of tens of thousands of P-TEFb molecules per cell (Gurney and Eliceiri, 1980; Wassarman and Steitz, 1991). As the amount of free P-TEFb is roughly equivalent to the number of engaged RNA Pol II molecules in a cell (Gurney and Eliceiri, 1980; Kimura et al., 1999; Nguyen et al., 2001; Yang et al., 2001), these changes in 7SK and P-TEFb availability likely play a major role in reshaping global transcription.

Numerous proteins have been implicated as "release factors" that stimulate P-TEFb release from 7SK, a group that includes both helicases and general RNA-binding proteins (Van Herreweghe et al., 2007; Ji et al., 2013; Calo et al., 2015; Mück et al., 2016; Bugai et al., 2019; Sithole et al., 2020). These proteins bind 7SK hundreds of nucleotides away from the SL1 hairpin binding site of HEXIM1/2 and P-TEFb (Egloff et al.,

control induced a 1.4-fold increase in reporter expression relative to no ASO controls, which could be due to off-target complementarity between MM-B and HIV activation pathways (see STAR Methods). Similar results were obtained when measuring expression of a CD8a reporter from the same HIV cassette (Figure S8). n = 12 for all except JQ1 (n = 9). Significance was determined using Welch's t test (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$).
See also Figure S8.

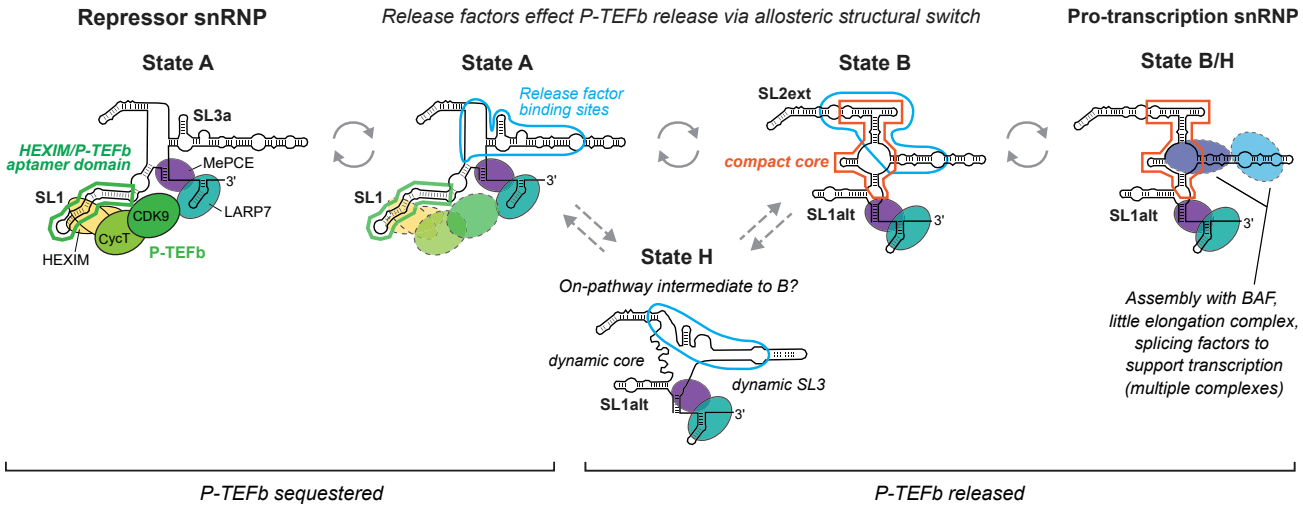


Figure 7. Model of the 7SK ensemble as a dual-function signal integrator

States A, B, and H are shown as schematic secondary structures. The compact core (red) corresponds to region of dense RINGs. Binding sites of core factors (P-TEFb, HEXIM1/2, MePCE, and LARP7), approximate regions bound by helicase and other release factors (blue outline), and postulated sites of protranscription factor binding are indicated.

2006; Krueger et al., 2008; Martinez-Zapien et al., 2016; Roder et al., 2020). HEXIM/P-TEFb release coincides with conformational changes throughout 7SK (Krueger et al., 2010; Brogie and Price, 2017). Binding of hnRNP proteins, which are potential effectors of P-TEFb release, also induce distal conformational changes in 7SK (Luo et al., 2021). However, the relationship between these conformational changes and P-TEFb release was unclear. Our data now support a model whereby release factors catalyze 7SK structural switching and thereby allosterically (via a topographically distinct site) effect P-TEFb release (Figure 7). Release factor binding sites directly overlap or are immediately adjacent to the SL2ext and compact core structures in state B, positioning them to influence structure switching. Given that unstructured regions of state H overlap these release factor binding sites, state H may represent an intermediate along the A-to-B pathway.

An RNA allostery, distance separated, mechanism of P-TEFb release offers several key mechanistic advantages: It allows 7SK to maintain a specialized "release domain" that can integrate cellular signals unencumbered by bound P-TEFb. It also prevents P-TEFb reassociation, once the structural switch is triggered. Less is known regarding how 7SK resequesters P-TEFb. We posit that helicases stimulate disassociation of hnRNPs and remodel 7SK to state A (the SL1-containing form), enabling HEXIM/P-TEFb to bind. Given that 7SK may be involved in transcription termination (Castelo-Branco et al., 2013), this process may be linked to RNA Pol II recycling. In this hypothetical model, distinct sets of proteins catalyze 7SK switching between structural states, allowing sequestering or release of P-TEFb.

The allosteric switching model also rationalizes the extreme sequence conservation of the first 100 7SK nucleotides across vertebrates and invertebrates (Yazbeck et al., 2018). This region must preserve the HEXIM/P-TEFb binding site and the ability to form the distinct SL1 and SL1alt structures (Figure 7). The core region is highly conserved among Tetrapoda, supporting its functional importance, but diverges outside of Tetrapoda, suggesting that there are multiple ways to create a P-TEFb-regulating allosteric switch. This pattern of a highly conserved P-TEFb aptamer and variable core is also observed in riboswitches, where conserved aptamer domains are often integrated with diverse expression domain architectures (Roth and Breaker, 2009).

7SK structural switch links P-TEFb release to protranscription functions

7SK is canonically considered a transcriptional repressor due to its P-TEFb sequestering function. However, the 7SK snRNP also has protranscription functions, including blocking convergent transcription via association with the BAF complex (Flynn et al., 2016) and facilitating spliceosome production (Egloff et al., 2017; Ji et al., 2021). These protranscription functions are specific to 7SK snRNPs not bound to P-TEFb. Based on our switching model, chemical probing data obtained on BAF-associated 7SK complexes (Flynn et al., 2016) can now be interpreted as corresponding to state B or H. We hypothesize that the 7SK structural switch integrates P-TEFb release with conversion of 7SK into a protranscription snRNP that scaffolds assembly of elongation-supporting factors (Figure 7, *right*). This model provides a mechanism for spatial and temporal coupling between RNA Pol II pause release and BAF-mediated inhibition of convergent transcription (Flynn et al., 2016). Our dual-function switching model also rationalizes observations that 7SK is not essential for basal P-TEFb regulation (Studniarek et al., 2021) but that 7SK depletion perturbs global chromatin structure (Prasanth et al., 2010) and compromises stress-induced transcriptional reprogramming (Studniarek et al., 2021). Most broadly, our model emphasizes how structural switching enables the 7SK snRNP to integrate diverse signals to cooperatively inactivate or activate transcription in response to cellular demand (Figure 7). RNAs are unique among biomolecules in their ability to encode large

but precise changes in structure (Breaker, 2012; Dethoff et al., 2012), making RNAs optimally suited to serve as molecular integrators. We speculate that similar switching mechanisms underlie regulatory functions of many noncoding RNAs.

7SK switch constitutes a therapeutic target for modulating transcription

Using an ASO that stabilizes state B, we showed that exogenously triggered switching of 7SK structure induces transcription of P-TEFb-sensitive targets. Development of compounds that induce P-TEFb release and activate transcription of latent HIV provirus represents a component of strategies to eradicate persistent HIV infection (Cary et al., 2016; Margolis et al., 2020). Our data thus provide proof of principle for the 7SK structural switch as a target for kick-and-kill therapies. Conversely, there is intense interest in developing pharmacological inhibitors of P-TEFb as a cancer therapeutic (Yang et al., 2020). Disruption of the 7SK/P-TEFb regulatory axis has been linked to tumorigenesis and cancer progression (Cheng et al., 2012; Ji et al., 2014; Tan et al., 2016). Intriguingly, we observed differences in the 7SK equilibrium between Jurkat (a leukemia line) and RPE-1 (an untransformed epithelial line) cells, suggesting that dysregulation of the 7SK structural equilibrium supports altered transcription in cancer cells. Small molecules or ASOs that reduce the cellular availability of P-TEFb by selectively stabilizing 7SK state A have potential for blocking transcription in cancer cells.

Limitations of the study

The DANCE-MaP technology has several limitations. Although this and other studies (Homan et al., 2014; Mustoe et al., 2019; Tomezsko et al., 2020; Luo et al., 2021; Morandi et al., 2021) indicate that multiple-hit DMS modification experiments accurately report native RNA structure, accumulated chemical damage may alter behavior of some RNAs. DANCE-MaP has a time resolution of ~5 min using DMS, but ~10-second resolution is possible with newer reagents (Ehrhardt and Weeks, 2020). DANCE-MaP can only resolve structural changes that involve >20 nucleotides with populations of at least ~5% and currently has a length limit of ~600 nucleotides across a single MaP-RT sequencing read. Reliable measurement of base pairing PAIR and tertiary RING interactions in DANCE-deconvoluted datasets ideally requires >1 million sequencing reads, making it cost prohibitive to apply on transcriptome-wide scales. DANCE-MaP requires that each sequenced DNA corresponds to a unique RNA molecule and will be more challenging to implement for low-abundance RNAs.

