HIDDEN STRUCTURAL DYNAMICS OF HUMAN 7SK RNA REVEALED BY DECONVOLUTION AND ANNOTATION OF RIBONUCLEIC CONFORMATIONAL ENSEMBLES (DANCE-MAP)

Samuel W. Olson

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the College of Arts and Sciences

Chapel Hill
2021

Approved by:

Kevin Weeks

Jeff Aubé

Alain Laederach

David Margolis

Gary Pielak

ABSTRACT

Samuel W. Olson: Hidden structural dynamics of human 7SK RNA revealed by deconvolution and annotation of ribonucleic conformational ensembles (DANCE-MaP)
(Under the direction of Kevin M. Weeks)

7SK is an essential non-coding RNA that regulates eukaryotic transcription by sequestering positive transcription elongation factor b (P-TEFb). 7SK regulatory function likely entails changes in RNA structure, but characterizing dynamic RNA-protein complexes in cells has remained a critical challenge. We describe a new chemical probing strategy (DANCE-MaP) that uses maximum likelihood deconvolution and probabilistic read assignment to define simultaneously (i) per-nucleotide reactivity profiles, (ii) direct base pairing interactions, and (iii) tertiary and higher-order interactions for each conformation of multi-state RNA structural ensembles, all from a single experiment. We show that human 7SK RNA, despite significant heterogeneity, intrinsically codes for 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 is regulated by cell type and dynamically shifts in response to stress. We further demonstrate an antisense oligonucleotide strategy for inducing 7SK structural switching to modulate transcription in cells. Collectively, our data indicate that the 7SK structural ensemble functions as an integrator of diverse cellular signals to control transcription elongation.

"Time is the best appraiser of scientific work, and I am aware that an industrial discovery rarely produces all its fruit in the hands of its first inventor"

– Louis Pasteur

ACKNOWLEDGEMENTS

This work is the culmination of more than a decade's worth of post-secondary education and research which would not have happened without the overwhelming support of my family, friends, and colleagues. Foremost, I need to express my gratitude and thanks to my wife, Melissa. She has stabilized my home life by dealing with the stress of moving ~1700 miles away from family, a global pandemic, and everyday life as a graduate student. I owe my balance to you.

I would also like to thank my whole family from my parents, in-laws, siblings, and soon to be 11 nieces and nephews. They have all expressed a huge amount of support for me moving to pursue my degree. I do not get to see them as much as I would like, but when we do get together, they always know how to make me relax and enjoy our time together.

I also need to thank all the current and past members of the Weeks lab during my time at UNC. Their leadership and guidance in lab have added expertise to my experiments and coding, and spending time talking about life, "fun facts", and pursuits outside of lab has blossomed into lifelong friendships. I will cherish each and every one of you.

Lastly, I would like to thank my advisor, Kevin. Thank you for encouraging me to always think critically about my project and the big picture in my approach to experimentation. Your optimism about this project and me as a scientist has made my career significantly more enjoyable.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

A	adenine
alt	alternate
ASO	antisense oligonucleotide
BAF	barrier-to-autointegration factor
BET	bromodomain and extraterminal domain
BME	2-mercaptoethanol
BzCN	benzoyl cyanide
C	cytosine
cdk9	cyclin dependent kinase 9
cDNA	complementary DNA
CMCT	1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide metho-p-toluenesulfonate
CO^2	carbon dioxide
CTD	C-terminal domain
DANCE	deconvolution and annotation of ribonucleic conformational ensembles
DMEM	dulbecco's modified eagle medium
DMS	dimethyl sulfate
dmso	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EM	expectation maximization
EtOH	ethanol

ext	extension
FBS	fetal bovine serum
G	guanosine
GAP	gapmer
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
H2O	water
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HEXIM1/2	hexamethylene bis-acetamide inducible protein 1/2
HIV	human immunodeficiency virus
hnRNP	heterogeneous nuclear ribonucleoproteins
hrs	hours
hTERT	human telomerase reverse transcriptase
KCl	potassium chloride
kethoxal	β-ethoxy-α-ketobutyraldehyde
LaRP7	La-related protein 7
M	molar
MaP	mutational profiling
MePCE	methylphosphate capping enzyme
MFE	minimum free energy
Mg^{2+}	magnesium, ion
$MgCl_2$	magnesium chloride
min	minute
miRNA	micro RNA

ML	maximum likelihood
mL	milliliter
mM	millimolar
MM	mismatched
mRNA	messenger RNA
ncRNA	non-coding RNA
ng	nanogram
nM	nanomolar
NMR	nuclear magnetic resonance
PAIR	pairing ascertained from interacting RNA strands
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Pen/Strep	penicillin-streptomycin
Poll II	RNA polymerase II
P-TEFb	positive transcription elongation factor b
R	pearson correlation coefficient
RCF	relative centrifugal force
RING	RNA interaction groups
RNA	ribonucleic acid
RNP	ribonucleoprotein
RPL13a	ribosomal protein L13a
RPMI	Roswell Park Memorial Institute
rRNA	ribosomal RNA

RT	reverse transcription
RT-qPCR	reverse transcription quantitative polymerase chain reaction
s	seconds
SHAPE	selective 2′ hydroxyl acylation analyzed by primer extension
siRNA	small interfering RNA
SL	stem loop
snRNP	small nuclear ribonucleoprotein
SPRI	solid phase reversible immobilization
std	standard dev
STR	short tandem repeat
TBP	tributyl phosphate
Tris	tris(hydroxymethyl)aminomethane
U	units
ug	microgram
uL	microliter
w/	with
w/o	without
XIST	X-inactive specific transcript
˚C	degree celsius

CHAPTER 1: PRINCIPLES OF RNA STRUCTURE AND FUNCTION

RNA structure and function relationships

Our appreciation of RNA has expanded over the past few decades. RNA was thought to serve primarily as an intermediate component between DNA and protein, but we now know that many RNAs act as vital regulators of cellular function (1, 2). Indeed, many of the RNA found within a eukaryotic cell is not transcribed into protein (3). These non-coding RNAs have been found to have a wide variety of functions or modes of action (2). On the simplest level, non-coding RNAs can function through their primary sequence. For example, small interfering RNAs and microRNAs can bind directly with complementary nucleotides on a target substrate, flagging the target for destruction or repression (4). Larger non-coding RNAs can fold back onto themselves to form complex secondary and tertiary structures, similarly to proteins, and these folded RNA structures can perform an equally diverse set of functions (2, 5).

In biology a molecules function is often determined by the structure the molecule creates, and RNA is not an exception. For some RNAs, flexibility is required for their function: The long non-coding RNA XIST contains a dynamic and flexible domain that binds many proteins, organizing a non-membrane-bound nuclear compartment around a single chromosome (6, 7). Some non-coding RNAs require precise and stable secondary structures to function: Hairpin loops, bulges, and multi-helix junctions in ribosomal RNAs form complex tertiary structures that physically catalyze peptide bond formation during protein translation (8, 9). Some RNAs swap

between distinct structures in response to stimuli: The three-dimensional structure of riboswitches change when bound by a cognate ligand (10). These varying structural behaviors enable RNA to accomplish a variety of functions in cells.

Challenges for determining RNA structure

Methods for determination of secondary and tertiary RNA structure has evolved rapidly, with new innovations revealing the structures of larger and more complex RNAs in increasing detail. With the advent of high-throughput sequencing technologies, sequences can be readily and rapidly obtained. Given a primary sequence, computational modeling can be performed to approximate the secondary structure of the RNA. Modern folding software can simulate physical interactions between nucleotides to create a minim free energy model for secondary structure (11). Because this can be done completely *in silico*, modeling has become the most common first step to determining a secondary structure. However, because these models rely on thermodynamic parameters derived from model sequences, their ability to accurately predict secondary structure is limited.

Improvements to computational models require physical experimentation to give insight into how the RNA is structured in the environment of interest. Technologies that directly measure RNA structure suffer from various limitations. X-ray crystallography can reveal atomic-resolution details but requires that the RNA assume a homogeneous state in crystals, and the technique is not well suited for multi-state RNAs (12). Nuclear magnetic resonance can reveal details of RNA structure at high-resolution and can be used to characterize dynamic structural states but is restricted in its application to very small RNAs in solution (13).

Chemical probing is most used method for informing computational models for RNA secondary structure. Modifying the nucleobase using a small chemical probe is easy to do on most RNAs, and the data obtained can help construct the secondary structure (14). Some probes such as dimethyl sulfide (DMS) and 1-cyclohexyl-(2-morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate (CMCT) react with the base-pairing faces of nucleobases resulting in modification of non-base paired nucleotides (15). The reactivities of each nucleotide can be used as restraints in computational folding software creating a better modeled structure (11). DMS and CMCT have large biases toward certain nucleotides, making them non-ideal for informing model construction. SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) reagents modify the ribose sugars at the 2′ hydroxyl position and correlate with local nucleotide flexibility and dynamics (16, 17). SHAPE probes have the advantage of reacting with all four nucleotides and therefore have less bias than other chemical probing methods. SHAPE reagents are very useful for determining secondary structure for RNA isolated from cells and RNA transcribed *in vitro*. Some SHAPE reagents are cell permeable, although permeability depends on the cell type used. SHAPE reagents tend to show lower reactivity in cells, meaning the information gathered from in-cell experiments can be less robust than that obtained *in vitro* (18). Additionally, SHAPE probes do not directly monitor base pairing. SHAPE reactivity is a measure of nucleotide flexibility, which can be constrained either by base pairing or by formation of other types of structure, making analysis of the information gained complex.

The workflow of a chemical probing experiment was originally limited to detection of a single mutation as a “stop” during reverse transcription. The truncated cDNA products were first detected by gel electrophoresis (19) and later by the semiautomated capillary electrophoresis (20). These techniques were revolutionary at the time but led to a relatively low signal and

required a multi-step library creation protocol. In the previous decade, this issue was solved by the creation of mutational profiling (MaP) (17, 21). MaP uses an reverse transcriptase that reads through modified nucleotides and encodes the modifications into the cDNA product as a mutation. The frequency of mutations allows for single-nucleotide resolution RNA structure mapping (17, 21). With MaP, structural information can be obtained on virtually any RNA, even very long transcripts.

A benefit of the MaP technique is that when an RNA is modified more than once this signal can inform directly about through space interactions (22). The propensity of certain nucleotides to be modified on an RNA can change depending on whether a modification has already occurred, and this causes correlations between modified nucleotides. If the correlation occurs between two nucleotides that were base paired, then the correlation informs on the secondary structure of the RNA. The PAIR-MaP strategy was developed to evaluate these correlations (23). Modifications can also be correlated because of through-space interactions resulting in information that can be used as restraints to determine tertiary structure as is done using RING-MaP (22).

DMS as a versatile probe of RNA structure in living cells

Although DMS has been used as a chemical probe for RNA structure for over 40 years, its inherent bias for reaction with adenine (A) and cytosine (C) makes it a suboptimal choice for structure probing. Because SHAPE probes are nucleotide-agnostic, many labs have adopted SHAPE reagents for determining secondary structure (19). However, the lower reactivities of many SHAPE reagents, especially in cells, make observation of low-frequency correlated

modification events difficult (18, 24). An ideal chemical probe for RNA structure in living cells would be nucleotide base-agnostic, highly reactive, and cell permeable.

Recent advancements have made DMS a much better choice in RNA structure probe. First, using a high-pH buffer during DMS probing results in modification of guanosine (G) and uracil (U), overcoming the largest impediment of DMS probing (23). Second, the computational advancements implemented in PAIR-MaP have led to refinement in the determination of secondary structure (23). Third, the RING analyses allows for creation of an accurate model of both secondary and tertiary structure based on data from a single DMS probing experiment (22). (4) Finally, this experiment can be performed in living cells without any additional permeabilizing reagents (25). Correlated chemical probing records structural snapshots of single RNA molecules, and these snapshots can be sorted into distinct groups that together represent an RNA structural ensemble. In one implementation, different subpopulations of RNA structure were identified by a process of deconvolution of the ensemble of states obtained by analysis of DMS data (26). This technique allowed for the determination of secondary structures of an RNA that adopts multiple conformational states. Currently no published deconvolution analysis incorporates PAIR-MaP (23) and RING-MaP (27) technologies. Incorporating both PAIR and RING data with deconvoluted RNA structural states would result in an ideal pipeline for determining RNA structures (**FIGURE 1.1**).

Global transcription regulation by a non-coding RNA

The process of transcription converts the information in a DNA sequence into RNA. In all organisms this is a highly regulated process. The complex responsible for the transcription of most genes is RNA polymerase II (Pol II) (28). Pol II is responsible for transcribing all protein-

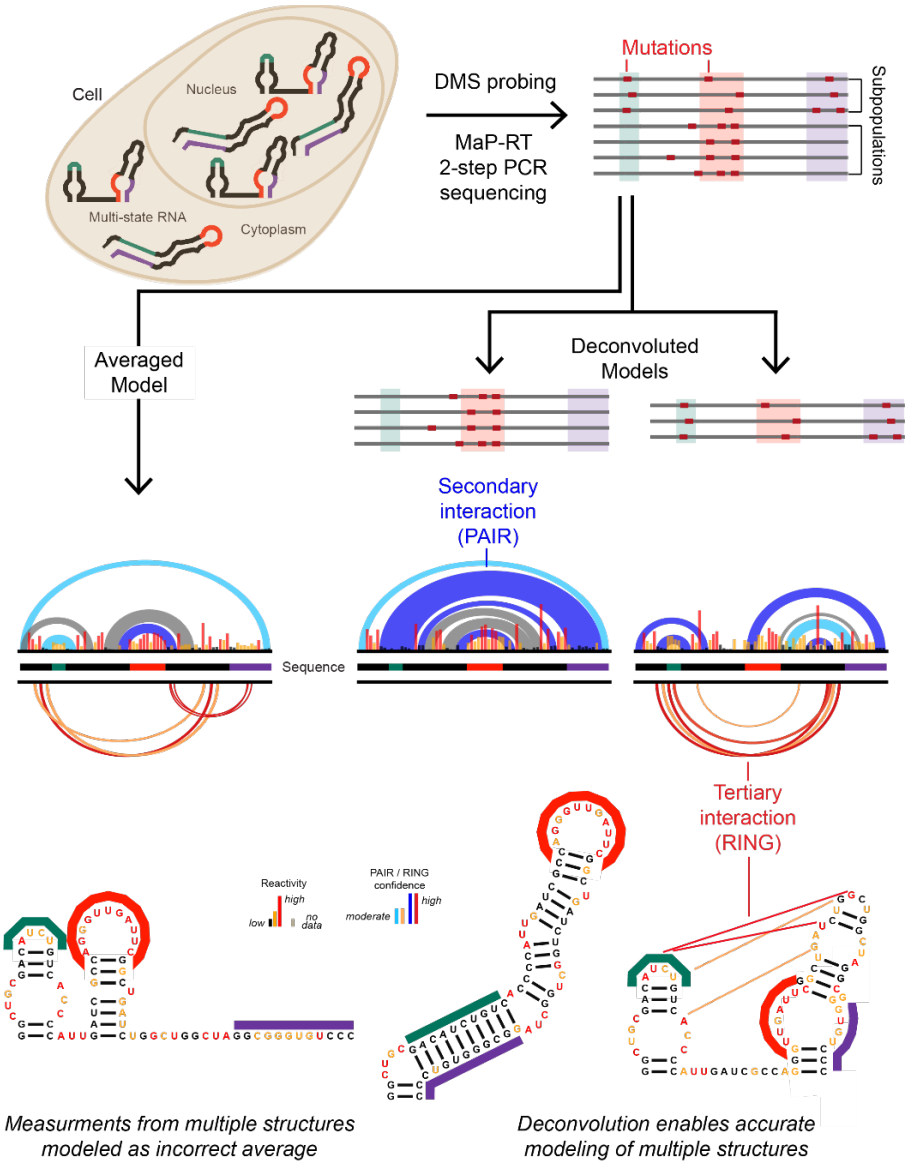

FIGURE 1.1: Explanation of deconvolution-based structure determination. RNA in cell is modified with DMS and subjected to reverse transcription under MaP conditions, where adduct-induced mutations are recorded and counted after library preparation and sequencing. RNAs with multiple structures generate distinctive chemical modification patterns in single-molecule MaP data. For an RNA sampling multiple states, an averaged, per-nucleotide reactivity profile may not be representative of any of the underlying structure states. Deconvolution-based analysis reveals both the individual reactivity profiles and populations of each ensemble state.

coding mRNAs and many non-coding RNAs (29). The elongation step for Pol II is tightly regulated by a protein dimer composed of cyclin T1 and the kinase cdk9, which form the positive transcription elongation factor b (P-TEFb) (30). P-TEFb phosphorylates the C-terminal domain (CTD) of Pol II, which is necessary for proper elongation of products (30). P-TEFb function is regulated by both the HEXIM1/2 protein and the 7SK non-coding RNA (31, 32) (**FIGURE 1.2)**. 7SK RNA is a 332-nucleotide RNA necessary for assembly of P-TEFb and HEXIM1/2 into a 7SK ribonucleoprotein (RNP) complex (33–35). When P-TEFb is sequestered in the 7SK RNP, cdk9 cannot phosphorylate the CTD, and transcription elongation is prevented (31, 32, 36). When the cell needs to ramp up transcription, the 7SK RNA undergoes a structural change that releases P-TEFb (31, 32).

Aside from P-TEFb and HEXIM1/2 association the 7SK RNA interacts with additional proteins such as LaRP7 and MePCE to stabilize the RNA from nuclease-mediated degradation. (37–39). Moreover many helicases, hnRNPs, and chromatin associated proteins have been proposed to interact with the non P-TEFb bound state of 7SK suggested to facilitate the release of P-TEFb (31, 32). Overall, many proteins are associated with the 7SK RNA but the mechanisms of binding and release are not well understood. A large reason for the confusion surrounding the 7SK RNA binding interactions is due to a lack of a well-defined structural has not been well characterized.

P-TEFb function is dysregulated in multiple diseases (40–42). In HIV the viral Tat protein directly competes with HEXIM1/2 for P-TEFb binding to be used in viral transcript elongation (43, 44). Antiretroviral treatments are used for treatment of HIV and target the quickly growing virally infected cells (45). In response to antiretroviral treatment the HIV in infected cells go into a latent form, P-TEFb availability is limited, creating a reservoir of slowly

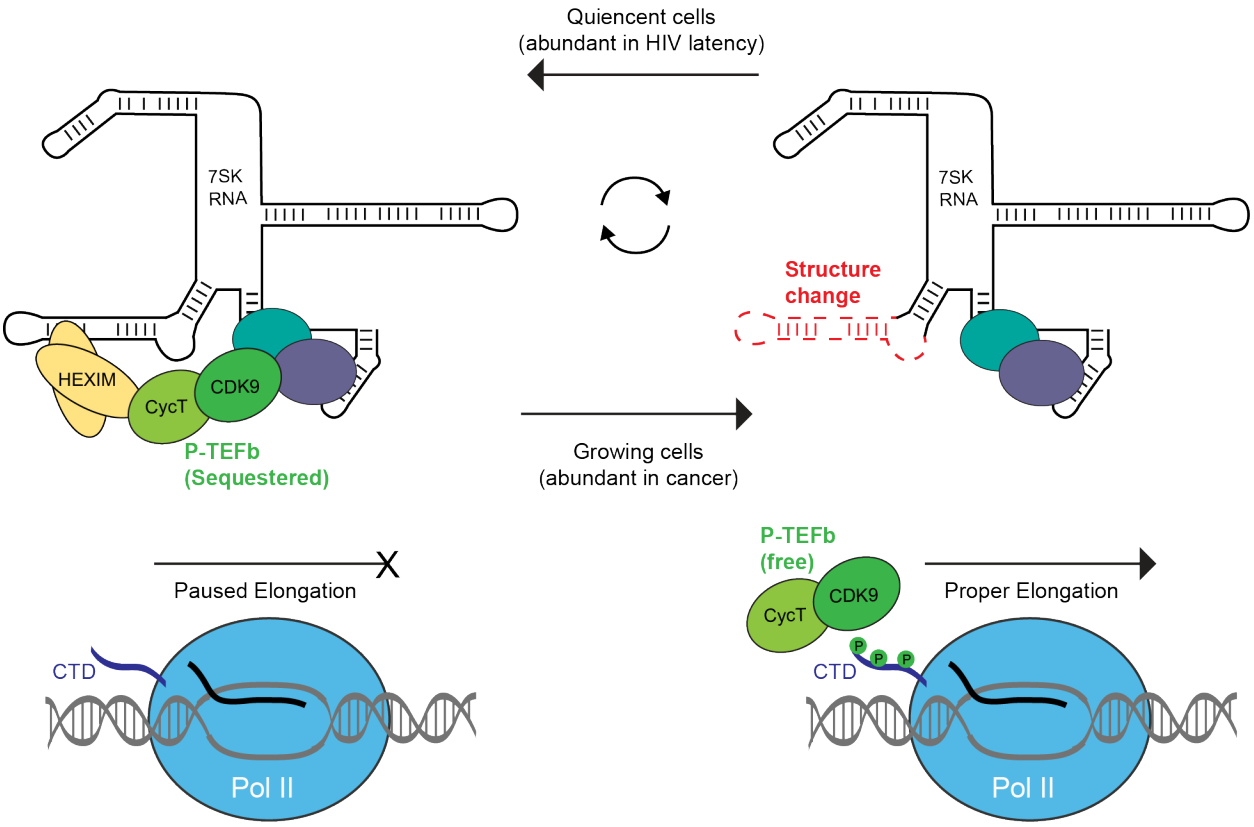

Figure 1.2: 7SK RNP is global transcription regulator. When P-TEFb is bound to 7SK RNA, its kinase function is inhibited and Pol II-mediated transcriptional elongation is paused. A structural change in the 7SK RNA (red) results in P-TEFb release, P-TEFb phosphorylates the CTD of Pol II, and transcription occurs.

dividing cells. These latent HIV infected cells can this evade traditional retroviral treatment and result in chronic disease that returns when retroviral treatments are stopped (46). In the 'kick and kill' strategy, the objective is to force more P-TEFb to be made available in latent cells (41, 42). The availability of P-TEFb in latent HIV cells has been shown to affect activation (47). It is possible that 7SK RNA could be targeted to induce the release of P-TEFb, selectively triggering activation of latent HIV. Activation of the latent HIV would be favored because the presence of the viral Tat protein creates a feedback loop quickly activating the latent HIV over other cellular transcripts. This would allow newly activated HIV to become susceptible to current HIV treatments (41, 42). Conversely, dysregulation of P-TEFb is one of many mechanisms transcriptional activation occurs in cancer cells. The 7SK/P-TEFb complex has been shown to be linked to cancer progression by effecting free P-TEFb levels and up regulating transcription (48–50). Small molecules that alter 7SK structure to induce P-TEFb sequestration could inhibit cell growth in various cancers and represents a possible target for cancer therapeutics (40).

Research overview

The primary goal of this work was to characterize the multiple structural states of the 7SK RNA and allow for informed perturbation of transcription. We used a new technology, deconvolution and annotation of ribonucleic conformational ensembles analyzed by mutational profiling (DANCE-MaP) to determine the structure of multi-state RNAs in living cells. The methodology is straightforward to implement and is applicable for any RNA in a single in-cell experiment.

In Chapter 2, I outline the DANCE-MaP pipeline used to deconvolute the data collected on 7SK RNA into its state subpopulations. Previously proposed 7SK RNA structures are likely

the products of averaged reactivity profiles that do not accurately represent the P-TEFb bound and unbound states. Using DANCE-MaP, I showed that the 7SK RNA adopts two stable conformations. I was able to manipulate the RNA structural states in predictable ways through targeted mutations. I found that state A contains a consensus HEXIM1/2 protein binding site (P-TEFb-bound state), and state B contains a truncated stem-loop supportive of P-TEFb release.

In Chapter 3, I used the two-state structure model for 7SK to structurally inform design of antisense oligonucleotides that induced P-TEFb release in cells. I first showed that flavopiridol, a small molecule known to cause P-TEFb release, does cause a switch in the amounts of each state present in the system. Then, based on the detailed structure of state B, I designed an antisense oligonucleotide that selectively switched the in-cell amount of 7SK exclusively to a single state. That antisense oligonucleotide caused a downstream effect on *HEXIM* mRNA levels, a validated method for determining P-TEFb release. Thus, comprehensive knowledge of the structure of the 7SK RNA allowed design of a global P-TEFb release approach by targeting the 7SK RNA.

Perspective

This work integrated biology, chemistry, and bioinformatics to determine the nucleotide-resolution structure of the complex and dynamic 7SK RNA. Through this work, I also developed a novel technology that can be applied to any RNA. RNA in general is an under-valued target of the academic world and pharmaceutical industry. Underutilization of RNA is due in large part to a lack of understanding of RNA structure. The DANCE-MaP strategy will allow study of RNAs to detect regions that could be targeted with small-molecules or nucleic acid-based drugs. I have attempted to lay out an easy-to-follow guide that will allow academic or industrial scientists to

determine the structures of RNAs that adopt more than one conformation and to selectively target a state of interest.

REFERENCES

1. S. R. Eddy, Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–929 (2001).

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

3. I. Dunham, *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

4. R. W. Carthew, E. J. Sontheimer, Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).

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

6. C. A. McHugh, *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).

7. C. Chu, *et al.*, Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).

8. J. A. Doudna, V. L. Rath, Structure and function of the eukaryotic ribosome: The next frontier. *Cell* **109**, 153–156 (2002).

9. P. B. Moore, T. A. Steitz, The involvement of RNA in Ribosome Function. *Nature* **418**, 229–235 (2002).

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

11. J. S. Reuter, D. H. Mathews, RNAstructure: Web servers for RNA secondary structure prediction and analysis. *BMC Bioinformatics* (2010) https:/doi.org/10.1093/nar/gkt290.

12. N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science (80-.).* **289**, 905–920 (2000).

13. A. Kotar, H. N. Foley, K. M. Baughman, S. C. Keane, Advanced approaches for elucidating structures of large RNAs using NMR spectroscopy and complementary methods. *Methods* **183**, 93–107 (2020).

14. C. Ehresmann, *et al.*, Probing the structure of RNAs in solution. *Nucleic Acids Res.* **15**, 9109–9128 (1987).

15. S. Stern, D. Moazed, H. F. Noller, Structural Analysis of RNA Using Chemical and Enzymatic Probing Monitored by Primer Extension. *METHODS Enzymol.* **164**, 481–489 (1988).