Our discovery of the 7SK structural equilibrium raises intriguing questions for future investigation. Our data indicate that each of the three 7SK states have residual heterogeneity; resolving this heterogeneity will be essential for fully understanding 7SK mechanisms and for therapeutic targeting efforts. Our experiments clearly indicate that state A can be converted to states B and H, but whether B can be converted to A is uncertain. Whether state H is an on-pathway intermediate between A and B, or serves some other function, is also unknown. Although we show that stabilizing state B via ASO-B induces transcription of P-TEFb-sensitive genes, further experiments are needed to confirm the ASO-B mechanism of action. We observed clear differences in the 7SK equilibrium between Jurkat and RPE-1 cells, but further experiments are needed to distinguish whether this observation reflects cell-type variation in 7SK regulation or differences in growth state. More broadly, delineating the pathways and factors that maintain and remodel the 7SK equilibrium represents a critical goal for future research.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Generation of 293Lat cell line
- METHOD DETAILS
  - DMS probing of the adenine riboswitch RNA
  - DMS probing of 7SK RNA in cells
  - DMS probing of cell-free 7SK RNA
  - DMS probing of *in vitro* transcribed 7SK RNA
  - DMS probing of contact-inhibited cells
  - DMS probing of flavopiridol treated cells
  - MaP reverse transcription
  - Sequencing library construction
  - ASO experiments
  - Gene expression analysis
  - HIV latency reversal experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - DANCE-MaP algorithm
  - Sequence alignment and data analysis
  - 7SK replicate analyses
  - Structure modeling
  - RT-qPCR Analysis
  - Fitting adenine riboswitch titration data

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2022.02.009.

ACKNOWLEDGMENTS

We thank Khoa Dhoa (Baylor College of Medicine, BCM) for exploratory analyses and Calla Olson (BCM), Kristen Karlin (BCM), and David Price (University of Iowa) for helpful discussions. Research reported in this publication was supported by the NIH (R35 GM122532 to K.M.W. and UM1AI164567 to D.M.M.), Qura Therapeutics (to D.M.M.), the Cancer Prevention & Research Institute of Texas (RR190054 to A.M.M.), and BCM seed funds (to A.M.M.). A.M.M. is a CPRIT Scholar in Cancer Research. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of Qura Therapeutics. The UNC Flow Cytometry Core Facility is supported in part by P30 CA016086 to the UNC Lineberger Comprehensive Cancer Center and the North Carolina Biotech Center Institutional Support Grant 2015-IDG-1001 for the iQue Screener.

AUTHOR CONTRIBUTIONS

K.M.W. and A.M.M. conceived the project. S.W.O., A.-M.W.T., I.S., C.A.W., D.M.M., K.M.W., and A.M.M. designed experiments. S.W.O., A.-M.W.T., J.W.A., I.S., and A.M.M. performed experiments and analysis. A.M.M. designed and wrote DanceMapper software. S.W.O., K.M.W., and A.M.M. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

A.M.M. is a consultant to and K.M.W. is an adviser to and holds equity in Ribometrix.

Received: August 18, 2021
Revised: November 29, 2021
Accepted: February 2, 2022
Published: March 22, 2022

REFERENCES

Barboric, M., Lenasi, T., Chen, H., Johansen, E.B., Guo, S., and Peterlin, B.M. (2009). 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. Proc. Natl. Acad. Sci. USA *106*, 7798–7803.

Bartholomeeusen, K., Xiang, Y., Fujinaga, K., and Peterlin, B.M. (2012). Bromodomain and extra-terminal (BET) bromodomain inhibition activate transcription via transient release of positive transcription elongation factor B (P-TEFb) from 7SK small nuclear ribonucleoprotein. J. Biol. Chem. *287*, 36609–36616.

Biglione, S., Byers, S.A., Price, J.P., Nguyen, V.T., Bensaude, O., Price, D.H., and Maury, W. (2007). Inhibition of HIV-1 replication by P-TEFb inhibitors DRB, seliciclib and flavopiridol correlates with release of free P-TEFb from the large, inactive form of the complex. Retrovirology *4*, 47.

Bishop, C. (2006). Pattern Recognition and Machine Learning (Springer).

Breaker, R.R. (2012). Riboswitches and the RNA world. Cold Spring Harb. Perspect. Biol. *4*, a003566.

Brogie, J.E., and Price, D.H. (2017). Reconstitution of a functional 7SK snRNP. Nucleic Acids Res. *45*, 6864–6880.

Bugai, A., Quaresma, A.J.C., Friedel, C.C., Lenasi, T., Düster, R., Sibley, C.R., Fujinaga, K., Kukanja, P., Hennig, T., Blasius, M., et al. (2019). P-TEFb activation by RBM7 shapes a pro-survival transcriptional response to genotoxic stress. Mol. Cell *74*, 254–267.e10.

Busan, S., and Weeks, K.M. (2018). Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. RNA *24*, 143–148.

Quaresma, A.J.C., Bugai, A., and Barboric, M. (2016). Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. Nucleic Acids Res. *44*, 7527–7539.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249–253.

Cary, D.C., Fujinaga, K., and Peterlin, B.M. (2016). Molecular mechanisms of HIV latency. J. Clin. Invest. *126*, 448–454.

Castelo-Branco, G., Amaral, P.P., Engström, P.G., Robson, S.C., Marques, S.C., Bertone, P., and Kouzarides, T. (2013). The non-coding snRNA 7SK controls transcriptional termination, poising, and bidirectionality in embryonic stem cells. Genome Biol. *14*, R98.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution – trashing old rules to forge new ones. Cell *157*, 77–94.

Chao, S.H., Fujinaga, K., Marion, J.E., Taube, R., Sausville, E.A., Senderowicz, A.M., Peterlin, B.M., and Price, D.H. (2000). Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. J. Biol. Chem. *275*, 28345–28348.

Cheng, C.Y., Kladwang, W., Yesselman, J.D., and Das, R. (2017). RNA structure inference through chemical mapping after accidental or intentional mutations. Proc. Natl. Acad. Sci. USA *114*, 9876–9881.

Cheng, Y., Jin, Z., Agarwal, R., Ma, K., Yang, J., Ibrahim, S., Olaru, A.V., David, S., Ashktorab, H., Smoot, D.T., et al. (2012). LARP7 is a potential tumor suppressor gene in gastric cancer. Lab. Invest. *92*, 1013–1019.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982.

Czudnochowski, N., Vollmuth, F., Baumann, S., Vogel-Bachmayr, K., and Geyer, M. (2010). Specificity of Hexim1 and Hexim2 complex formation with cyclin T1/T2, Importin α and 7SK snRNA. J. Mol. Biol. *395*, 28–41.

Dethoff, E.A., Boerneke, M.A., Gokhale, N.S., Muhire, B.M., Martin, D.P., Sacco, M.T., McFadden, M.J., Weinstein, J.B., Messer, W.B., Horner, S.M., et al. (2018). Pervasive tertiary structure in the dengue virus RNA genome. Proc. Natl. Acad. Sci. USA *115*, 11513–11518.

Dethoff, E.A., Chugh, J., Mustoe, A.M., and Al-Hashimi, H.M. (2012). Functional complexity and regulation through RNA dynamics. Nature *482*, 322–330.

Dobrowolski, C., Valadkhan, S., Graham, A.C., Shukla, M., Ciuffi, A., Telenti, A., and Karn, J. (2019). Entry of polarized effector cells into quiescence forces HIV latency. mBio *10*, e00337–19.

Egloff, S., Van Herreweghe, E., and Kiss, T. (2006). Regulation of polymerase II transcription by 7SK snRNA: two distinct RNA elements direct P-TEFb and HEXIM1 binding. Mol. Cell. Biol. *26*, 630–642.

Egloff, S., Vitali, P., Tellier, M., Raffel, R., Murphy, S., and Kiss, T. (2017). The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. EMBO J. *36*, 934–948.

Ehrhardt, J.E., and Weeks, K.M. (2020). Time-resolved, single-molecule, correlated chemical probing of RNA. J. Am. Chem. Soc. *142*, 18735–18740.

Eichhorn, C.D., Yang, Y., Repeta, L., and Feigon, J. (2018). Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7 [Biophysics and Computational Biology]. Proc. Natl. Acad. Sci. USA *115*, E6457–E6466.

Eilebrecht, S., Brysbaert, G., Wegert, T., Urlaub, H., Benecke, B.J., and Benecke, A. (2011). 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. Nucleic Acids Res. *39*, 2057–2072.

Flynn, R.A., Do, B.T., Rubin, A.J., Calo, E., Lee, B., Kuchelmeister, H., Rale, M., Chu, C., Kool, E.T., Wysocka, J., et al. (2016). 7SK-BAF axis controls pervasive transcription at enhancers. Nat. Struct. Mol. Biol. *23*, 231–238.

Fu, Y., Deiorio-Haggar, K., Anthony, J., and Meyer, M.M. (2013). Most RNAs regulating ribosomal protein biosynthesis in *Escherichia coli* are narrowly distributed to Gammaproteobacteria. Nucleic Acids Res. *41*, 3491–3503.

Fujinaga, K., Luo, Z., and Peterlin, B.M. (2014). Genetic analysis of the structure and function of 7SK small nuclear ribonucleoprotein (snRNP) in cells. J. Biol. Chem. *289*, 21181–21190.

Fujinaga, K., Luo, Z., Schaufele, F., and Peterlin, B.M. (2015). Visualization of positive transcription elongation factor B (P-TEFb) activation in living cells. J. Biol. Chem. *290*, 1829–1836.

Gurney, T., and Eliceiri, G.L. (1980). Intracellular distribution of low molecular weight RNA species in hela cells. J. Cell Biol. *87*, 398–403.

Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H., and Weeks, K.M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proc. Natl. Acad. Sci. USA *110*, 5498–5503.