16. E. J. Merino, K. A. Wilkinson, J. L. Coughlan, K. M. Weeks, RNA structure analysis at single nucleotide resolution by Selective 2′-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).

17. M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, 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 (2015).

18. M. J. Smola, J. M. Calabrese, K. M. Weeks, Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).

19. K. M. Weeks, Advances in RNA structure analysis by chemical probing. *Curr. Opin. Struct. Biol.* **20**, 295–304 (2010).

20. F. Karabiber, J. L. McGinnis, O. V. Favorov, K. M. Weeks, QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *Rna* **19**, 63–73 (2013).

21. N. a Siegfried, S. Busan, G. M. Rice, J. a E. Nelson, K. M. Weeks, RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–65 (2014).

22. P. J. Homan, *et al.*, Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13858–63 (2014).

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

24. E. Calo, *et al.*, RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).

25. M. Zubradt, *et al.*, DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* **14**, 75–82 (2016).

26. P. J. Tomezsko, *et al.*, Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).

27. A. Krokhotin, A. M. Mustoe, K. M. Weeks, N. V Dokholyan, Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* (2016).

28. M. Levine, C. Cattoglio, R. Tjian, Looping back to leap forward: Transcription enters a new era. *Cell* **157**, 13–25 (2014).

29. R. G. Roeder, 50+ Years of Eukaryotic Transcription: an Expanding Universe of Factors and Mechanisms. *Nat. Struct. Mol. Biol.* **26**, 783–791 (2019).

30. Q. Zhou, T. Li, D. H. Price, RNA polymerase II elongation control. *Annu. Rev. Biochem.* **81**, 119–143 (2012).

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

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

33. B. M. Peterlin, D. H. Price, Controlling the Elongation Phase of Transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).

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

35. D. Martinez-Zapien, *et al.*, The crystal structure of the 5' functional domain of the transcription riboregulator 7SK. *Nucleic Acids Res.* **45**, 3568–3579 (2016).

36. N. Czudnochowski, C. A. Bösken, M. Geyer, Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* **3** (2012).

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

38. C. D. Eichhorn, Y. Yang, L. Repeta, J. Feigon, Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7 [Biophysics and Computational Biology]. *Pnas*, 1–10 (2018).

39. B. J. Krueger, *et al.*, 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 (2008).

40. C. M. Olson, *et al.*, Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. *Nat. Chem. Biol.* **14**, 163–170 (2018).

41. D. D. Richman, *et al.*, The challenge of finding a cure for HIV infection. *Science (80-.).* **323**, 1304–1307 (2009).

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

43. L. Muniz, S. Egloff, B. Ughy, B. E. Jády, T. Kiss, Controlling cellular P-TEFb activity by the HIV-1 transcriptional transactivator tat. *PLoS Pathog.* **6** (2010).

44. J. H. N. Yik, R. Chen, A. C. Pezda, C. S. Samford, Q. Zhou, A Human Immunodeficiency Virus Type 1 Tat-Like Arginine-Rich RNA-Binding Domain Is Essential for HEXIM1 To Inhibit RNA Polymerase II Transcription through 7SK snRNA-Mediated Inactivation of P-TEFb. *Mol. Cell. Biol.* **24**, 5094–5105 (2004).

45. S. Broder, The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral Res.* **85**, 1–18 (2010).

46. R. T. Davey, *et al.*, HIV-1 and T cell dynamics after interruption of highly active antiretroviral therapy (HAART) in patients with a history of sustained viral suppression. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 15109–15114 (1999).

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

48. Y. Cheng, *et al.*, LARP7 is a potential tumor suppressor gene in gastric cancer. *Lab. Investig.* **92**, 1013–1019 (2012).

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

50. J. L. Tan, *et al.*, Stress from Nucleotide Depletion Activates the Transcriptional Regulator HEXIM1 to Suppress Melanoma. *Mol. Cell* **62**, 34–46 (2016).

CHAPTER 2: MULTIPLE STRUCTURES OF THE 7SK RNA REVEALED BY DANCE MAP

Introduction

RNA molecules fold back on themselves into complex secondary and tertiary structures that provide the basis of specific protein recognition, ligand binding, and broad gene regulatory functions (1, 2). Most RNA elements likely sample more than one structure, and this underlying structural complexity enables RNAs to function as regulatory switches (3). mRNA-based switches have been identified that regulate transcription, splicing, and translation of specific genes in response to metabolites (riboswitches) (4) and protein binding (5, 6). Large-scale RNA structural dynamics also underpin function of ribonucleoprotein (RNP) complexes such as the ribosome (7, 8) and the spliceosome (9). Nevertheless, and despite their broad importance, RNA switches remain exceedingly difficult to identify, quantify in terms of their structure and in-cell equilibria, and link to functional outcomes.

The 7SK RNA is highly structured (10) and several studies support the model that P-TEFb binding and release involves remodeling of 7SK RNA structure (11, 12), or that the 7SK RNA exists in distinct conformations depending which proteins are bound (11–13). Multiple models for the 7SK RNA structure have been proposed (10, 12, 14, 15), but the accuracy of these models, whether they represent distinct co-existing states, and how these states might differentially modulate function remains unknown. To date, 7SK structure has primarily been studied using ensemble-average chemical probing approaches that are poorly suited for identifying coexisting RNA conformations or resolving structural dynamics. Typical of most

non-coding RNAs, 7SK sequences show weak sequence covariation, precluding informative evolutionary analysis (16, 17). The 7SK RNP thus encapsulates broad features illustrating how RNA structural complexity endows functional complexity, and how such structural complexity deeply frustrates mechanistic understanding.

Single-molecule chemical probing is emerging as a transformative technology for characterizing RNA structure and dynamics in living cells. The foundational conceptual advance is mutational profiling (MaP) reverse transcription, whereby a polymerase reads through and measures multiple chemical adducts per RNA molecule, recording them as mutations in complementary DNA (18). Massively parallel sequencing enables measurement of correlated modification events across hundreds of thousands of molecules, which encode rich information regarding RNA ensemble composition (18–20), and through-space secondary (21, 22) and tertiary (7, 18, 23) structure interactions. However, existing single-molecule analysis frameworks only extract one type of information at a time (ensemble composition, base paring, or tertiary interactions). Probing data can be deconvolved into multiple co-existing reactivity profiles; however, the structure of each state is not measured directly but rather is merely inferred based on compatibility with the reactivity data. This inference problem is typically ambiguous and becomes increasingly so for longer and more complex RNAs, particularly in cells. Moreover, the existence of multiple RNA structural states has made direct measurement and assignment of base pairing and tertiary interactions infeasible. To resolve complex RNA ensembles and their biological functions accurately, especially in cells, it is thus critical to develop integrated strategies that both deconvolute per-nucleotide reactivity profiles and directly measure state-specific secondary and tertiary interactions.

Here I took advantage of a maximum likelihood (ML) strategy, DANCE-MaP

(deconvolution and annotation of ribonucleic conformational ensembles) that extracts and annotates a large fraction of the total information from a single-molecule chemical probing experiment (**FIGURE 2.1**). DANCE -MaP was developed by Dr. Anthony Mustoe and extensively benchmarked using the known two state adenine riboswitch system. DANCE-MaP directly visualizes complex RNA ensembles from MaP probing data, including direct detection of base pairs and tertiary interactions for each sub-state, at nucleotide resolution in a single experiment (**FIGURE 2.2**). Dr. Anthony Mustoe found while benchmarking the adenine riboswitch using DANCE-MaP, consistency with the ON and OFF states are defined by the presence of the aptamer domain and the SD-sequestering helix, respectively, but additionally substantial heterogeneity elsewhere in the molecule. These data are consistent with and clarify observations from prior biophysical and chemical probing studies (19, 24–26). Therefore, I applied DANCE-MaP to discover a large-scale, sequence-encoded structural switch in the 7SK RNA. The structural-switch model developed here rationalizes a large body of prior data and directly links P-TEFb release to concerted remodeling of 7SK structure in a cellular environment and cell type specific way. This work helps to establish DANCE-MaP as a powerful framework for directly resolving complex ensembles in cells and rationalized all the previous 7SK models into a new consensus model.

Results

Native 7SK RNA exists as a multi-state structural ensemble

Motivated by the fundamental role of 7SK RNA in transcriptional regulation and prior

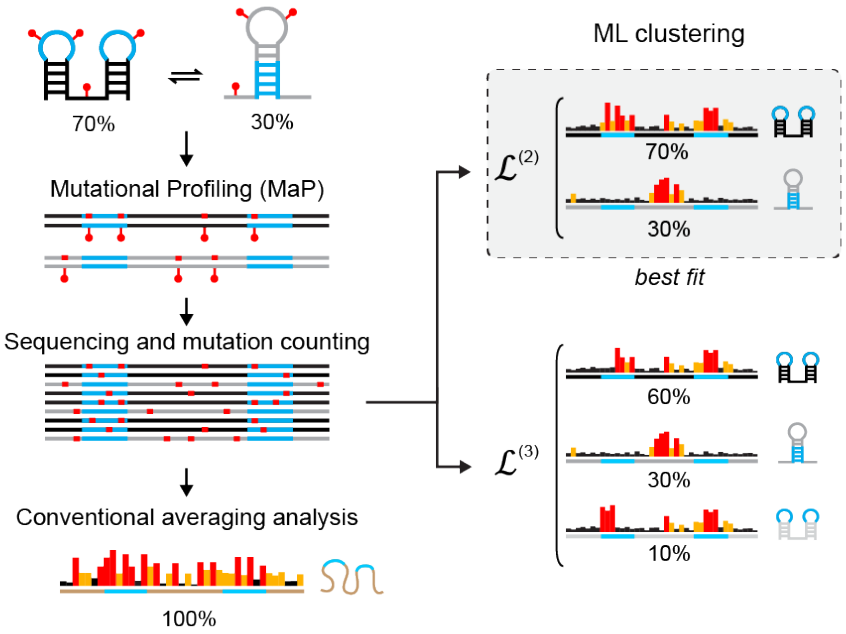

FIGURE 2.1: Schematic of ML ensemble deconvolution. RNAs with multiple structures generate distinctive chemical modification patterns in single-molecule MaP data.. Typically, these data are averaged together into a single composite 1-dimensional reactivity profile, which can be non-representative of either underlying ensemble state. ML analysis of the single-molecules reads reveals both the of individual reactivity profiles and populations of the underlying generating RNA ensemble.ML analysis reveals both the individual reactivity profiles and populations of each ensemble state. Using the individual reactivity for each state to create a better-defined structural model.

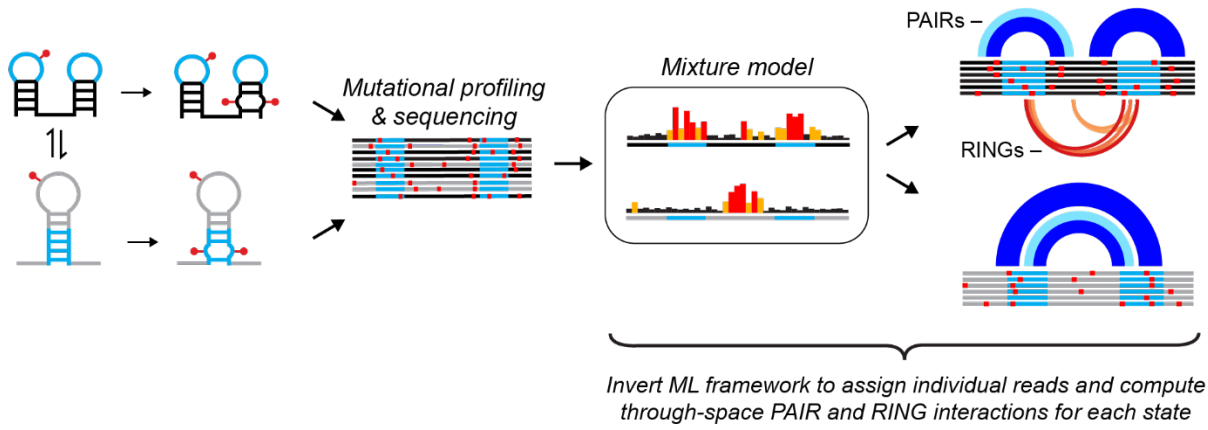

FIGURE 2.2: DANCE-MaP enables direct detection of state-specific base pairs and tertiary interactions through inverted ML framework. DMS induces correlated disruptions of base pairing and tertiary interactions, measurable by MaP. Following ML deconvolution of per-nucleotide reactivity profiles, individual 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.

evidence of 7SK dynamics (11–13), I sought to define the 7SK structural ensemble in-cell and its role in regulating transcription. Previous studies suggest that the 7SK nucleotide reactivity changes upon release of P-TEFb and changes to the RNP complex (11–13). For this reason, I believe that deconvoluting the 7SK RNA to gain complete picture of the dynamics is key to understanding the role 7SK plays in transcriptional regulation. To test my hypothesis I performed DMS-MaP experiments on living human Jurkat cells and used amplicon sequencing to obtain high-coverage single-molecule DMS probing data for the 7SK RNA. Conventional averaged analysis (without deconvolution) yielded per-nucleotide reactivity profiles consistent with prior studies, and generally compatible with previously proposed SL1, SL3, and SL4 stem-loop structures (10, 15) (**FIGURE 2.3**, *top*). However, as observed in previous probing studies (10–12, 27), many nucleotides exhibit intermediate reactivities, consistent with significant, unresolved structural heterogeneity. ML single-molecule analysis indicated that 7SK structural heterogeneity reproducibly resolves into three states: A, B, and H (populations of 39% ± 3, 46% ± 2, 15% ± 2, respectively; **TABLE 2.1**). State populations and reactivity profiles ($R > 0.96$) were highly reproducible over 10 biological replicates performed months apart. The minority H (heterogenous) state shares some features with B, but generally has high reactivity across the RNA. By comparison, the predominant A and B states show punctate regions of high and low nucleotide reactivity, consistent with these states representing well-defined structural states (**FIGURE 2.3**). Notably, nucleotides throughout the SL1 region including U28, U30, U66, and U68 are unreactive in state A, but reactive in states B and H, corresponding precisely to nucleotides previously identified as changing conformation upon P-TEFb release (11, 12). Numerous additional differences occur throughout the 7SK RNA, indicative of a concerted global structural switch.

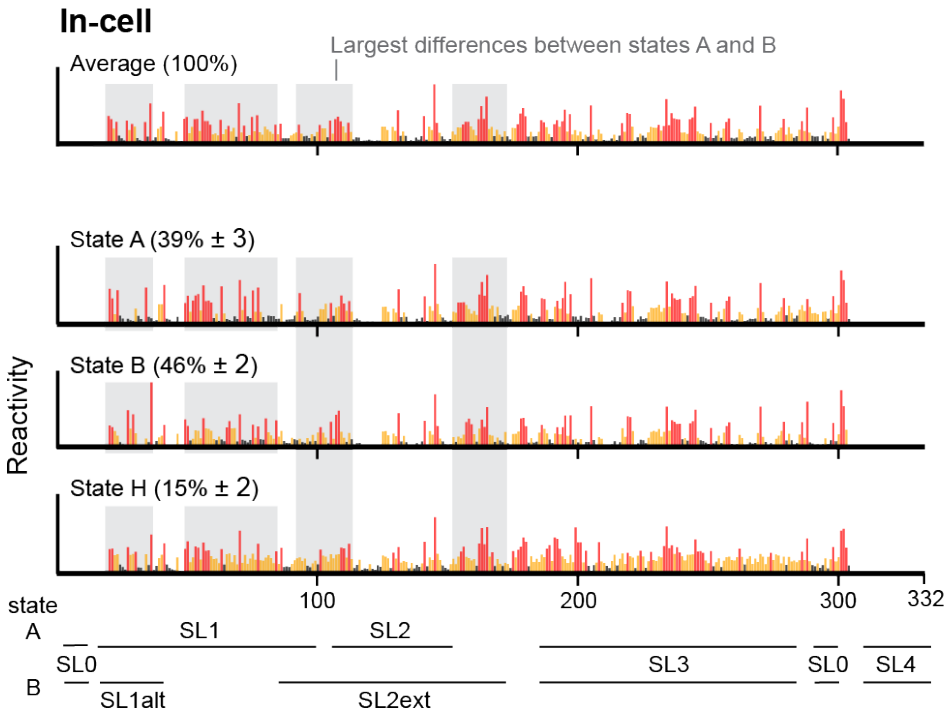

FIGURE 2.3: Per Nucleotide Reactivity of 7SK states. Averaged and DANCE-deconvoluted reactivity profiles for 7SK RNA in cells. Major differences between states A and B are highlighted with gray shading. State H (heterogenous) has high reactivity throughout. Population averages and standard deviations are computed over 10 replicates. Stem loop (SL) structural landmarks are indicated at bottom.

Sample	A%	B%	H%	number of reads
1	41	47	12	3,792,127
2	40	44	16	405,390
3	43	42	15	620,897
4	34	48	18	532,913
5	39	46	15	564,212
6	38	48	14	809,449
7	40	45	15	306,960
8	35	49	16	376,969
9	58	42	--	736,405
10	40	47	13	612,621
11	42	44	14	594,668
Consolidated (2-9)*	46	43	11	4,425,454
Mean (std) (samples 1-8,10-11)[#]	39 (±3)	46 (±2)	15 (±2)	

TABLE 2.1: Jurkat in-cell replicates: * Reads from samples 2-9 were aggregated into a single "consolidated replicate" used for replicate PAIR and RING in downstream analysis. [#] State population means and standard deviations were computed across individual biological replicates, excluding sample 9 which only clustered into 2 states.

To understand the extent to which these conformational dynamics depend on proteins that bind 7SK in cells, DMS probing experiments were repeated on total RNA from Jurkat cells that was heat denatured and refolded. ML deconvolution revealed that cell-free 7SK also populates 3 states: A and B, and a mixed (M) state that shares features of both A and B (populations 43% ± 8, 29% ± 6, 28% ± 1; R > 0.98 between reactivity profiles; two consolidated replicates, see **TABLE 2.2**). States A and B are the same as observed in-cell (R = 0.96 and R = 0.95 for states A and B, respectively **FIGURE 2.4**). Only diffuse reactivity protections and enhancements are observed relative to the in-cell RNA, which implies that the A and B states are dynamically rather than stably bound by proteins in cells. The lack of an H state under cell-free conditions is consistent with H representing a state heterogeneously bound by proteins in cells. Conversely, the lack of a mixed state in cells suggests that bound cellular factors specifically favor the A and B states. Thus, the 7SK RNA sequence intrinsically codes for two energetically balanced states, specifying a large-scale structural switch that behaves similarly with or without bound proteins denatured and refolded. ML deconvolution revealed that cell-free 7SK also populates 3 states: A and B, and a mixed (M) state that shares features of both A and B (populations 43% ± 8, 29% ± 6, 28% ± 1; R > 0.98 between reactivity profiles; two consolidated replicates, see **TABLE 2.2**). States A and B are the same as observed in-cell (R = 0.96 and R = 0.95 for states A and B, respectively **FIGURE 2.4**). Only diffuse reactivity protections and enhancements are observed relative to the in-cell RNA, which implies that the A and B states are dynamically rather than stably bound by proteins in cells. The lack of an H state under cell-free conditions is consistent with H representing a state heterogeneously bound by proteins in cells. Conversely, the lack of a mixed state in cells suggests that bound cellular factors specifically favor the A and B states. Thus, the 7SK RNA sequence intrinsically codes for two energetically balanced states,

Sample	A%	B%	M%	number of reads
1	37	34	29	3,920,245
2	59	41	0	745,299
3	44	39	17	701,912
4	61	39	0	474,976
Consolidated (2-4)*	48	25	27	1,935,468
Mean (std) (sample 1, consolidated) [#]	43 (±8)	29 (±6)	28 (±1)	

TABLE 2.2: Jurkat cell-free replicates. * Reads from samples 2-4 were aggregated into a single "consolidated replicate" used for replicate PAIR and RING analysis in downstream anlysis. [#] Due to unreliable 3-state clustering at lower read-depths, population means and standard deviations were computed using sample 1 and the consolidated replicate.

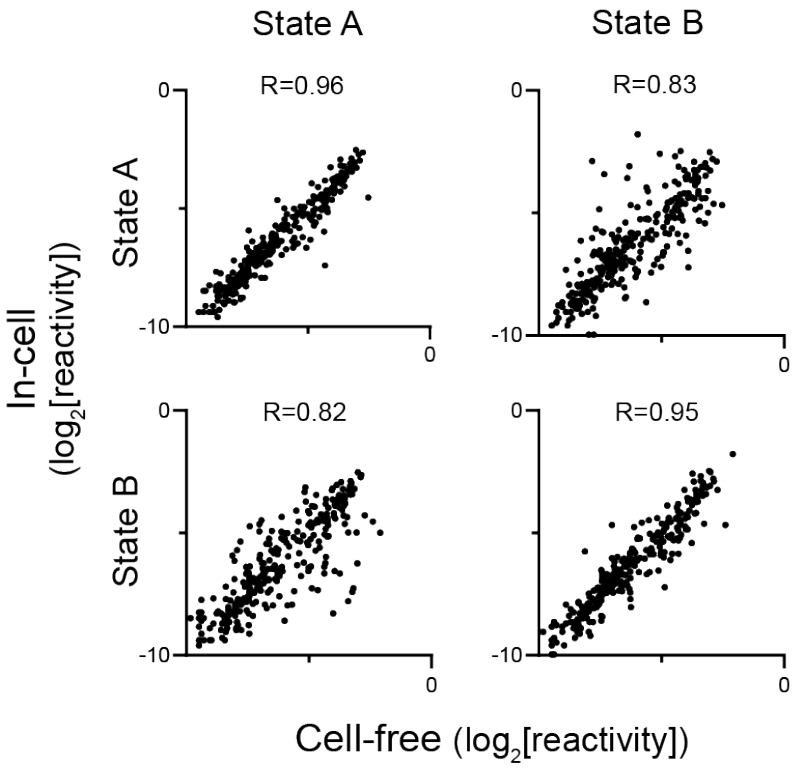

FIGURE 2.4: **In-cell and cell-free per nucleotide reactivity comparison** Comparison of state A and B per-nucleotide reactivities for in-cell and cell-free 7SK. Person's R is shown.

specifying a large-scale structural switch that behaves similarly with or without bound proteins.

Direct base pair mapping and structure modeling reveals 7SK architecture

To fully resolve the secondary structure and potential tertiary structures of each state, high depth sequencing datasets (>3 million reads) were obtained that provide power sufficient to detect through-space PAIRs and RINGs across the 332 nt long 7SK RNA. These data reveal numerous PAIR signals that directly report base-paired structural elements distinctive to each state (**FIGURE 2.5**), reproducible between in-cell and cell-free environments. The structural features were additionally consistent between replicates created by combining biological replicates to create a similar read depth sample (**FIGURE 2.6**). Using these PAIR data in combination with per-nucleotide reactivity profiles to build detailed secondary structure models for the A and B states (**FIGURE 2.7**). The resulting structural models reveal that 7SK folds into two globally different conformations, each of which is supported by distinctive per-nucleotide reactivities and state-specific, direct PAIR correlations. Both states also show alternative predicted pairing possibilities and PAIRs suggestive of residual heterogeneity (**FIGURE 2.5**), indicating that states A and B should be interpreted as "macro" states rather than pure states. Nevertheless, each macro state possesses key defining structural features, and these state-specific structures clearly link 7SK conformational dynamics to P-TEFb binding and release.

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

State A largely recapitulates classic models of 7SK structure, blending features predicted by early probing studies (10) and more recent evolutionary analyses (15) (**FIGURE 2.8**). The SL1 helix is the defining structural feature of state A, and is directly supported by PAIRs both in

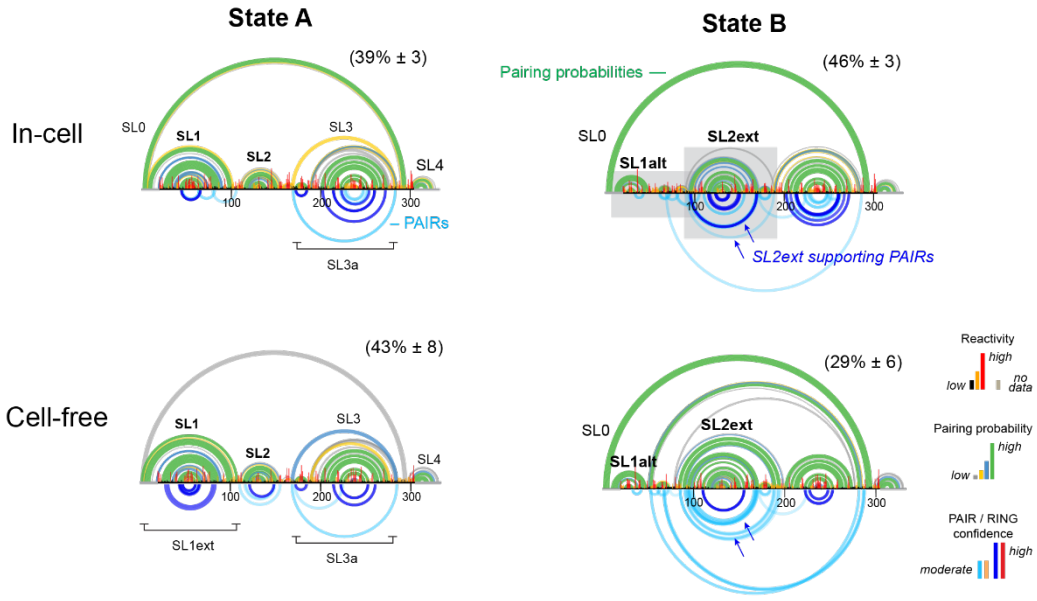

FIGURE 2.5 In-cell and cell-free structural models for states A and B. Modeled base pairing probabilities (*top*) and directly measured PAIRs (*bottom*) are shown as arcs.