He, N., Pezda, A.C., and Zhou, Q. (2006). Modulation of a P-TEFb functional equilibrium for the global control of cell growth and differentiation. Mol. Cell. Biol. *26*, 7068–7076.

Herrmann, C.H., and Rice, A.P. (1995). Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat cofactor. J. Virol. *69*, 1612–1620.

Homan, P.J., Favorov, O.V., Lavender, C.A., Kursun, O., Ge, X., Busan, S., Dokholyan, N.V., and Weeks, K.M. (2014). Single-molecule correlated chemical probing of RNA. Proc. Natl. Acad. Sci. USA *111*, 13858–13863.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Thérien, C., Bergeron, D., Bourassa, S., Greenblatt, J., et al. (2007). Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol. Cell *27*, 262–274.

Ji, C., Bader, J., Ramanathan, P., Hennlein, L., Meissner, F., Jablonka, S., Mann, M., Fischer, U., Sendtner, M., and Briese, M. (2021). Interaction of 7SK with the Smn complex modulates snRNP production. Nat. Commun. *12*, 1278.

Ji, X., Lu, H., Zhou, Q., and Luo, K. (2014). LARP7 suppresses P-TEFb activity to inhibit breast cancer progression and metastasis. eLife *3*, e02907.

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell *153*, 855–868.

Kalvari, I., Nawrocki, E.P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., Griffiths-Jones, S., Toffano-Nioche, C., Gautheret, D., Weinberg, Z., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res. *49*, D192–D200.

Kashanchi, F., Piras, G., Radonovich, M.F., Duvall, J.F., Fattaey, A., Chiang, C.M., Roeder, R.G., and Brady, J.N. (1994). Direct interaction of human TFIID with the HIV-1 transactivator Tat. Nature *367*, 295–299.

Kimura, H., Tao, Y., Roeder, R.G., and Cook, P.R. (1999). Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. Mol. Cell. Biol. *19*, 5383–5392.

Krokhotin, A., Mustoe, A.M., Weeks, K.M., and Dokholyan, N.V. (2017). Direct identification of base-paired RNA nucleotides by correlated chemical probing. RNA *23*, 6–13.

Krueger, B.J., Jeronimo, C., Roy, B.B., Bouchard, A., Barrandon, C., Byers, S.A., Searcey, C.E., Cooper, J.J., Bensaude, O., Cohen, E.A., et al. (2008). LARP7 is a stable component of the 7SK snRNP while P-TEFb, HEXIM1 and hnRNP A1 are reversibly associated. Nucleic Acids Res. *36*, 2219–2229.

Krueger, B.J., Varzavand, K., Cooper, J.J., and Price, D.H. (2010). The mechanism of release of P-TEFb and HEXIM1 from the 7SK snRNP by viral and cellular activators includes a conformational change in 7SK. PLoS One *5*, e12335.

Lebars, I., Martinez-Zapien, D., Durand, A., Coutant, J., Kieffer, B., and Dock-Bregeon, A.C. (2010). HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. Nucleic Acids Res. *38*, 7749–7763.

Leontieva, O.V., Demidenko, Z.N., and Blagosklonny, M.V. (2014). Contact inhibition and high cell density deactivate the mammalian target of rapamycin pathway, thus suppressing the senescence program. Proc. Natl. Acad. Sci. USA *111*, 8832–8837.

Li, B., Cao, Y., Westhof, E., and Miao, Z. (2020). Advances in RNA 3D structure modeling using experimental data. Front. Genet. *11*, 574485.

Liu, B., Shi, H., and Al-Hashimi, H.M. (2021). Developments in solution-state NMR yield broader and deeper views of the dynamic ensembles of nucleic acids. Curr. Opin. Struct. Biol. *70*, 16–25.

Liu, P., Xiang, Y., Fujinaga, K., Bartholomeeusen, K., Nilson, K.A., Price, D.H., and Peterlin, B.M. (2014). Release of positive transcription elongation factor B (P-TEFb) from 7SK small nuclear ribonucleoprotein (snRNP) activates hexamethylene bisacetamide-inducible protein (HEXIM1) transcription. J. Biol. Chem. *289*, 9918–9925.

Luo, L., Chiu, L.Y., Sugarman, A., Gupta, P., Rouskin, S., and Tolbert, B.S. (2021). HnRNP A1/A2 proteins assemble onto 7SK snRNA via context dependent interactions. J. Mol. Biol. *433*, 166885.

Margolis, D.M., Archin, N.M., Cohen, M.S., Eron, J.J., Ferrari, G., Garcia, J.V., Gay, C.L., Goonetilleke, N., Joseph, S.B., Swanstrom, R., et al. (2020). Curing HIV: seeking to target and clear persistent infection. Cell *181*, 189–206.

Margolis, D.M., Garcia, J.V., Hazuda, D.J., and Haynes, B.F. (2016). Latency reversal and viral clearance to cure HIV-1. Science *353*, aaf6517.

Martinez-Zapien, D., Legrand, P., McEwen, A.G., Proux, F., Cragnolini, T., Pasquali, S., and Dock-Bregeon, A.C. (2016). The crystal structure of the 5′ functional domain of the transcription riboregulator 7SK. Nucleic Acids Res. *45*, 3568–3579.

Marz, M., Donath, A., Verstraete, N., Nguyen, V.T., Stadler, P.F., and Bensaude, O. (2009). Evolution of 7SK RNA and its protein partners in metazoa. Mol. Biol. Evol. *26*, 2821–2830.

Matson, J.P., House, A.M., Grant, G.D., Wu, H., Perez, J., and Cook, J.G. (2019). Intrinsic checkpoint deficiency during cell cycle re-entry from quiescence. J. Cell Biol. *218*, 2169–2184.

Mbonye, U., and Karn, J. (2017). The molecular basis for human immunodeficiency virus latency. Annu. Rev. Virol. *4*, 261–285.

McLachlan, G., and Peel, D. (2000). Finite Mixture Models (Wiley).

Morandi, E., Manfredonia, I., Simon, L.M., Anselmi, F., van Hemert, M.J., Oliviero, S., and Incarnato, D. (2021). Genome-scale deconvolution of RNA structure ensembles. Nat. Methods *18*, 249–252.

Mück, F., Bracharz, S., and Marschalek, R. (2016). DDX6 transfers P-TEFb kinase to the AF4/AF4N (AFF1) super elongation complex. Am. J. Blood Res. *6*, 28–45.

Muniz, L., Egloff, S., and Kiss, T. (2013). RNA elements directing in vivo assembly of the 7SK/MePCE/Larp7 transcriptional regulatory snRNP. Nucleic Acids Res. *41*, 4686–4698.

Mustoe, A.M., Lama, N.N., Irving, P.S., Olson, S.W., and Weeks, K.M. (2019). RNA base-pairing complexity in living cells visualized by correlated chemical probing. Proc. Natl. Acad. Sci. USA *116*, 24574–24582.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SKsmall nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322–325.

Olson, C.M., Jiang, B., Erb, M.A., Liang, Y., Doctor, Z.M., Zhang, Z., Zhang, T., Kwiatkowski, N., Boukhali, M., Green, J.L., et al. (2018). Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. Nat. Chem. Biol. *14*, 163–170.

Peterlin, B.M., Bragie, J.E., and Price, D.H. (2012). 7SK snRNA: a noncoding RNA that plays a major role in regulating eukaryotic transcription. Wiley Interdiscip. Rev. RNA *3*, 92–103.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.

Prasanth, K.V., Camiolo, M., Chan, G., Tripathi, V., Denis, L., Nakamura, T., Hübner, M.R., and Spector, D.L. (2010). Nuclear organization and dynamics of 7SK RNA in regulating gene expression. Mol. Biol. Cell *21*, 4184–4196.

Raha, T., Cheng, S.W.G., and Green, M.R. (2005). HIV-1 Tat stimulates transcription complex assembly through recruitment of TBP in the absence of TAFs. PLoS Biol. *3*, e44.

Ray, P.S., Jia, J., Yao, P., Majumder, M., Hatzoglou, M., and Fox, P.L. (2009). A stress-responsive RNA switch regulates VEGFA expression. Nature *457*, 915–919.

Reining, A., Nozinovic, S., Schlepckow, K., Buhr, F., Fürtig, B., and Schwalbe, H. (2013). Three-state mechanism couples ligand and temperature sensing in riboswitches. Nature *499*, 355–359.

Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: web servers for RNA secondary structure prediction and analysis. BMC Bioinformatics *11*, 129.

Rivas, E., Clements, J., and Eddy, S.R. (2017). A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat. Methods *14*, 45–48.

Roder, K., Stirnemann, G., Wales, J., and Pasquali, S. (2020). Structural transitions in the RNA 7SK 5 hairpin and their effect on HEXIM binding. Nucleic Acids Res. *48*, 370–389.

Rodnina, M.V., Fischer, N., Maracci, C., and Stark, H. (2017). Ribosome dynamics during decoding. Philos. Trans. R. Soc. Lond. B Biol. Sci. *372*, 20160182.

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

Sengupta, A., Rice, G.M., and Weeks, K.M. (2019). Single-molecule correlated chemical probing reveals large-scale structural communication in the ribosome and the mechanism of the antibiotic spectinomycin in living cells. PLoS Biol. *17*, e3000393.

Sharp, P.A. (2009). The centrality of RNA. Cell *136*, 577–580.

Shaw, L.M., Blanchard, A., Chen, Q., An, X., Davies, P., Tötemeyer, S., Zhu, Y.G., and Stekel, D.J. (2019). DirtyGenes: testing for significant changes in gene or bacterial population compositions from a small number of samples. Sci. Rep. *9*, 2373.

Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A., and a Weeks, K.M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat. Methods *11*, 959–965.

Sithole, N., Williams, C.A., Abbink, T.E.M., and Lever, A.M.L. (2020). DDX5 potentiates HIV-1 transcription as a co-factor of Tat. Retrovirology *17*, 6.