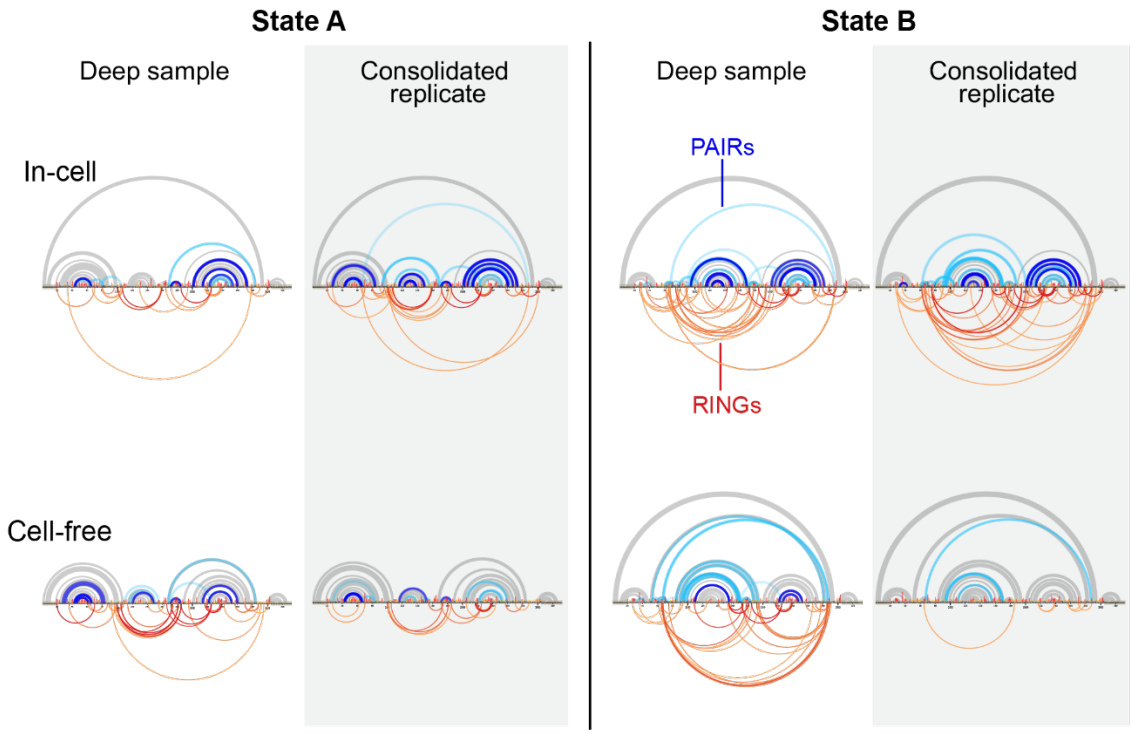

FIGURE 2.6: State-specific PAIR and RING correlations measured for 7SK RNA. PAIR and RING correlations for in-cell and cell-free RNA measured from a single deeply sequenced sample (same as shown in **FIGURE 2.5**) and a consolidated replicate constructed by pooling multiple lower depth independent replicates (see **TABLE 2.1**). Minimum free energy structures are shown as arcs (top, gray). High- and moderate-confidence PAIRs are shown as dark and light blue arcs, respectively. Through-space RING correlations are shown as high (dark red, G>100) and moderate (light red, G>20) confidence arcs. RINGs were filtered for contact distance (>15) (23, 28), and only positive correlations are shown (22).

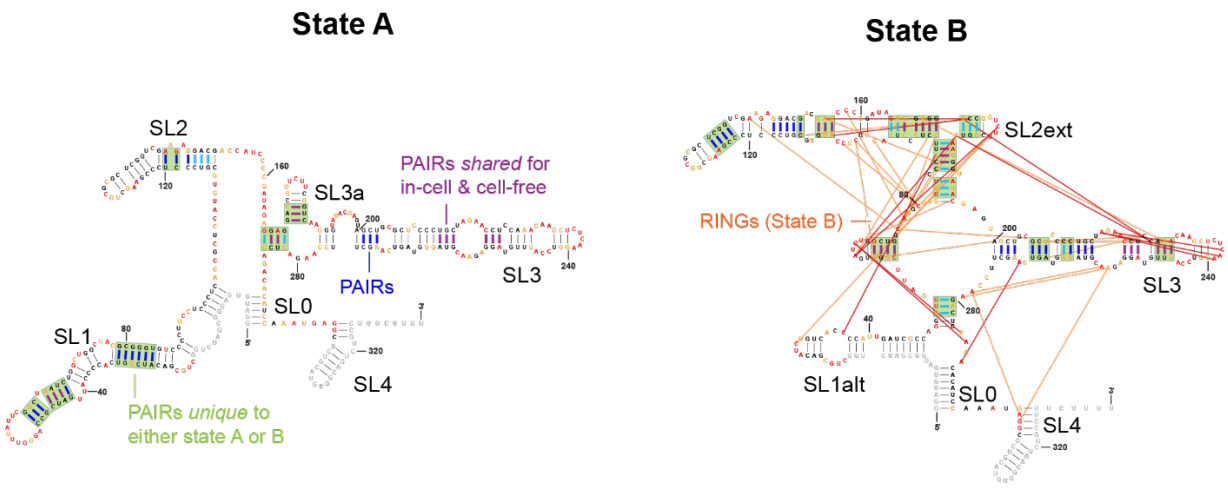

FIGURE 2.7: Secondary structure models shown with supporting PAIR and RING data. Consensus secondary structure models, shown in individual base pair format. Per-nucleotide reactivities are colored as per panel A. RINGs observed for state B are shown with orange-red lines, consistent with through-space structural communication and tertiary interactions. Measured PAIRs that directly support either state A or B are boxed in green.

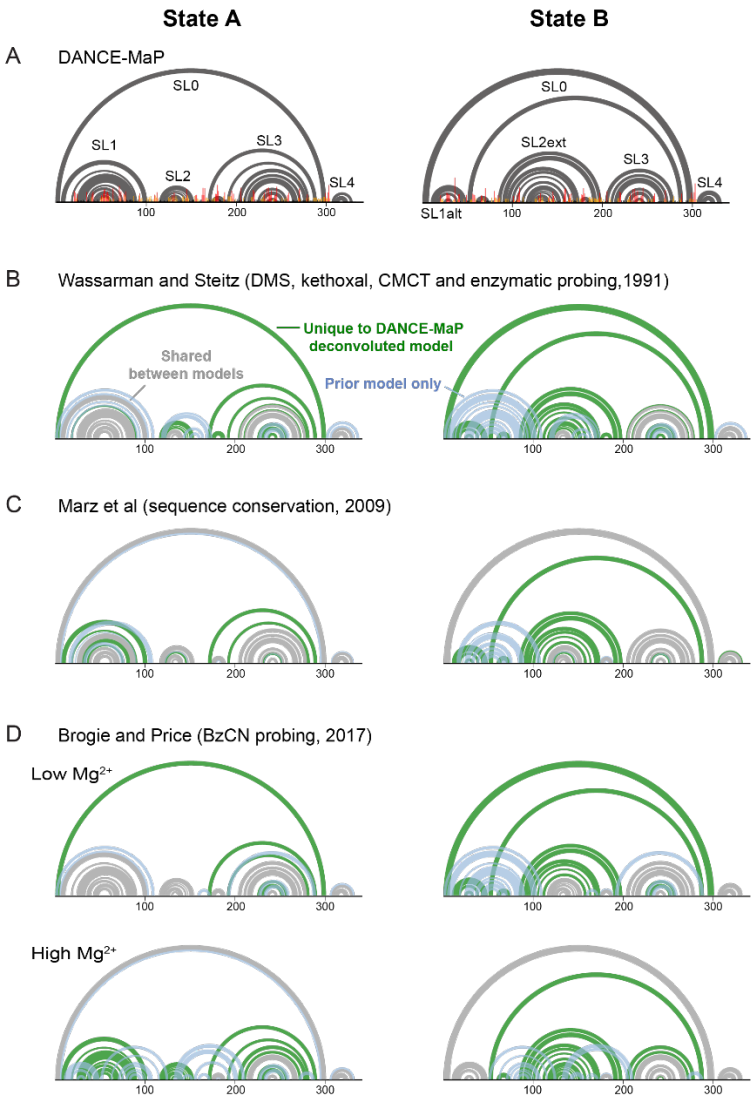

FIGURE 2.8: Comparison of states A and B, defined by DANCE-MaP, with prior models of the 7SK RNA structure. (10, 12, 15) (A) DANCE-MaP consensus secondary structures models for states A and B. Base pairs are shown as arcs; structural landmarks are labeled for each state. (B-D) Comparison of DANCE-deconvoluted states A and B with previously described models for the 7SK RNA. Base pairs shared between DANCE-deconvoluted and prior models, unique to DANCE models, and unique to prior models are gray, green and blue, respectively. The Brogie and Price high Mg^{2+} model (panel D) was inferred using the partial structure published by the authors (in **FIGURE 2B**), and using reactivities (provided in the supplementary information) to model remaining positions (12).

cells and for the cell-free extracted RNA (**FIGURE 2.5**). SL1 has been extensively validated as the primary recognition site for HEXIM1/2 and P-TEFb, based on in vitro binding assays (29, 30), analysis of P-TEFb–bound 7SK fractions from cells (12), and in-cell functional assays (31, 32). The population of state A in cells, ~40%, is also consistent with the estimated fraction of 7SK bound by P-TEFb (33). Thus, state A was assigned as the P-TEFb bound (sequestered) state.

Structure modeling indicates that this P-TEFb sequestered state contains the long-range SL0 pairing interaction between the 5' and 3' ends that "circularizes" the RNA. While the lack of data for the 5' strand of SL0 due to overlap with the primer binding site, the 3' strand of SL0 is lowly-to-moderately reactive in-cells, consistent with formation of a dynamic, partially stable stem (**FIGURE 2.5, 2.9**). By contrast, the alternative extended form of SL1, which out-competes SL0 in the cell-free RNA (SL1ext, see **FIGURE 2.5**), is reactive in cells, arguing against the

State B constitutes a P-TEFb released state with remodeled SL1 and central domains

State B constitutes a novel structure without close literature precedent (**FIGURE 2.5, 2.8**). Most notably, SL1 is absent. Instead, this region folds into the previously postulated SL1alt stem (11, 12). Although overlap with the primer binding site precludes measurement of SL1alt-specific PAIRs, the disappearance of the SL1 PAIRs (as observed in state A) implies that this region adopts an alternative structure in state B. SL1alt, and not SL1, is also clearly supported by per-nucleotide DMS reactivities and by pairing probabilities (**FIGURE 2.3**). P-TEFb does not bind SL1alt (32, 34) and, indeed, P-TEFb binding converts SL1alt to SL1 *in vitro* (12). Conversely, release of P-TEFb induces conversion of SL1 to SL1alt (11, 12). The 45% population of state B in cells is also consistent with the fraction of 7SK that is in a P-TEFb-

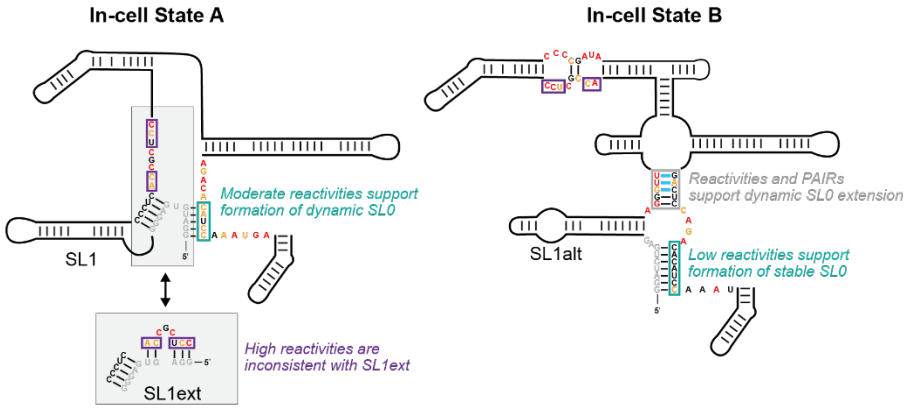

FIGURE 2.9: Determination of consensus 7SK RNA structures. Expanded view of DMS reactivities for the SL0 region for states A and B, in cells. The state A inset shows the alternative SL1ext pairing. Green and purple boxes indicate nucleotides paired in SL0 and SL1ext, respectively. For state B, reactivities and PAIRs supporting formation of an extension of SL0 and SL2ext are shown.

released state in cells (33). Thus, state B constitutes a P-TEFb released state.

SL1alt is coupled to formation of a major extension of SL2, which was termed SL2ext, that has not been observed previously (**FIGURE 2.5**). Re-pairing to form SL2ext is directly supported by PAIRs in both in-cell and cell-free RNAs (**FIGURE 2.5**, blue arrows). Indeed, PAIR analysis was essential for resolving these interactions: SL2ext is not predicted when structure is modeled only on the basis of per-nucleotide reactivities (**FIGURE 2.10 A,B**). Moderate DMS reactivities indicate that SL2ext is dynamic, and these dynamics are enhanced in cells, consistent with this region being bound by diverse proteins (13, 35, 36). Thus, while SL2ext is modeled as lowly probable in cells, the overall consistency between cell-free and in-cell PAIRs leads us to conclude SL2ext is present in state B in cells (**FIGURE 2.7**). The interdependence between SL1alt and SL2ext, visualized in this study, rationalizes prior observations that P-TEFb binding induces structural changes in the 7SK central region, located up to 200 nts away from SL1 (12). As discuss below, this structural reorganization overlaps the principal regions bound by P-TEFb release factors, consistent with allosteric coupling between SL1, SL2ext, and release factor binding sites. Strikingly, RING analysis revealed a dense network of correlations for both cell-free and in-cell RNA (**FIGURE 2.6, 2.7, 2.11**). Prior *in vitro* studies observed salt-dependent formation of a B-like state, consistent with B potentially being stabilized by tertiary interactions (12). 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 tertiary interactions.