Smola, M.J., Calabrese, J.M., and Weeks, K.M. (2015b). Detection of RNA-protein interactions in living cells with SHAPE. Biochemistry *54*, 6867–6875.

Smola, M.J., Rice, G.M., Busan, S., Siegfried, N.A., and Weeks, K.M. (2015a). Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. Nat. Protoc. *10*, 1643–1669.

Studniarek, C., Tellier, M., Martin, P.G.P., Murphy, S., Kiss, T., and Egloff, S. (2021). The 7SK/P-TEFb snRNP controls ultraviolet radiation-induced transcriptional reprogramming. Cell Rep. *35*, 108965.

Tan, J.L., Fogley, R.D., Flynn, R.A., Ablain, J., Yang, S., Saint-André, V., Fan, Z.P., Do, B.T., Laga, A.C., Fujinaga, K., et al. (2016). Stress from nucleotide depletion activates the transcriptional regulator HEXIM1 to suppress melanoma. Mol. Cell *62*, 34–46.

Tian, S., Kladwang, W., and Das, R. (2018). Allosteric mechanism of the V. vulnificus Adenine riboswitch resolved by four-dimensional chemical mapping. eLife *7*, 1–36.

Tomezsko, P.J., Corbin, V.D.A., Gupta, P., Swaminathan, H., Glasgow, M., Persad, S., Edwards, M.D., Mcintosh, L., Papenfuss, A.T., Emery, A., et al. (2020). Determination of RNA structural diversity and its role in HIV-1 RNA splicing. Nature *582*, 438–442.

Van Herreweghe, E., Egloff, S., Goiffon, I., Jády, B.E., Froment, C., Monsarrat, B., and Kiss, T. (2007). Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. EMBO J. *26*, 3570–3580.

Wang, P.Y., Sexton, A.N., Culligan, W.J., and Simon, M.D. (2019). Carbodiimide reagents for the chemical probing of RNA structure in cells. RNA *25*, 135–146.

Warhaut, S., Mertinkus, K.R., Höllthaler, P., Fürtig, B., Heilemann, M., Hengesbach, M., and Schwalbe, H. (2017). Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. Nucleic Acids Res. *45*, 5512–5522.

Wassarman, D.A., and Steitz, J.A. (1991). Structural analyses of the 7SK ribonucleoprotein (RNP), the most abundant human small RNP of unknown function. Mol. Cell. Biol. *11*, 3432–3445.

Weeks, K.M. (2021). SHAPE directed discovery of new functions in large RNAs. Acc. Chem. Res. *54*, 2502–2517.

Wilkinson, M.E., Charenton, C., and Nagai, K. (2020). RNA splicing by the spliceosome. Annu. Rev. Biochem. *89*, 359–388.

Yang, W., Wang, Y., Lai, A., Clark, C.G., Corte, J.R., Fang, T., Gilligan, P.J., Jeon, Y., Pabbisetty, K.B., Rampulla, R.A., et al. (2020). Discovery of a high affinity, orally bioavailable macrocyclic FXIa inhibitor with antithrombotic activity in preclinical species. J. Med. Chem. *63*, 7226–7242.

Yang, Y., Eichhorn, C.D., Wang, Y., Cascio, D., and Feigon, J. (2019). Structural basis of 7SK RNA 5′-γ-phosphate methylation and retention by MePCE. Nat. Chem. Biol. *15*, 132–140.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317–322.

Yazbeck, A.M., Tout, K.R., and Stadler, P.F. (2018). Detailed secondary structure models of invertebrate 7SK RNAs. RNA Biol. *15*, 158–164.

Zhang, J., and Ferré-D'Amaré, A.R. (2014). Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. Structure *22*, 1363–1371.

Zhao, Y., Karijolich, J., Glaunsinger, B., and Zhou, Q. (2016). Pseudouridylation of 7SK snRNA promotes 7SK snRNP formation to suppress HIV-1 transcription and escape from latency. EMBO Rep. *17*, 1441–1451.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
RPMI 1640	Gibco	Cat# 21875034
Fetal Bovine Serum (FBS; used for Jurkat and HEK293T cells)	Millipore	Cat# TMS-013-B
DMEM/F-12+HEPES	Gibco	Cat# 12430112
Pen/Strep	Gibco	Cat# 15140122
Sodium pyruvate	Gibco	Cat# 11360070
Fetal Bovine Serum (FBS; used for RPE-1 cells)	Gibco	Cat# 10437036
MEM non-essential amino acids	Gibco	Cat# 11140050
T7 polymerase	K. Weeks Lab	N/A
TURBO DNase	Invitrogen	Cat# AM2238
TRIzol reagent	Invitrogen	Cat# 15596018
Bicine	Sigma-Aldrich	Cat# B3876
Superscript II Reverse transcriptase	Invitrogen	Cat# 18064014
Dimethyl sulfate (DMS)	Thermo Scientific	Cat# 430831000
2-mercaptoethanol (BME)	Sigma-Aldrich	Cat# M6250
Q5 HotStart polymerase	NEB	Cat# M0493L
RNasin	Promega	Cat# N2511
Yeast inorganic pyrophospatase	NEB	Cat# M2403L
Flavopiridol	Selleckchem	Cat# L86-8275
JQ1	MedChemExpress	Cat# HY-13030
HMBA	Sigma	Cat# H4663
TNFα	R&D Systems	Cat# 210-TA-100
TransIT-oligo	Mirus Bio	Cat# MIR 2164
anti-CD8a-PacBlue, clone 53-6.7	BioLegend	Cat# 100725 (RRID:AB_493425)
Kapa Hot Start HiFi Polymerase	Roche	Cat# 7958935001
Paraformaldehyde	Electron Microscopy Sciences	Cat# 15710
5-ethynyl-2′-deoxyuridine (EdU)	Millipore-Sigma	Cat# 900584
5-FAM azide	LumiProbe	Cat# C4130
DAPI	Millipore-Sigma	Cat# MBD0015
Critical commercial assays		
PureLink PCR column	Invitrogen	Cat# K310001
Qubit RNA BR assay	Invitrogen	Cat# Q10210
Agencourt RNAclean XP beads	Beckman Coulter	Cat# A63987
RNeasy mini	Qiagen	Cat# 74004
MagBind TotalPure NGS beads	Omega BioTek	Cat# M1378-01
Bioanalyzer (HS DNA)	Agilent	Cat# 5067-4626
G-50 Sephadex	Cytiva	Cat# 27534001
Illumina MiSeq v2 chemistry (300-cycle)	Illumina	Cat# MS-102-2002
Illumina MiSeq v2 chemistry (500-cycle)	Illumina	Cat# MS-102-2003
Illumina MiSeq v3 chemistry (600-cycle)	Illumina	Cat# MS-102-3003
Quick-RNA 96-well	Zymo	Cat# R1052
Maxima First Strand cDNA Synthesis kit for RT-qPCR	ThermoFisher	Cat# K1641

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FastStart Universal SYBR Green Master	Roche	Cat# 4913850001
PrestoBlue	Life Technologies	Cat# A13261
Mission Lentiviral Packaging Mix	Millipore-Sigma	Cat# SHP001
EasySep Mouse CD8a Positive Selection Kit	StemCell Technologies	Cat# 18953
Deposited data		
Adenine riboswitch DMS-MaP data	This study	GSE182552
7SK DMS-MaP data	This study	http://www.ncbi.nlm.nih.gov/bioproject/741330
Experimental models: Cell lines		
Human: Jurkat E6-1 (blood T-lymphoblast)	ATCC	TIB-152
Human: hTERT RPE-1 (retina epithelial)	Gift from W. Marzluff	CRL-4000
Human: HEK293T/17 (kidney epithelial)	ATCC	CRL-11268
Human: 293LatF3	This study	N/A
Oligonucleotides		
primers	Table S2	N/A
g-block templates	Table S2	N/A
Antisense oligonucleotides	Table S2	N/A
Recombinant DNA		
pHR-H13LTat-CD8a/d2eGFP-IRES-Nef	Dobrowolski et al. (2019); Addgene	Plasmid # 126552
Software and algorithms		
Prism 9	Gaphpad Software	https://www.graphpad.com/scrientific-sofware/prism/
ShapeMapper (v2.1.5)	Busan and Weeks (2018)	https://weekslab.com/software/
DanceMapper (v1.0)	This study	https://github.com/MustoeLab/DanceMapper https://doi.org/10.5281/zenodo.5847984
RingMapper (v1.1)	Mustoe et al. (2019)	https://github.com/Weeks-UNC/RingMapper
arcPlot		https://github.com/Weeks-UNC/arcPlot
RNAstructure (v6.2)	Reuter and Mathews (2010)	https://rna.urmc.rochester.edu/RNAstructure.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents may be directed to and will be fulfilled by the lead contact, Kevin Weeks (weeks@unc.edu).

Materials availability

Commercially available reagents are listed in the key resources table. 293LatF3 cells are available from the lead contact and A.W.T. upon reasonable request.

Data and code availability

- Adenine riboswitch probing data have been deposited at the GEO (GSE182552), and 7SK probing data have been deposited at the NCBI BioProject database (http://www.ncbi.nlm.nih.gov/bioproject/741330).
- The *DanceMapper* pipeline, along with scripts for performing structure modeling and visualizing DANCE solutions, is available at https://github.com/MustoeLab/DanceMapper and has also been deposited at Zenodo (DOI listed in key resources table).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

Jurkat E6-1 cells were obtained from ATCC (TIB-152) and cultured in suspension using RMPI 1640 media (Gibco) supplemented with 10% FBS (Millipore), 100 U/mL Pen/Strep (Gibco) at 37 °C and 5% CO_2. hTERT RPE-1 (RPE-1) cells were a gift from W. Marzluff (UNC) and were authenticated by STR profiling and confirmed to be free of mycoplasma contamination. RPE-1 cells were maintained in DMEM/F-12 + HEPES (Gibco) with 10% FBS (Gibco), 100 U/mL Pen/Strep (Gibco), 2 mM sodium pyruvate (Gibco), and MEM non-essential amino acids (Gibco) at 37 °C and 5% CO_2. HEK293T/17 cells were obtained from ATCC (CRL-11268) and maintained in DMEM (Gibco) supplemented with 10% FBS (Millipore) and 100 U/mL Pen/Strep at 37 °C and 5% CO_2.