State H constitutes a heterogenous P-TEFb released state

The (heterogenous) state H features well-defined SL0, SL1alt, and SL2 stems, but is otherwise highly reactive and contains no other stable structural elements (**FIGURE 2.3, 2.10**).

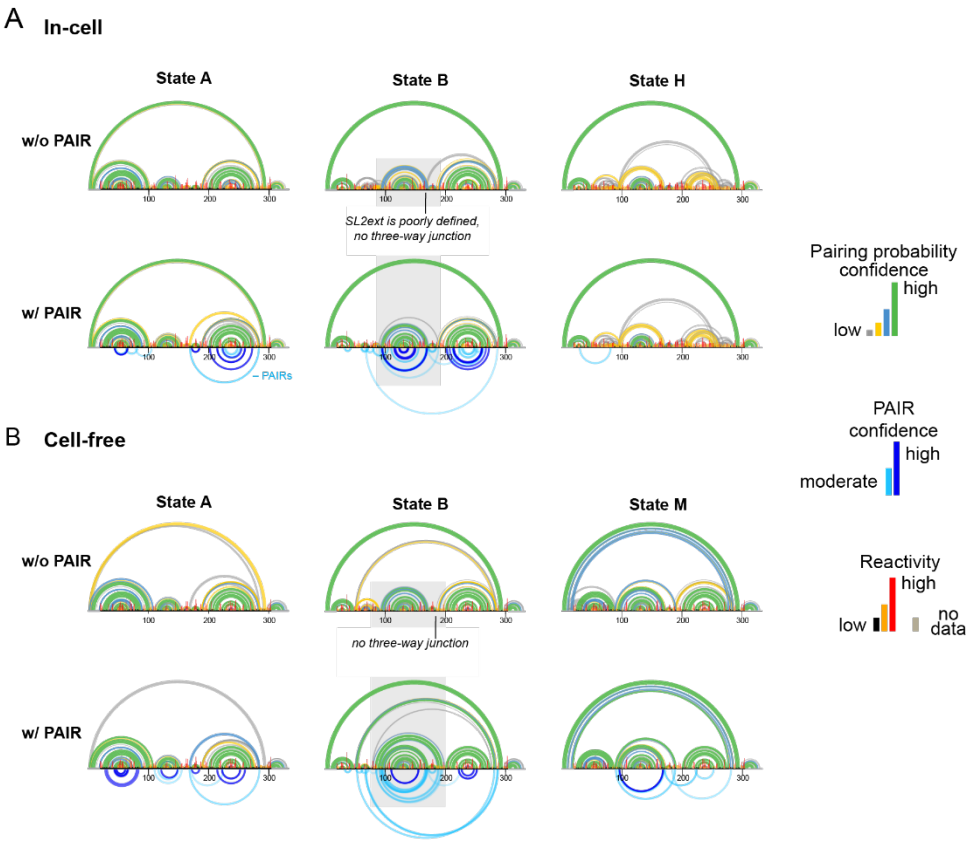

FIGURE 2.10: Comparisons of in-cell and cell-free structures modeled with and without PAIR data. (A) In-cell and (B) Cell-free structure models are shown. SL2ext region, which is poorly defined in the absence of PAIR data, is highlighted.

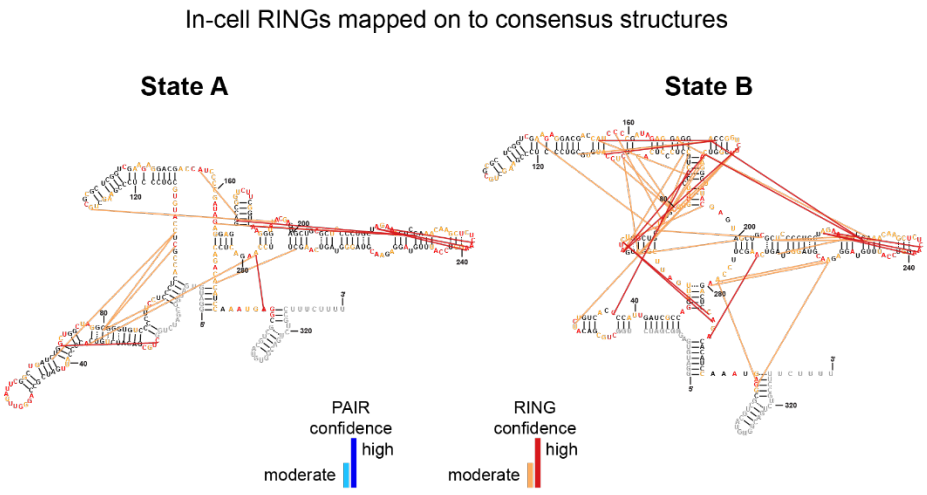

FIGURE 2.11: RING correlations from the deeply sequenced sample superimposed on the consensus models for states A and B. In addition to state-specific contact distance filtering used in A, RINGs were additionally filtered by contact distance (<15) in relation to the other structural state to exclude signals originating from residual secondary structures. Correlations correspond to deeply sequenced data.

The clear presence of SL1alt indicates that, like B, state H is a P-TEFb released state. However, the elevated reactivities within the central domain indicate this state is highly heterogenous. The interpretation of this state is a composite of diverse lowly populated protein-bound structures. Alternatively, H may represent a transition state between A and B that is stabilized by helicases, which are known to bind to unstructured regions in RNA.

The in-cell structure of the 7SK RNA

In sum, this data reveal that 7SK folds into at least three structures that comprise P-TEFb–competent and –released states. PAIR and RING measurements, measured individually for each state, provide pivotal and direct evidence of new structural elements (SL2ext) and a compact core in state B that are invisible to per-nucleotide analyses. This multi-state ensemble rationalizes a large compendium of biochemical and functional data on the 7SK RNP, and implies that 7SK contains an allosteric switch that structurally couples release factor binding sites in the central domain to HEXIM1/2–P-TEFb binding in SL1.

Mutational analysis validates importance of SL2ext in 7SK structural switching

To validate the proposed DANCE-MaP-resolved models and to define the role of individual structural elements in 7SK switching, mutants were designed to probe the A and B states in collaboration with J. Winston Arney. As an initial control, DANCE-MaP experiments were performed on *in vitro* transcripts of the native sequence RNA. The native RNA folds into a two-state ensemble, consisting of the A and B states with populations 71 ± 4 and 29 ± 4%, respectively ($R > 0.95$ between reactivity profiles; N=3). The two states are almost identical to the A and B states observed for the cell-free RNA (**FIGURE 2.12**), with the exception that SL0

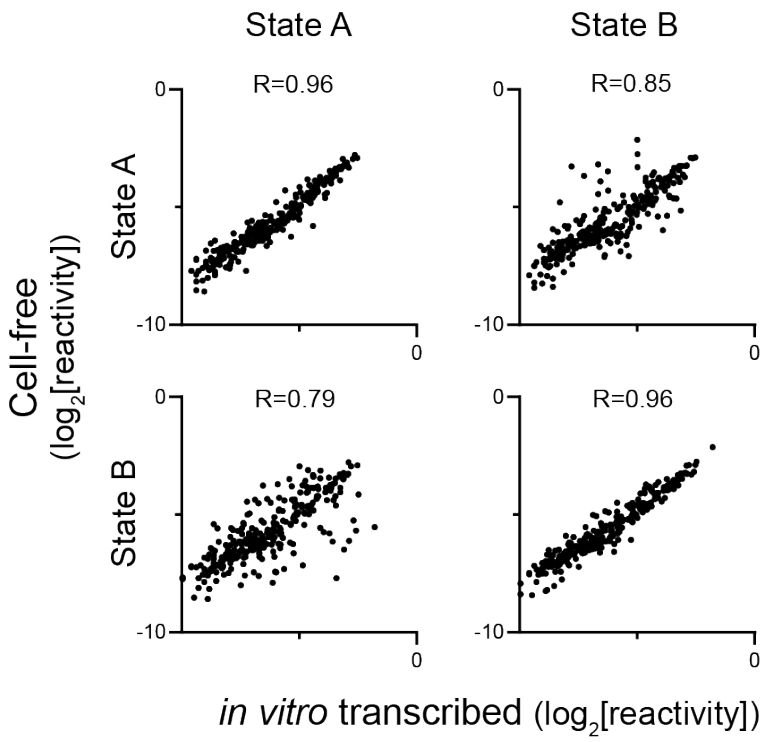

FIGURE 2.12: Cell-free and *in vitro* per nucleotide reactivity comparison. DMS reactivities for states A and B for native *in vitro* transcribed versus cell-free 7SK RNA. Pearson's R is shown.

is further destabilized in State A. The lack of an M state in the *in vitro* transcribed RNA is also consistent with SL0 destabilization. This destabilization may reflect increased formal charge at the 5'-phosphate compared the endogenous transcript, which is 5'-γ-methylated (37, 38). These data from *in vitro* transcripts further validate that the A and B states are intrinsic features of the 7SK RNA sequence (**TABLE 2.3**).

To validate the B state structure model, three mismatches in SL1 were introduced while leaving SL1alt pairing intact (mutant M1, **FIGURE 2.13, 2.14A**). DANCE-MaP experiments showed that M1 completely ablated state A, with the RNA clustering into 2-3 B or B-like states (N=3; **FIGURE 2.14B**, **2.15**). Detection of multiple B/B-like states is consistent with the native B state representing a composite of multiple similar structures, and with the M1 mutation modestly destabilizing SL1alt. Conventional probing studies have also observed that mutational disruption of SL1 results in B-like structures (12). Rescue of the M1 mutation by restoring base pairing complementarity in SL1 (M1+M2) recovers the native A:B equilibrium (78 ± 2 and 22 ± 2% populations, respectively; N=3; **FIGURE 2.14, 2.15**).

Next the role of SL0 in 7SK switching was investigated. While in-cell data support SL0 formation in both states, SL0 is clearly more stable in state B, and others have proposed that SL0 drives 7SK dynamics (12). Ablation of SL0 via the M3 mutation (**FIGURE 2.13**) had minimal impact on the 7SK ensemble: M3 populates an 80 ± 3, 20 ± 3% equilibrium consisting of 1-2 A/A-like states and a heterogenous B-like state that replaces SL0 with other interactions (N=3, **FIGURE 2.14, 2.15**). Mutation of only three of seven base pairs in SL0 gave similar results (not shown). Previous studies of M3-like mutants are also consistent with SL0 disruption having minimal impact on SL1:SL1alt dynamics (12). Thus, the stability of SL0 does not drive the A:B equilibrium.

Sample	A/A-like%	B/B-like%	number of reads
Native			
1	67	33	93,136
2	72	28	219,330
3	74	26	356,685
Mean (std)	71 (±4)	29 (±4)	
Consolidated	73	27	668,821
M1			
1		57, 22, 21	159,840
2		39, 36, 25	303,036
3		58, 42	320,229
Mean (std)		100 (±0)	
Consolidated		41, 32, 27	784,010
M1+M2			
1	77	23	210,719
2	49, 31	20	291,448
3	49, 28	23	142,086
Mean (std)	78 (±2)	22 (±2)	
Consolidated	46, 24	30	644,165
M3			
1	77	14,9	226,157
2	60, 21	19	368,473
3	61, 21	18	207,740
Mean (std)	80 (±3)	20 (±3)	
Consolidated	61, 20	19	910,460
M4			
1	100		638,079
2	55, 45		575,840
Mean (std)	100 (±0)		
Consolidated	63, 37		1,015,106

TABLE 2.3: IVT 7SK and mutants: State population means and standard deviations were computed across individual biological replicates

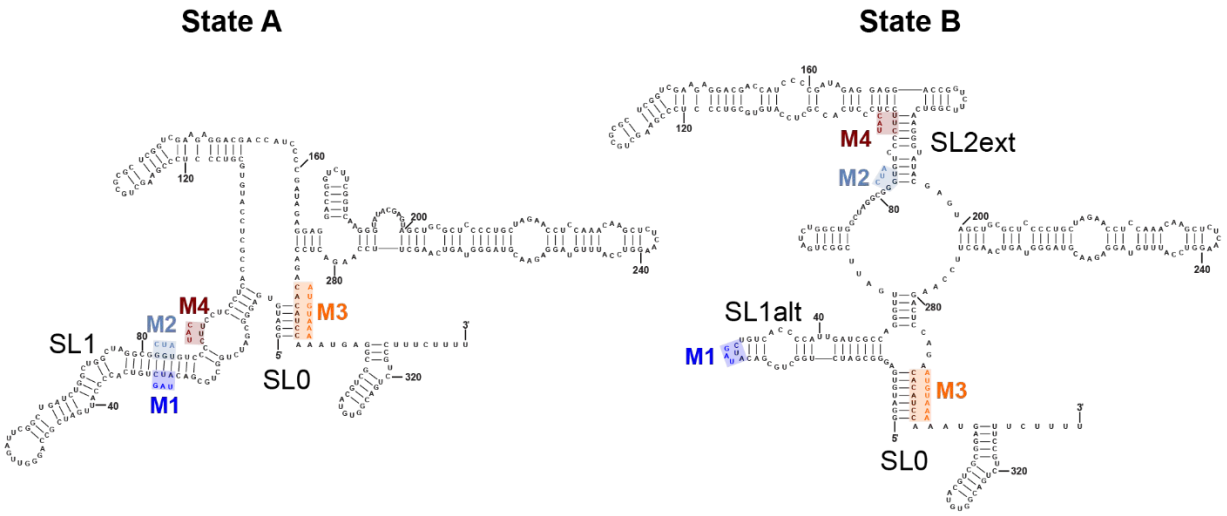

FIGURE 2.13: Assessment and validation of 7SK states A and B by mutational analysis.

RNA mutants. Mutations are shown superimposed on consensus state A and B structural models. Mutants and native sequence RNA were produced as *in vitro* transcripts.

A

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

B

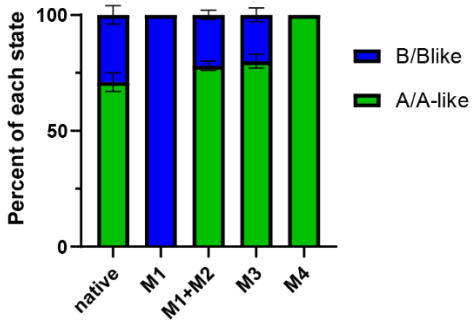

FIGURE 2.14: Mutations to the 7SK RNA results in single state shifts. (A) Summary of designed structural impact for each mutant. (B) Ensemble distribution observed for the native sequence RNA and each mutant.

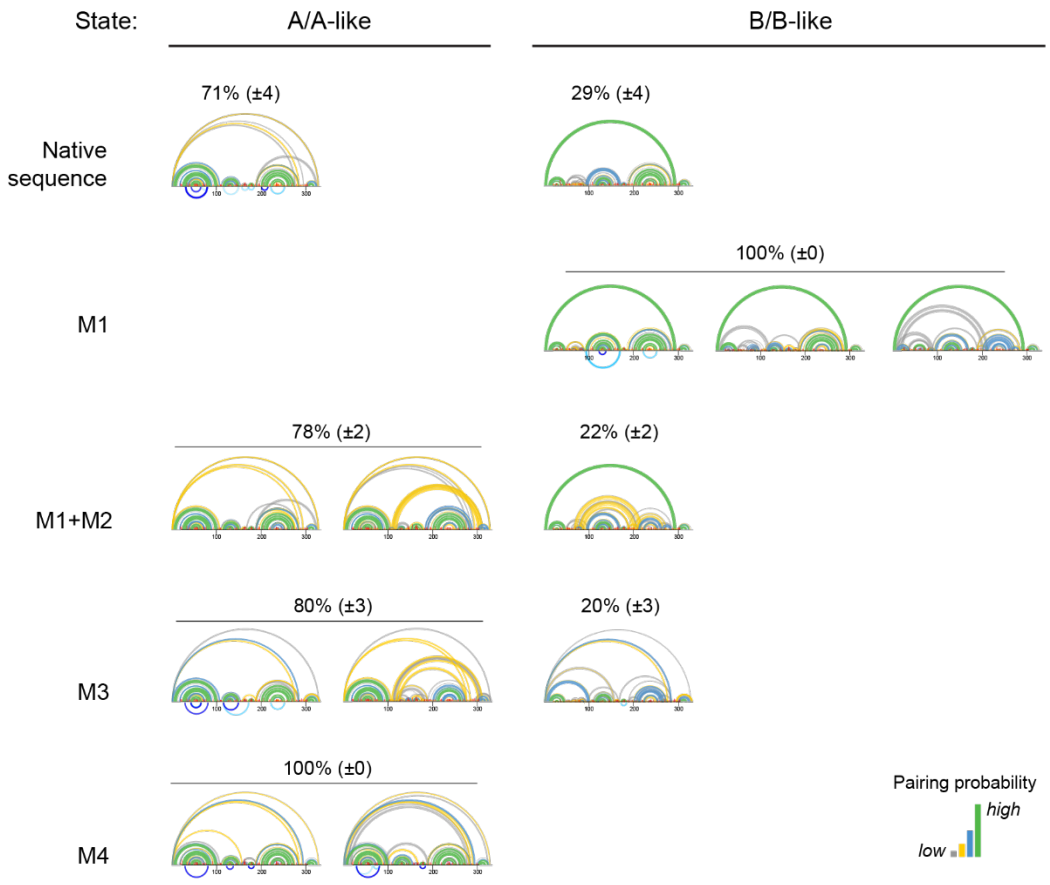

FIGURE 2.15: DANCE-MaP characterization of *in vitro* transcribed native sequence and mutant 7SK RNAs. Derived structural ensembles for each *in vitro* RNA. States were assigned to A/A-like and B/B-like based on comparison to the native sequence RNA. States were assigned by the following criteria: has clear SL1: A or A-like; has clear SL1alt and/or SL2ext: B or B-like. Percentages assigned to each state are shown. Probing experiments were performed in triplicate (native, M1, M1+M2, M3) or duplicate (M4). Structural models and PAIRs (limited due to low sequencing depth) obtained from DANCE analysis from combined replicate datasets are shown with population and errors obtained from the average and standard deviation across individual replicates.

Finally, role of SL2ext in A:B switching was examined. PAIR interactions directly support formation of the SL2ext stem in both cell-free and in-cell RNA (**FIGURE 2.5, 2.7**). Mutant, M4, was designed to disrupt the PAIR-supported 3-helix junction at the base of SL2ext in state B, but not perturb SL1 pairing in state A (**FIGURE 2.13**). Strikingly, this three-nucleotide mutation fully shifts the ensemble to A/A-like states (N=2; **FIGURE 2.14, 2.15**). Prior studies have also observed that mutations in the SL2ext region induce global remodeling of 7SK structure (12, 14), although the mechanistic basis was not resolved. Thus, even though SL2ext shows intermediate stability, this region is critical to 7SK A-B switching.

In sum, the 7SK ensemble can be rationally perturbed via targeted structure-informed mutations, validating the DANCE-resolved A and B states. Moreover, a concise mutation to SL2ext (M4) is sufficient to drive SL1/SL1alt switching: The data establish that the central core is an energetically accessible platform for (allosterically) modulating 7SK structure and activity.

Discussion

DANCE-MaP enables complete analysis of RNA structural ensembles

Most RNAs have the potential to fold into multiple structures, which creates numerous opportunities to regulate RNA biology. However, authoritatively defining RNA structural ensembles in cells has remained an unresolved challenge. *Bona fide* functional elements in both mRNAs and lncRNAs often show low sequence conservation, and span complex mixtures of well-structured and dynamic regions, ultimately meaning that establishing their structures and responses to cellular stimuli has remained a critical knowledge gap. Utilizing DANCE-MaP, a single-molecule chemical probing technology that simultaneously measures per-nucleotide reactivities, through-space base pairs (PAIRs), and tertiary interactions (RINGs) for up to three

co-existing structural states in a single, concise experiment in cells. DANCE-MaP further measures populations with thermodynamic precision enabling subtle, but impactful, measurements of ligand binding affinity and of RNA structural differences between cell types. Collectively, these integrated measurements now enable definitive and comprehensive characterization of RNA structural ensembles.

This work emphasizes the intrinsic complexity of RNA structural ensembles. For 7SK, the A, B, and H states are each distinguished by major structural landmarks, but ultimately represent class averages, or macrostates, 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 macrostate is a defining and crucial advance, enabling us to resolve otherwise invisible dynamic structures in cells and to model global RNA architectures with confidence. Direct PAIR measurements of base pairing were essential for resolving the SL2ext structure unique to state B of 7SK. State B contains significant through-space RINGs in the compact core, emphasizing that the in-cell ensembles are distinct in terms of both base pairing and higher-order structure. Integrated DANCE-MaP analysis represents a powerful advance over single-purpose deconvolution or duplex detection strategies and will broadly enable studies of RNA dynamic complexity.

DANCE-MaP does have several limitations. While our data and other studies (14, 18–20, 22, 39) support that multiple-hit DMS modification experiments accurately report native RNA structure, accumulated chemical damage may alter behavior of some RNAs. Working in collaboration with Dr. Anthony Mustoe, I show that DANCE-MaP can only resolve structural changes that involve >20 nucleotides with populations of at least ~10%, and currently has a length limit of ~500 nucleotides across a single strand of RNA. DANCE-MaP has a time-

resolution of roughly 5 minutes using DMS, but ~10 second resolution is possible with newer reagents (40). DANCE-MaP further requires that each read originate from a unique RNA molecule, and thus will be more challenging to implement for low abundance RNAs.

Ultimately, DANCE-MaP provides many of the same measurements previously accessible only using state-of-art the NMR experiments, which have provided the primary prior ground-truth reference for RNA ensembles and can be applied to short RNAs in cell-free contexts (41). The RINGs measured in 7SK State B are challenging to interpret due to the residual dynamics of this state, but such dynamics would similarly challenge established biophysical techniques. Uniquely, DANCE-MaP is readily performed in cells, interrogates large RNAs and RNA domains, and requires modest experimental effort. DANCE-MaP thus paves the way for a new generation of biophysical studies in living systems.

Methods

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 (LifeTech) at 37 °C and 5% CO_2.

DMS probing of 7SK RNA in cells

Jurkat 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, Omega BioTek; 1.8× bead:volume ratio) 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 (SPRI beads, Beckman), 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 $MgCl_2$] (22). 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 (**TABLE 2.4**) and amplified by PCR [Q5 HotStart polymerase (NEB), supplemented with 1.0 M betaine]. DNA templates were purified (SPRI beads, Mag-Bind TotalPure NGS; 0.7× bead:volume 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

G-Block	Sequence
Native 7SK sequence	GAAATTAATACGACTCACTATAGGGATGTGAGGGCGATCTGGCTGCGACATCTGTCACC CCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTTCCTC CCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACCA TCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTG CTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAGC TTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTTTT
7SK M1	GAAATTAATACGACTCACTATAGGGATGTGAGGGCGATCTGGCTGCGAC**TAG**TGTCACC CCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTTCCTC CCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACCA TCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTG CTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAGC TTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTTTT
7SK M1+M2	GAAATTAATACGACTCACTATAGGGATGTGAGGGCGATCTGGCTGCGAC**TAG**TGTCACC CCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGCTAGGCG**CTA**GTCCCCTTCCTC CCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACCA TCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTG CTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAGC TTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTTTT
7SK M3	GAAATTAATACGACTCACTATAGGGATGTGAGGGCGATCTGGCTGCGACATCTGTCACC CCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTTCCTC CCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACCA TCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTG CTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAGC TTCCAAGACTCCAGAA**TGTAAA**AAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTTTT
7SK M4	GAAATTAATACGACTCACTATAGGGATGTGAGGGCGATCTGGCTGCGACATCTGTCACC CCATTGATCGCCAGGGTTGATTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCC**TAC**CCT CCCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACC ATCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCT GCTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGGTAGTCAAG CTTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGTGGCAGTCTGCCTTTCTTTT
Primer	**Sequence**
7SK-Template-F	GAAATTAATACGACTCACTATAGGGATGTGAG
7SK-Template-R	AmAAAGAAAGGCAGACTGCCAC

TABLE 2.4: IVT sequences for gene block amplification. Mutations are in **bold** and primer binding regions are underlined. mA is a 2′-OMe Adenosine.

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 (SPRI beads, Mag-Bind TotalPure NGS, Omega BioTek; 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 EtOH), 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 EtOH for the DMS solution.

MaP reverse transcription

Mutational profiling (MaP) reverse transcription (RT) was performed exactly as described (7, 22). 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. Primer sequences were designed to bind to the exact ends of the RNA (**TABLE 2.5**). RT products were purified (Mag-Bind TotalPure NGS beads, Omega BioTek; 1.8× ratio) or G-50 Sephadex columns (Cytiva).

Sequencing library construction

Sequencing libraries were generated using the two-step PCR approach (42). 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

Primer	Sequnce
7SK-RT	AAAAGAAAGGCAGACTGCCAC
7SK-PCR1-F	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNN GGATGTGAGGGCGATCTG
7SK-PCR1-R	CCCTACACGACGCTCTTCCGATCTNNNNNAAAAGAAA GGCAGACTGCCACATG

TABLE 2.5: Primers for 7SK. Sequences are designed for 2-step amplicon PCR.

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.

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. 7SK data were aligned against NR_001445.2. DANCE-MaP analysis was performed using the *DanceMapper* (v1.0) software. For 7SK, *DanceMapper* was run allowing a maximum of 3 clusters (--maxc=3). PAIR and RING analyses were performed via *DanceMapper* using default options.

Structure modeling

Structure modeling was performed using *RNAstructure* (v6.2) (43). The *partition* module was modified to enable DMS-guided pairing probably calculations using nucleotide- specific DMS reactivity restraint functions (22); this modified code is available from the authors 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 2.10** 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*, the script *foldClusters*.py is provided

that automates structure modeling and visualization for all states of a deconvoluted ensemble.