Generation of 293Lat cell line

The pHR-H13LTat-CD8a/d2eGFP-IRES-Nef plasmid encoding a replication incompetent HIV and a polycistronic transcript expressing mouse CD8a, GFP, and Nef was obtained as a gift from Jonathan Karn (Addgene plasmid #126552). The plasmid was modified to add a T2A motif to decouple CD8a and GFP expression using incomplete primer extension (primers listed in Table S2; Integrated DNA technologies) and Kapa Hot Start HiFi Polymerase (Roche), generating pHR-H13LTat-CD8a-T2A-d2eGFP-IRES-Nef. Successful addition of the T2A was confirmed by sanger sequencing (GeneWiz). Lentiviral particles containing pHR-H13LTat-CD8a-T2A-d2eGFP-IRES-Nef were generated using the Mission Lentiviral Packaging Mix (Millipore-Sigma) and HEK293T/17 cells were transduced at an MOI of 0.1. After 3 days, cells positive for mouse CD8a were isolated using the EasySep Mouse CD8a Positive Selection Kit (StemCell Technologies). Isolated cells were subject to limiting dilution to generate monoclonal lines. Successful monoclonal lines were screened for low basal GFP expression and ability to be reactivated by TNFα by flow cytometry. Clone 293LatF3 was expanded and tested based on these criteria.

METHOD DETAILS

DMS probing of the adenine riboswitch RNA

Native sequence and mutant *Vibrio vulnificus add* adenine riboswitches containing 5' and 3' structure cassettes were transcribed *in vitro* as described (Mustoe et al., 2019). Briefly, templates were synthesized as gBlocks [IDT; (Mustoe et al., 2019) and Table S2], amplified by PCR (Q5 DNA polymerase, NEB), and purified (PureLink PCR column, Invitrogen). RNA was transcribed *in vitro* [400 μL; 40 mM Tris (pH 8.0), 25 mM $MgCl_2$, 2.5 mM Spermidine, 0.01% (vol/vol) Triton X-100, 10 mM DTT, 5 mM each NTP, ~4 μg DNA template, 0.05 mg/mL T7 RNA polymerase (lab made), 0.2 U pyrophosphatase (NEB); 37 °C; 4h], treated with DNase (TURBO DNase, Invitrogen), purified (Agencourt RNAclean XP beads; Beckman Coulter), and stored at -20 °C. RNA size and purity was confirmed using Bioanalyzer analysis (Agilent) and concentration was quantified (Qubit RNA BR assay, Invitrogen).

For probing experiments, RNA [4 pmol in 2 μL volume] was denatured at 95 °C for 2 min followed by snap cooling on ice for 2 min. RNA was folded by adding 7 uL of 1.43× adenine-containing folding buffer [1× buffer: 300 mM bicine (pH 8.0), 100 mM NaCl, 5 mM $MgCl_2$, variable adenine] and incubated at 30 °C for 30 min. Folded RNA was added to 1 μL of DMS solution (1.7 M in ethanol), allowed to react for 10 min at 30 °C, and then quenched via addition of an equal volume of 20% 2-mercaptoethanol (vol/vol in H_2O) and placed on ice. RNA was purified by precipitation with ethanol. No-reagent control RNA was prepared identically, substituting neat ethanol for the DMS solution.

DMS probing of 7SK RNA in cells

For Jurkat cells, cells were pelleted, washed with PBS, and counted. 1-2×10^6 cells were resuspended in 450 μL fresh media supplemented with 200 mM Bicine (pH 8.0). Cells were then treated with 50 μL of 1.7 M DMS in ethanol or 50 μL ethanol for 6 min at 37 °C. Reactions were quenched with 500 μL 20% 2-mercaptoethanol and placed on ice. Cells were pelleted and RNA extracted using 1 mL TRIzol reagent (Invitrogen). Residual DNA was removed by treating with 2 units of TURBO DNase (Ambion) for 30 min at 37 °C, followed by spike-in of 2 additional units and further 30 min incubation (1 hour total). RNA was purified by SPRI beads (MagBind TotalPure NGS beads; Omega BioTek) and quantified by UV absorbance (Nanodrop).

For RPE-1 cells, 1.5×10^6 cells were seeded into a 10 cm dish 48 h prior to probing. Media was removed and 5.4 mL fresh media, supplemented with 200 mM Bicine (pH 8.0), was added and incubated at 37 °C for 3 min. Cells were treated with 600 μL of 1.7 M DMS or neat ethanol for 6 min at 37 °C, followed by quenching using 6 mL of 20% 2-mercaptoethanol on ice. Cells were scraped and pelleted, RNA was extracted using TRIzol (as described for Jurkat cells) or column (RNeasy mini; Qiagen), and quantified by UV absorbance (Nanodrop).

DMS probing of cell-free 7SK RNA

Total RNA was extracted from 2×10^6 Jurkat cells using TRIzol reagent (Invitrogen). RNA was DNase treated, purified (Mag-Bind TotalPure NGS beads; 1.8× ratio), and quantified as described above for in-cell RNA. 2 μg RNA in 50 μL in water was denatured at 98 °C for 1 min, snap cooled at 4 °C for 1 min, and then refolded via addition of 50 μL of 2× Bicine RNA folding buffer and incubation at 37 °C for 20 minutes [1× folding buffer: 200 mM Bicine (pH 8.0), 200 mM potassium acetate (pH 8.0) and 5 mM $MgCl_2$] (Mustoe et al., 2019). Samples were split into two 45 μL aliquots and treated with either 5 μL 1.7 M DMS in ethanol or 5 μL neat ethanol at 37 °C for 6

minutes. Following treatment, samples were quenched with 1 volume of 20% 2-mercaptoethanol, placed on ice, and purified by isopropanol precipitation.

DMS probing of *in vitro* transcribed 7SK RNA

DNA templates were synthesized as gBlocks (Integrated DNA technologies; Table S2) and amplified by PCR [Q5 HotStart polymerase (NEB), supplemented with 1.0 M betaine]. DNA templates were purified (Mag-Bind TotalPure NGS beads; 0.7× ratio). RNA was transcribed *in vitro* [400 μL; 40 mM Tris (pH 8.0), 25 mM $MgCl_2$, 2.5 mM spermidine, 0.01% (vol/vol) Triton X-100, 10 mM DTT, 5 mM each NTP, 200 ng DNA template, 95 μg T7 RNA polymerase (lab made), 20 U RNasin (Promega), 50 U yeast inorganic pyrophosphatase (NEB); 37 °C; 4h]. Transcription reactions were treated with 16 U TURBO DNase (Thermo) for 30 min at 37 °C and purified (Mag-Bind TotalPure NGS beads; 1.8× bead:volume ratio) and stored at -20 °C. RNA size and purity were confirmed using Bioanalyzer analysis (Agilent) and concentration was quantified (Qubit RNA BR assay, Invitrogen).

For probing experiments, RNA [10 μg in 50 μL] was denatured at 95 °C for 2 min followed by snap cooling on ice for 2 min. 50 uL of 2× folding buffer was then added and the RNA folded at 37 °C for 30 min [1× buffer: 200 mM Bicine (pH 8.0), 200 mM potassium acetate (pH 8.0) and 5 mM $MgCl_2$]. 45 μL of folded RNA was added to 5 μL of DMS solution (1.7 M in ethanol), allowed to react for 6 min at 37 °C, quenched via addition of an equal volume of 20% 2-mercaptoethanol, and placed on ice. RNA was purified by precipitation with isopropanol. No-reagent control RNA was prepared identically, substituting neat ethanol for the DMS solution.

DMS probing of contact-inhibited cells

RPE-1 cells were growth-arrested by contact inhibition as described (Matson et al., 2019). For each replicate experiment, three 10 cm dishes were seeded with 6×10^6 RPE-1 cells and grown to 100% confluency. Medium was then exchanged and cells were incubated for an additional 96 h to allow for complete growth arrest. Two dishes of arrested cells were modified with DMS or ethanol as described for proliferating RPE-1 cells. The third dish was used to confirm growth arrest by flow cytometry analysis. Cells were incubated with 10 μM 5-ethynyl-2′-deoxyuridine (EdU) for 1 h prior to harvesting. Cells were then harvested with trypsin and fixed in 4% formaldehyde. Cells were labeled with FAM-azide and nuclei were stained with DAPI. Less than 1% of cells were in S phase (positive for FAM), and 90% of cells were in G1/G0 (2*n* DNA content measured by DAPI).

DMS probing of flavopiridol treated cells

Jurkat cells (3 million cells in 10 mL fresh media) were seeded 23 h prior to treatment and were then treated with either vehicle (0.01% DMSO) or with 1 μM flavopiridol (in DMSO) for 1 h. RPE-1 cells were seeded 23 h prior to be 70% confluent on the day of experiment and were treated with 0.01% DMSO or 1 μM flavopiridol for 1 h. Cells were then treated with DMS and RNA was extracted identically as described above for in-cell probing experiments.

MaP reverse transcription

Mutational profiling (MaP) reverse transcription (RT) was performed as described (Mustoe et al., 2019; Sengupta et al., 2019). For adenine riboswitch experiments, one-half of the purified DMS reaction was input into RT. For in-cell and cell-free 7SK experiments, 1 μg total cellular RNA was input into RT. For *in vitro* 7SK experiments, 100 ng RNA was input into RT. RT products were purified by beads (Mag-Bind TotalPure NGS beads; 1.8× ratio) or column (G-50 Sephadex; Cytiva).

Sequencing library construction

Sequencing libraries were generated using the two-step PCR approach (Smola et al., 2015a).

For the adenine riboswitch, one-seventeenth of the purified RT reaction was input to PCR1, performed [98 °C for 30 s, 10 cycles of (98 °C for 8 s, 66 °C for 20 s, 72 °C for 20 s), and 72 °C for 2 min]. PCR1 product was purified (Mag-Bind TotalPure NGS beads; 0.8× ratio). 2.5 ng product was input to PCR2 [98 °C for 30 s, 10 cycles of (98 °C for 8 s, 68 °C for 20 s, 72 °C for 20 s), and 72 °C for 2 min]. PCR2 product was purified (Mag-Bind TotalPure NGS beads; 0.8× ratio), and sequenced with an Illumina MiSeq instrument using 2×150 paired-end sequencing (v2 chemistry). Data used for high-depth PAIR-MaP/RING-MaP analysis in Figures 2 and S2 were obtained by resequencing libraries from a previously published experiment (Mustoe et al., 2019).

For 7SK, one-fifth of the purified RT reaction was input to PCR1 [98 °C for 30 s, 10 cycles of (98 °C for 10 s, 68 °C for 20 s, 72 °C for 20 s), and 72 °C for 2 min]. PCR1 product was purified (Mag-Bind TotalPure NGS beads; 0.8× ratio). 1-2 ng product was input to PCR2 [98 °C for 30 s, 10-14 cycles of (98 °C for 10 s, 65 °C for 30 s, 72 °C for 20 s), and 72 °C for 2 min]. PCR2 product was purified (Mag-Bind TotalPure NGS beads; 0.8× ratio) and sequenced with an Illumina MiSeq instrument using 2×250 (v2 chemistry) or 2×300 (v3 chemistry) paired-end sequencing.

ASO experiments

The ASO-B antisense oligonucleotide was designed to bind to nucleotides 64-82 of 7SK to stabilize state B and contained complete 2′-O-methyl modifications to render it insensitive to RNase H. The mismatch MM-B ASO contains 5 central mismatches to reduce binding affinity. The positive control gapmer ASO (GAP-B) targets the 64-78 region but lacks central 2′-O-methylation and hence targets 7SK for RNase H degradation. ASOs were synthesized (IDT) with phosphorothioate backbones for added stability in cells (Table S2).

ASO-B and other ASO designs targeting the same general region (complementary to nucleotides 49-64, 52-67, 59-71, 73-83, 78-89, 80-95, 84-93, and 85-96; see Figure S8A) were evaluated for their ability to engage with 7SK and stabilize state B using DANCE-MaP experiments. 4 μg total RNA from Jurkat cells in 100 μL H_2O was denatured at 98 °C for 1 min, snap cooled on 4 °C for 1 min, and then folded via addition of 100 μL of 2× folding buffer [1×: 200 mM Bicine (pH 8.0), 200 mM potassium acetate (pH 8.0) and 5 mM $MgCl_2$] and incubated at 37 °C for 15 min. 99 μL folded RNA was then added to 1 μL of 100 μM ASO and incubated for an additional 15 min at 37 °C. Samples were then split in to two 45 μL samples and treated with DMS or ethanol as described for cell-free experiments. ASO-B was identified as the design that most efficiently stabilized state B without disrupting global structure.

HEK293T cells were seeded at 25,000 cells/well in a 96-well flat bottom plate 18 h prior to transfection. 100 nM of an ASO or gapmer were transfected using TransIT-Oligo (Mirus Bio). For JQ1 combination experiments, 100 nM JQ1 was added 4 h post transfection for a total incubation time of 20 h. 250 nM JQ1 was added for 24 h. After 24 h, cells were lysed in lysis buffer (Quick RNA 96-well RNA kit; Zymo) and RNA was either immediately isolated or lysed samples were flash frozen and stored at -80 °C for no longer than 48 h prior to RNA isolation.

Gene expression analysis

Total RNA was isolated (Quick RNA 96-well; Zymo) and cDNA was generated (Maxima First Strand cDNA Synthesis Kit for RT-qPCR; with dsDNase, Thermo Fisher). Gene expression was assayed by RT-qPCR (using FastStart Universal SYBR Green Master; Roche) on an QuantStudio 5 instrument (Applied Biosystems). Primer sets are listed in Table S2. Primer efficiency for all targets was quantified for each run using a standard curve derived from a DNA gene fragment (gBlock; Integrated DNA Technologies; Table S2) designed to mimic the target amplicon.

HIV latency reversal experiments

293LatF3 cells were plated at 25,000 cells/well in a 96-well flat bottom plate 18 h prior to drug/ASO treatment. Response to canonical latency reversal agents was assessed by treating cells with JQ1, TNFα, or HMBA (DMSO stock solutions, max final DMSO concentration of 0.1%). Transfections of ASO-B, MM-B, and GAP-B ASOs were performed using TransIT-Oligo Reagent (Mirus Bio). 24 h after drug/ASO treatment, cells were treated with PrestoBlue (Life Technologies) for 30 min and fluorescence measured (555 nm excitation, 585 nm emission, 570 nm cutoff; SpectraMax M3, Molecular Devices). PrestoBlue was removed, cells washed, stained with anti-CD8a-PacBlue (Biolegend), and fixed in 1.5% paraformaldehyde (Electron Microscopy Sciences). GFP and cell surface CD8a levels were assayed using the iQue Screener Plus (Intellicyt) and data processed using ForeCyt analysis software (Intellicyt). Similar results were observed for replicate experiments performed on a second 293Lat clone (not shown).

MM-B demonstrated a minor 1.4-fold increase in GFP expression (Figure 6F) but failed to induce *HEXIM1* mRNA expression (Figure 6E), suggesting a potential off-target hit. NCBI BLAST of MM-B against the human G+T database identified 89% query coverage (nt 2-18 complementary with 1 mismatch) to the TAF10 mRNA, a TATA-binding protein associated factor that modulates activity of transcription factor IID (TFIID). TFIID is known to be recruited by Tat (Kashanchi et al., 1994; Raha et al., 2005) and thus potential off-target impacts of MM-B in this pathway could result in HIV activation but would be unlikely to activate *HEXIM1*.

QUANTIFICATION AND STATISTICAL ANALYSIS

DANCE-MaP algorithm

Data pre-processing

Aligned DMS-MaP sequencing reads are processed into two binary vectors. The mutation string, $\boldsymbol{x_n} = (x_{n,1}, \ldots, x_{n,D})$, encodes whether a match (0) or a mutation (1) is observed at each nucleotide, where n indexes an individual read and D is the length of the amplicon. The data string, $\boldsymbol{\delta_n} = (\delta_{n,1}, \ldots, \delta_{n,D})$, encodes whether the nucleotide was measured (1) or not (0); example "no measurement" scenarios include low quality score, a deletion, or masking due to minimum spacing requirements between mutations (Busan and Weeks, 2018). Reads with fewer than 75% of positions defined, $\sum_i^D \delta_{n,i} < 0.75 \cdot D$, are discarded. Positions with mutation rates >0.02 in the ethanol-treated control sample $(r_{etoh,i})$, or with average mutation rates less than 0.0001 in the DMS-treated sample $(r_{DMS,i})$, are considered invalid and excluded from further analysis ($\delta_{n,i} = 0$ for all n).

To prevent individual nucleotides from dominating the maximum likelihood (ML) clustering outcome and to improve clustering power, nucleotide positions are split into "active" and "inactive" categories. Active nucleotides are included in likelihood calculations during primary ML clustering, whereas inactive nucleotides are excluded and solved via a second constrained ML-optimization (Figure S1). Active versus inactive status is specified using the binary vector $\phi = (\phi_1, \ldots, \phi_D)$, where $\phi_i = 1$ denotes active status. Nucleotides where $(r_{DMS,i} - r_{etoh,i}) < 0.002$ are set to inactive upon model initialization. As described below, additional nucleotides may be inactivated over the course of model solution.

Primary ML clustering

Data are fit to a Bernoulli mixture model using the expectation-maximization (EM) algorithm (Bishop, 2006). Fitting is performed for sequentially larger numbers of model components (structural states), beginning with 1, until the best fit is identified (Figure S1).

A model consisting of K components is specified by two parameter vectors:

$$\text{the populations of each state, } \boldsymbol{\pi} = (\pi_1, \ldots \pi_K); \sum \pi_k = 1$$

$$\text{and the mutation rates of each state, } \boldsymbol{\mu} = \begin{pmatrix} \mu_{1,1}, \ldots \mu_{1,D} \\ \vdots \\ \mu_{K,1}, \ldots \mu_{K,D} \end{pmatrix}$$

The fitting process begins with random initialization of $\boldsymbol{\mu}$ (drawn from {Beta(1,40)+0.001}) and $\boldsymbol{\pi} = (1/K, \ldots 1/K)$. Solutions for $\{\boldsymbol{\pi}, \boldsymbol{\mu}\}$ are then obtained via EM iteration. The appropriate equations for EM iteration incorporating missing data and active/inactive positions are:

$$z_{n,k} = \frac{\pi_k \prod_{i=1}^{D} \left[\mu_{k,i}^{x_{n,i}} (1 - \mu_{k,i})^{1 - x_{n,i}} \right]^{\phi_i \ \delta_{n,i}}}{\sum_{m=1}^{K} \pi_m \prod_{i=1}^{D} \left[\mu_{m,i}^{x_{n,i}} (1 - \mu_{m,i})^{1 - x_{n,i}} \right]^{\phi_i \ \delta_{n,i}}} \qquad \text{(Equation 1)}$$

$$\mu_{k,i} = \frac{\sum_{n=1}^{N} z_{n,k} \ \delta_{n,i} \ x_{n,i} + \alpha_i - 1}{\sum_{n=1}^{N} z_{n,k} \ \delta_{n,i} + \alpha_i + \beta_i - 2} \text{for } i \in \{active\} \qquad \text{(Equation 2.1)}$$

$$\pi_k = \frac{\sum_{n=1}^{N} z_{n,k}}{N} \qquad \text{(Equation 2.2)}$$

N is the total number of reads being clustered. α_i and β_i are parameters for a nucleotide-specific beta prior, which are set to:

$$\alpha_i = 1 + 0.01 \cdot r_{etoh,i} \cdot \sum_{n=1}^{N} \delta_{n,i}; \beta_i = 2$$

This prior encourages convergence to solutions with $\mu_{k,i} \geq r_{etoh,i}$ (preventing convergence to non-physical solutions where $\mu_{k,i} \rightarrow 0$), and contributes approximately $^1/_{100}$ the weight relative to the input data in determining the final value of $\mu_{k,i}$.

Equations 1, 2.1, and 2.2 are iterated until convergence, defined as $\max(|\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}|) < 10^{-4}$ and $\max(|\boldsymbol{\pi}^{(t+1)} - \boldsymbol{\pi}^{(t)}|) < 10^{-4}$, where $\{\boldsymbol{\pi}^{(t+1)}, \boldsymbol{\mu}^{(t+1)}\}$ denotes the parameters at iteration $t+1$. Converged solutions are assessed for validity as defined below. If an invalid solution is obtained repeatedly, then the nucleotides causing the invalid failure are inactivated ($\phi_i = 0$).

EM fitting is repeated from different random initializations until 3 identical converged solutions are found, up to a maximum of 50 attempts. Identical solutions are defined as $\max(|\boldsymbol{\pi}^a - \boldsymbol{\pi}^b|) < 0.03$ and $\max(|\boldsymbol{\mu}^a - \boldsymbol{\mu}^b|) < 0.01$, where a and b denote two different valid K-component solutions. If 3 solutions are not identified, then the model search terminates, selecting the K-1 model. If 3 identical solutions are identified, then the Bayesian information criteria (BIC) is used to evaluate whether the K-component model is significantly better than the K-1 model. Specifically, we compute:

$$\Delta BIC = (q_K - q_{K-1}) \ln(N) - 2 \ln(\mathcal{L}_K / \mathcal{L}_{K-1})$$

where q_K and $\mathcal{L}_K$ denote the number of parameters and the likelihood of the K-component model, respectively. If $\Delta BIC \leq -46$, indicating $\sim 10^{10}$ greater evidence for the K versus K-1 model, then the K-component model is accepted, K is incremented, and the EM fitting process repeats to a default maximum of K=5. Otherwise, the K-1 model is selected and the search terminates. We denote the final selected parameters as $\{\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\mu}}\}$.

Note that final fitted parameters can occasionally exhibit small run-to-run variation, caused by different sets of nucleotides being inactivated during each stochastic fitting trajectory. When observed, this variation is within the expected precision of the overall experiment: $\widehat{\boldsymbol{\pi}}$ correlate with R>0.995, and $\widehat{\boldsymbol{\mu}} \pm 3\%$.

Model validity criteria

Initial testing revealed that the EM algorithm occasionally converged to non-physical or degenerate solutions, particularly for K>2. We thus we perform several tests to ensure model validity; solutions that fail any of these validity criteria are rejected.

First, we require that $\{\boldsymbol{\pi}, \boldsymbol{\mu}\}$ fall within the physical boundaries:

$$\min(\boldsymbol{\pi}) \geq 0.001$$

$$\min(\boldsymbol{\mu}) \geq 10^{-5}$$

$$\max(\boldsymbol{\mu}) \leq \min(0.5, 3 \cdot \max(\boldsymbol{r_{DMS}}))$$

$$\mu_{k,i}/\mu_{l,i} \leq 200 \text{ for } i \in \{active\}; k,l \in \{1,\ldots K\}, k \neq l$$

Second, to prevent selection of degenerate models, we require the root mean square (RMS) difference between reactivity profiles be ≥ 0.005, excluding the top percentile of $\boldsymbol{\mu}$ differences:

$$RMS_{k,l} = \sqrt{\left(\mu_{l,i} - \mu_{k,i}\right)^2_{i \in \{active\}; \left(\mu_{l,i}-\mu_{k,i}\right)^2 \leq P_{99}}} \geq 0.005 \text{ for all } k,l \in \{1,\ldots K\}, k \neq l$$

where P_{99} denotes the 99th percentile of $(\mu_{l,i} - \mu_{k,i})^2$ values.

Third, because the minimum spacing required between mutations can favor artifactual solutions featuring anticorrelated $\boldsymbol{\mu}$ parameters over short-length scales, we discriminate against such solutions by computing the scores

$$s_{1,i} = \sum_{m=0}^{2} \phi_{i+m} \log\left(\frac{\mu_{k,i+m}}{\mu_{l,i+m}}\right) + \sum_{m=3}^{7} \phi_{i+m} \log\left(\frac{1-\mu_{k,i+m}}{1-\mu_{l,i+m}}\right) + \sum_{m=8}^{10} \phi_{i+m} \log\left(\frac{\mu_{k,i+m}}{\mu_{l,i+m}}\right)$$

$$s_{2,i} = \sum_{m=0}^{2} \phi_{i+m} \log\left(\frac{1-\mu_{l,i+m}}{1-\mu_{k,i+m}}\right) + \sum_{m=3}^{7} \phi_{i+m} \log\left(\frac{\mu_{l,i+m}}{\mu_{k,i+m}}\right) + \sum_{m=8}^{10} \phi_{i+m} \log\left(\frac{1-\mu_{k,i+m}}{1-\mu_{l,i+m}}\right)$$

for all positions i. Solutions are rejected if $|s_{1,i}| > 4.6$ and $|s_{2,i}| > 4.6$ for any i, equivalent to a >100-fold anticorrelated likelihood difference over a 10-nt window.

Finally, to prevent selection of poorly defined models, we require that the information matrix for $\{\boldsymbol{\pi}, \boldsymbol{\mu}\}$ be invertible. The information matrix is computed from the observed data as described (McLachlan and Peel, 2000).

Solving for inactive positions

Following identification of a valid converged solution, $\mu_{k,i}$ parameters of inactive nucleotides are obtained via constrained EM fitting. π_k and active $\mu_{k,i}$ are held fixed while iterating the equations

$$z'_{n,k} = \frac{\pi_k \prod_{i=1}^{D} \left[\mu_{k,i}^{x_{n,i}} \left(1-\mu_{k,i}\right)^{1-x_{n,i}}\right]^{\delta_{n,i}}}{\sum_{m=1}^{K} \pi_m \prod_{i=1}^{D} \left[\mu_{m,i}^{x_{n,i}} \left(1-\mu_{m,i}\right)^{1-x_{n,i}}\right]^{\delta_{n,i}}}$$

$$\mu_{k,i} = \frac{\sum_{n=1}^{N} z'_{n,k}\, \delta_{n,i}\, x_{n,i} + \alpha_i - 1}{\sum_{n=1}^{N} z'_{n,k}\, \delta_{n,i} + \alpha_i + \beta_i - 2} \text{ for } i \in \{inactive\}$$

until the inactive $\mu_{k,i}$ converge (defined as $\max(|\boldsymbol{\mu}^{(t+1)} - \boldsymbol{\mu}^{(t)}|) < 10^{-4}$).

Final model quality assessment

In silico benchmarking revealed that ML fitting occasionally yielded "valid" solutions that were nonetheless inaccurate in relation to the known generating ensemble. These inaccurate solutions were reliably identified as having poorly differentiated reactivity profiles, with relatively few nucleotides differing between model states. Thus, we perform a final quality assessment and warn users of potential low-quality solutions. These quality checks consist of:

i) Confirming that the RMS difference between all profiles is sufficiently different,

$$RMS_{k,l} = \sqrt{\left(\hat{\mu}_{l,i} - \hat{\mu}_{k,i}\right)^2_{i \in \text{ active}}} \geq 0.01 \text{ for all } k,l \in \{1,\ldots K\}, k \neq l$$

ii) Confirming that at least 20 nucleotide positions in each profile have distinct reactivities by computing the number distinct score (ND)

$$ND_{k,l} = \left[\sum_i \phi_i \cdot 1_{\boldsymbol{A}}\left(\left|\hat{\mu}_{l,i} - \hat{\mu}_{k,i}\right| > 0.01\right)\right] \geq 20 \text{ for all } k,l \in \{1,\ldots K\}, k \neq l$$

where 1_A is the indicator function.

iii) Confirming that the populations are well-defined according to the information matrix

$$\max(\sigma(\widehat{\pi}_k)) < 0.01$$

where $\sigma(\widehat{\pi}_k)$ is the standard deviation of $\widehat{\pi}_k$ obtained from the square root of the inverse of the information matrix.

Reactivity normalization

Following clustering, $\widehat{\boldsymbol{\mu}}$ parameters are transformed into normalized DMS reactivities that can be used as input for structure modeling. Raw reactivity profiles $(r_{raw,k})$ for each state k are obtained by subtracting r_{etoh} from $\widehat{\boldsymbol{\mu}}$. Nucleotide-specific normalization factors (n_i) that scale A/C and U/G reactivities to similar 0 to $\approx$2 ranges are computed as described (Mustoe et al., 2019) based on the maximum raw reactivities observed over all states. Normalized reactivities (r_k) are then obtained as $r_{raw,k}/n_i$.

RING- and PAIR-MaP analysis

Given a converged Bernoulli mixture model $\{\widehat{\boldsymbol{\pi}}, \widehat{\boldsymbol{\mu}}\}$, individual reads can be assigned *a posteriori* to the component (structure) from which they were derived. These assigned reads can then be input to PAIR and RING analyses, which identify correlated modifications between pairs of nucleotides that are indicative of through-space base pairing and tertiary interactions (Homan et al., 2014; Mustoe et al., 2019).

In principle, each read n can be assigned to its most probable parent structure using the complete data vector [$\mathbf{x_n} = (x_{n,1}, x_{n,2}, \ldots, x_{n,D})$]. However, in practice, the modification status of a nucleotide $(x_{n,i})$ can bias read assignment and thereby impose correlations in the assigned data. We address this issue by performing read assignment independently for each pair of nucleotides, excluding the considered nucleotides from the posterior probability calculation (Figure S1). Specifically, for the nucleotide pair (v, w), the posterior probability of a read $\mathbf{x_n}$ being derived from a structure k is computed as

$$z_{n,k}(v,w) = \frac{\widehat{\pi}_k \prod_{\forall\, i \neq v,w} \left[\widehat{\mu}_{k,i}^{x_{n,i}} \left(1-\widehat{\mu}_{k,i}\right)^{1-x_{n,i}}\right]^{\delta_{n,i}}}{\sum_{m=1}^{K} \widehat{\pi}_m \prod_{\forall\, i \neq v,w} \left[\widehat{\mu}_{m,i}^{x_{n,i}} \left(1-\widehat{\mu}_{m,i}\right)^{1-x_{n,i}}\right]^{\delta_{n,i}}}$$

Only reads that can be confidently assigned to a parent structure, defined as $\{\mathbf{X_k}(v,w)\} = \{\mathbf{x_n} | z_{n,k}(v,w) \geq 0.9\}$, are used for correlation analysis. The modification status of v and w is then tabulated across $\{\mathbf{X_k}(v,w)\}$ to obtain the {(*unmod*, *unmod*), (*mod*, *unmod*), (*unmod*, *mod*), (*mod*, *mod*)} contingency table (Homan et al., 2014; Mustoe et al., 2019). Note that this scheme means that the same read $\mathbf{x_n}$ can be assigned to different parent structures for different (v, w). For PAIR-MaP analysis, which considers correlations between windows of 3 nucleotides, nucleotides (*v, v+1*, v+*2, w, w+1, w+2*) are excluded from the products in the $z_{n,k}(v,w)$ equation.

As an additional control to eliminate artifactual correlations from biased read assignment, we perform an identical read-assignment and correlation analysis on a matched uncorrelated synthetic dataset (Figure S1). The synthetic dataset is generated from the experimentally fitted $\widehat{\pi}_k$ and $\widehat{\mu}_{k,i}$ parameters, treating all nucleotides as independent, and contains an identical number of reads as the experimental dataset. Any nucleotide pairs (v, w) observed correlated in this null dataset (P<0.001, G-test) are removed from the set of experimental correlations. Additionally, the contingency table for each (v, w) pair measured for the experimental data is required to be significantly different than the contingency table obtained for the null dataset (P<0.001, G-test); (v, w) pairs that fail this test are likewise removed from the set of experimental correlations.

In developing the read-assignment algorithm, we also tested the following alternative strategies: assigning reads using the complete data vector (not excluding (v, w)); using maximum *a posteriori* assignment; and assigning reads using Monte Carlo selection. Benchmarking tests on synthetic datasets unequivocally showed the strategy described above to have the best sensitivity and specificity.

After read assignment, RING and PAIR analysis are performed using v1.1 of RingMapper (Mustoe et al., 2019). RINGs were filtered for contact distance (>15) (Hajdin et al., 2013; Dethoff et al., 2018), and only positive correlations are shown (Mustoe et al., 2019).

Sequence alignment and data analysis

ShapeMapper (v2.1.5) was used to align and parse mutations from DMS-MaP sequencing experiments using the –amplicon and –output-parsed-mutations options. Adenine riboswitch data were aligned against the synthesized template sequence, and 7SK data were aligned against NR_001445.2. DANCE-MaP analysis was performed using the *DanceMapper* (v1.0) software. For the adenine riboswitch, *DanceMapper* was run with default options, which allows clustering into a maximum of 5 components. For 7SK, *DanceMapper* was run allowing a maximum of 3 clusters (–maxc=3) to save computational time searching for higher-order solutions; the absence of higher-order clustering solutions was confirmed via analysis of selected datasets. PAIR and RING analyses were performed via *DanceMapper* using default options.

7SK replicate analyses

7SK per-nucleotide reactivity data, PAIRs, RINGs, and structure models are derived from single experiments and are representative of at least 2 independent replicate datasets and analyses. Note that the sensitivity of PAIR and RING analysis depends strongly on read depth. The high-depth Jurkat in-cell and cell-free datasets shown in Figures 3, S3, and S4 were deliberately sequenced to high depths (>3 million), and the comparative lack of PAIRs and RINGs in other samples is attributable to lower sequencing coverage (0.3 to 1 million reads per sample). Ensemble populations are reported as the mean and standard deviation across replicates.

PAIR and RING reproducibility for in-cell Jurkat data were assessed by pooling reads from 8 independent replicates to create the "consolidated replicate" shown in Figure S4.

For cell-free Jurkat probing experiments, a total of 4 independent replicates were collected. Data did not reliably cluster into 3 states at read-depths below 1 million (2 of 4 replicates clustered into two-state ensembles consisting of A and B). Population means and errors were thus computed by comparing results from the single deeply sequenced dataset to a consolidated replicate constructed from the remaining 3 replicates. This consolidated replicated was also used to assess PAIR and RING reproducibility in Figure S4.

Structure modeling

Structure modeling was performed using *RNAstructure* (v6.2) (Reuter and Mathews, 2010). The *partition* module was modified to enable DMS-guided pairing probably calculations using nucleotide-specific DMS reactivity restraint functions (Mustoe et al., 2019); this modified code is available upon request and will be distributed in future releases of *RNAstructure*. Normalized DMS and PAIR restraints output by *DanceMapper* were passed to *fold* and *partition* using the –dmsnt and –x flags, respectively. Pairing probabilities shown in Figure 1E were computed using DMS reactivities only. All other structure modeling was performed using both DMS reactivities and PAIR restraints (when available). As part of *DanceMapper*, we provide the script *foldClusters.py* that automates structure modeling and visualization for all states of a deconvoluted ensemble. Arc diagrams were composed using the *ArcPlot* software.

RT-qPCR Analysis

Expression was standardized to indicated control gene (TBP) using the Pfaffl method (Pfaffl, 2001). Standardization of *HEXIM1* and 7SK expression to two other housekeeping genes yielded similar results. We also repeated analysis on 7SK using a different primer set, and again observed similar results. *CDK9* and *CCNT1* expression were also measured in the same experiments and no changes in these genes were observed. Data in Figure 6D corresponds to three biological replicates from 2 independent experiments (n = 6), except the 200 nM GAP-B sample which corresponds to three biological replicates from one independent experiment (n = 3) and 250nM JQ1 is the result of 3 independent experiments (n = 9). Data in Figure 6E correspond to three biological replicates from 3 independent experiments (n = 9). Analyses were performed using Prism9 (GraphPad) and error bars correspond to standard deviation.

Fitting adenine riboswitch titration data

Following prior studies (Reining et al., 2013), adenine riboswitch titration data were fit assuming the three-state equilibrium

$$\mathbf{OFF} \underset{K_{pre}}{\rightleftharpoons} \mathbf{ON_{-L}} \overset{\text{ade}}{\underset{K_d}{\rightleftharpoons}} \mathbf{ON_{+L}}$$

where ON_{-L} and ON_{+L} correspond to adenine-free and adenine-bound ON states, respectively. The equilibrium constant K_{pre} describes the pre-equilibrium conversion between the OFF and the ligand-free ON state, and K_d is the dissociation constant of ligand binding to the ON state:

$$K_{pre} = \frac{[ON_{-L}]}{[OFF]}; \mathrm{K_d} = \frac{[ON_{-L}][ade]}{[ON_{+L}]}$$

The total RNA concentration [RNA] is the sum of the three species:

$$[RNA] = [OFF] + [ON_{-L}] + [ON_{+L}] = [OFF] + K_{pre}\,[OFF] + \frac{K_{pre}}{\mathrm{K_d}}[OFF][ade] \qquad \text{(Equation 3)}$$

The amount of free ligand, [ade], can be further expressed as a function of the total ligand [ade_0]:

$$[ade] = [ade_0] - [ON_{+L}] = [ade_0] - [RNA] + [OFF] + K_{pre}[OFF] \qquad \text{(Equation 4)}$$

Substituting Equation 3 into Equation 4 and solving for [OFF] (Mathematica) yields two solutions, the valid one is:

$$[OFF] = \frac{\sqrt{4\,K_d\left(K_{pre}+K_{pre}^2\right)[RNA] + \left(K_d + K_d\,K_{pre} + K_{pre}[ade_0] - K_{pre}[RNA]\right)^2}}{2\,K_{pre}(K_{pre}+1)} + \frac{-K_d - K_d\,K_{pre} - K_{pre}[ade_0] + K_{pre}[RNA]}{2\,K_{pre}(K_{pre}+1)}$$

The combined fraction of species in the ON state, f_{ON}, is then:

$$f_{ON} = 1 - f_{OFF} = 1 - \frac{[OFF]}{[RNA]}$$

To account for incomplete saturation, we fit to the equation:

$$f_{ON} = m * \left(1 - \frac{[OFF]}{[RNA]}\right)$$

where m is the fractional saturation, [RNA] and [ade_0] are the input concentrations, and f_{ON} is the population measured by DANCE-MaP. Fits were obtained using the curve_fit module of SciPy in Python. The modest deviations of the data from the expected curve are likely to due to DMS modification of the adenine ligand, which reduces effective ligand concentration, and a central tendency bias of DANCE-MaP for clustering into more-equally-weighted groups (for example, a 75:25 ratio may be resolved as a 70:30 ratio). The fitted saturation fraction (m) values were 0.77 for both replicates, consistent with prior studies (Warhaut et al., 2017).