DANCE-MaP clustering

Data are fit to a Bernoulli mixture model using the expectation-maximization (EM) algorithm (44). Fitting is performed for sequentially larger numbers of model components (structural states), beginning with 1, until the best fit is identified. Given a converged Bernoulli mixture model, individual reads can be assigned to a 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 (18, 22).

REFERERENCES

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

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

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

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

5. P. S. Ray, *et al.*, A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915–919 (2009).

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

7. A. Sengupta, G. M. Rice, K. M. Weeks, 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**, 1–19 (2019).

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

9. M. E. Wilkinson, C. Charenton, K. Nagai, RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).

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

11. B. J. Krueger, K. Varzavand, J. J. Cooper, D. H. Price, 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** (2010).

12. J. E. Brogie, D. H. Price, Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res.*, 1–17 (2017).

13. R. A. Flynn, *et al.*, 7SK-BAF axis controls pervasive transcription at enhancers. *Nat. Struct. Mol. Biol.* **23**, 231–238 (2016).

14. L. Luo, *et al.*, HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions. *J. Mol. Biol.*, 166885 (2021).

15. M. Marz, *et al.*, Evolution of 7SK RNA and its protein partners in metazoa. *Mol. Biol.*

Evol. **26**, 2821–2830 (2009).

16. I. Kalvari, *et al.*, Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **49**, D192–D200 (2021).

17. E. Rivas, J. Clements, S. R. Eddy, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45–48 (2016).

18. P. J. Homan, *et al.*, Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 13858–63 (2014).

19. P. J. Tomezsko, *et al.*, Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **582**, 438–442 (2020).

20. E. Morandi, *et al.*, Genome-scale deconvolution of RNA structure ensembles. *Nat. Methods* **18** (2021).

21. A. Krokhotin, A. M. Mustoe, K. M. Weeks, N. V Dokholyan, Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* (2016).

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

23. E. A. Dethoff, *et al.*, Pervasive tertiary structure in the dengue virus RNA genome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 11513–11518 (2018).

24. A. Reining, *et al.*, Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355–359 (2013).

25. S. Warhaut, *et al.*, Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. *Nucleic Acids Res.* **45**, 5512–5522 (2017).

26. S. Tian, W. Kladwang, R. Das, Allosteric Mechanism of the V. vulnificus Adenine Riboswitch Resolved by Four-dimensional Chemical Mapping. *Elife*, 1–36 (2018).

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

28. C. E. Hajdin, *et al.*, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5498–5503 (2013).

29. I. Lebars, *et al.*, HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. *Nucleic Acids Res.* **38**, 7749–7763 (2010).

30. D. Martinez-Zapien, *et al.*, The crystal structure of the 5’ functional domain of the

transcription riboregulator 7SK. *Nucleic Acids Res.* **45**, 3568–3579 (2016).

31. S. Egloff, E. Van Herreweghe, T. Kiss, Regulation of Polymerase II Transcription by 7SK snRNA: Two Distinct RNA Elements Direct P-TEFb and HEXIM1 Binding. *Mol. Cell. Biol.* **26**, 630–642 (2006).

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

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

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

35. X. Ji, *et al.*, SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. *Cell* **153**, 855–868 (2013).

36. E. Van Herreweghe, *et al.*, Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.* **26**, 3570–3580 (2007).

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

38. C. Jeronimo, *et al.*, 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 (2007).

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

40. J. E. Ehrhardt, K. M. Weeks, Time-Resolved, Single-Molecule, Correlated Chemical Probing of RNA. *J. Am. Chem. Soc.* **142**, 18735–18740 (2020).

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

42. M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, 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 (2015).

43. J. S. Reuter, D. H. Mathews, RNAstructure: Web servers for RNA secondary structure prediction and analysis. *BMC Bioinformatics* (2010) https:/doi.org/10.1093/nar/gkt290.

44. C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006) https:/doi.org/10.1007/978-3-030-57077-4_11.

CHAPTER 3: LARGE-SCALE ALLOSTERIC SWITCH IN THE HUMAN 7SK RNA SUPPORTS APPROACH FOR TARGETING RNA STRUCTURE WITH ANTISENSE OLIGONUCLEOTIDES

Introduction

The 7SK RNA is an abundant 332 nucleotide long non-coding RNA, forms the key architectural component of the 7SK small non-coding RNA-protein complex (snRNP), and serves as a major nexus of transcriptional control (1, 2). 7SK is canonically thought to function by sequestering and inhibiting Cdk9/Cyclin T1 (together termed positive transcription elongation factor b, P-TEFb), a kinase required for phosphorylation and release of RNA polymerase II (Pol II) complexes paused at promoter-proximal regions. The P-TEFb-free form of 7SK appears to play additional roles in facilitating productive elongation (1, 2), including modulating splicing (3–5) and chromatin remodeling (6, 7). These diverse functions of the 7SK snRNP are driven by coordinated changes to 7SK protein components (8). P-TEFb is sequestered via interactions with the accessory protein dimer hexamethylene bis-acetamide inducible protein 1 or 2 (HEXIM1/2), which binds at a high-affinity stem-loop structure, SL1, in the 7SK RNA (9–11). Other core members of the inhibitory 7SK snRNP include methylphosphate capping enzyme (MePCE) (12) and La related protein 7 (LaRP7) (13, 14), which stabilize the RNP. Under transcription stimulatory conditions, P-TEFb and HEXIM1/2 are liberated from 7SK by various release factors, including the bromodomain protein Brd4, and several helicases (1, 2). The P-TEFb-free form of 7SK is in turn bound by diverse heterogenous ribonucleoproteins (hnRNPs), splicing factors, and chromatin-remodelers that interact with sites at the 3' end the 7SK RNA (1, 2).

However, this remodeling process remains poorly understood. Defining 7SK regulatory mechanisms will both illuminate fundamental aspects of transcriptional control and also inform ongoing efforts to inhibit transcription in disease settings, especially cancer (15), and, conversely, to activate transcription as part of "kick-and-kill" HIV therapies (16, 17).

Using DANCE-MaP, I showed in Chapter 2 that the 7SK RNA contains 2 macrostates state A and state B. State A is most likely to be bound by P-TEFb because it contains the HEXIM 1/2 recognition site, agreement with previous binding studies (11, 18–21). Whereas state B contains a truncated SL1alt and should have weak if any binding to HEXIM and so it would be considered the P-TEFb release state (9, 21). Leveraging this information a prototype, anti-sense oligonucleotide (ASO) based, strategy was designed that alters the 7SK ensemble and upregulates transcription in cells. This work establishes DANCE-MaP as a powerful framework for directly resolving complex ensembles in cells and explains diverse features of 7SK non-coding RNA biology.

Results

The 7SK structural equilibrium is regulated by cell type and responds to transcriptional stress

The 7SK-P-TEFb axis is a global regulator of transcription, a process likely to be heavily regulated by the gene expression and growth needs of a cell. Therefore an exploration on whether the 7SK equilibrium is regulated based on cell type and state is essential. Jurkat cells, like many other tumor cells, exhibit altered P-TEFb regulation and aberrantly upregulated transcription compared to primary cells (22, 23). In collaboration with Dr. Anthony Mustoe we examined whether the 7SK ensemble differs in normal, non-transformed cells, using human RPE-1 cells (24) as a model. DANCE-MaP experiments performed on living RPE-1 cells

revealed that 7SK adopts precisely the same three structures as observed for Jurkat cells (**FIGURE 3.1**) but with significantly different populations: states A, B, and H have populations of 47 ± 2, 33 ± 2, and 20% ± 1, respectively, in RPE-1 cells compared to 39 ± 3, 46 ± 2, 15% ± 2 in Jurkat cells (**FIGURE 3.2**). This shift represents a ~20% relative increase in the fraction of 7SK competent to bind P-TEFb. In absolute molecular terms, ~10,000 7SK snRNPs per cell have shifted conformation to a P-TEFb binding state (25, 26). These data are consistent with P-TEFb being more sequestered in RPE-1 cells (or conversely, aberrantly released in Jurkat tumor cells), and establish that the 7SK structural equilibrium is regulated in a cell-type-specific manner.

7SK and P-TEFb are also dynamically regulated in response to transcriptional stress (1, 2). Flavopiridol is a pan-CDK inhibitor that suppresses transcription by inhibiting CDK9 (27). To compensate for reduced CDK9 activity, flavopiridol induces cells to release P-TEFb from 7SK, which conventional probing experiments have indicated induces structural changes in the 7SK RNA (8, 28). In collaboration with Dr. Anthony Mustoe we directly visualized these structural changes by performing DANCE-MaP experiments in Jurkat and RPE-1 cells treated with either vehicle (DMSO, 0.01%) or 1 μM flavopiridol (saturating concentration) (28). In both cell types, flavopiridol treatment dramatically remodels the 7SK structural ensemble. In Jurkat cells, state A is completely converted to B/B-like states (**TABLE 3.1**), consistent with total P-TEFb release (**FIGURE 3.3, 3.4A**). However, in RPE-1 cells, state A is significantly, but signifigantly depopulated (47% to 26%). Further, instead of increasing the population of B, flavopiridol stimulates conversion to state H (also a non-P-TEFb-binding state) (**FIGURE 3.3, 3.4B**). This switch supports the physiological relevance of state H, indicates that P-TEFb release is governed multiple cell-type-dependent pathways, and reveals a prominent role for cell type-

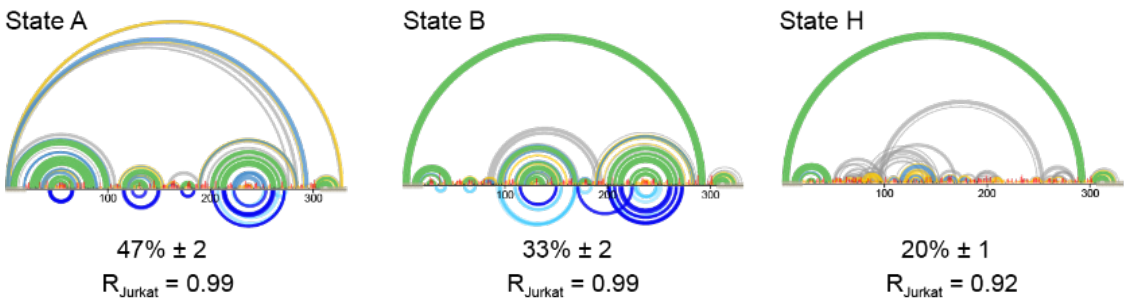

FIGURE 3.1: 7SK structural ensemble in proliferating RPE-1 cells resolved by DANCE-MaP. Structure models for 7SK in RPE-1 cells are shown. Pearson's R, comparing RPE-1 and Jurkat cell reactivities, are shown.

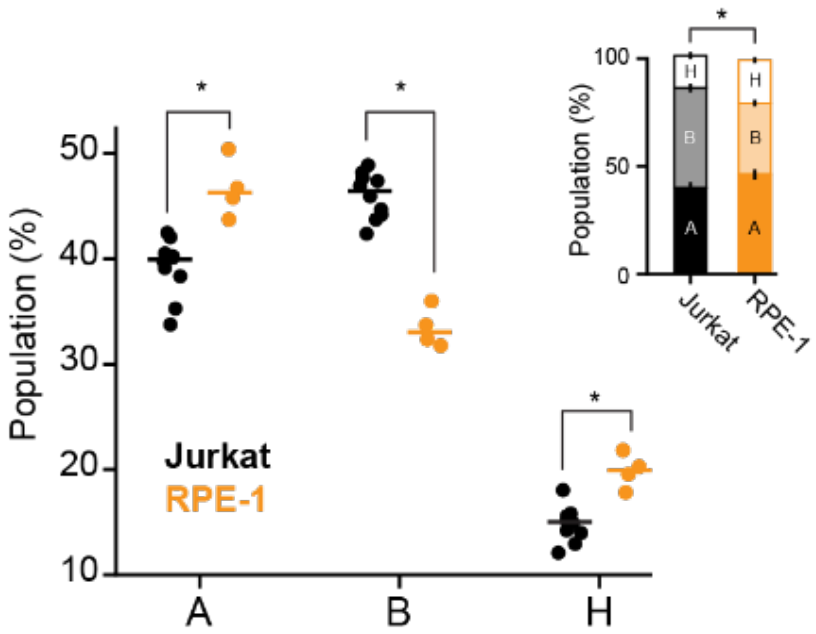

FIGURE 3.2: 7SK ensemble populations for Jurkat tumor cells and proliferating and quiescent normal RPE-1 cells. Comparisons between individual state populations were evaluated using a two-sided Mann-Whitney U test. Inset, population data shown in complete distribution format. Comparisons between complete ensembles were performed using a Dirichlet likelihood ratio test (29). N = 10, 7, and 3 for Jurkat, proliferating RPE-1, and quiescent RPE-1 cells, respectively.

Sample	A%	B%	M%	number of reads
DMSO				
1	46	44	10	396,792
2	43	48	9	577,110
Mean (std)	45 (±2)	46 (±3)	9 (±1)	
Consolidated	45	46	9	973,935
(+) Flavopiridol				
1		50, 36	14	413,497
2		53, 32	15	435,550
Mean (std)		86 (±1)	14 (±1)	
Consolidated		51, 36	13	849,401

TABLE 3.1: Flavopiridol treated Jurkat cell replicates. Populations and standard deviation are based on 2 independent replicates.

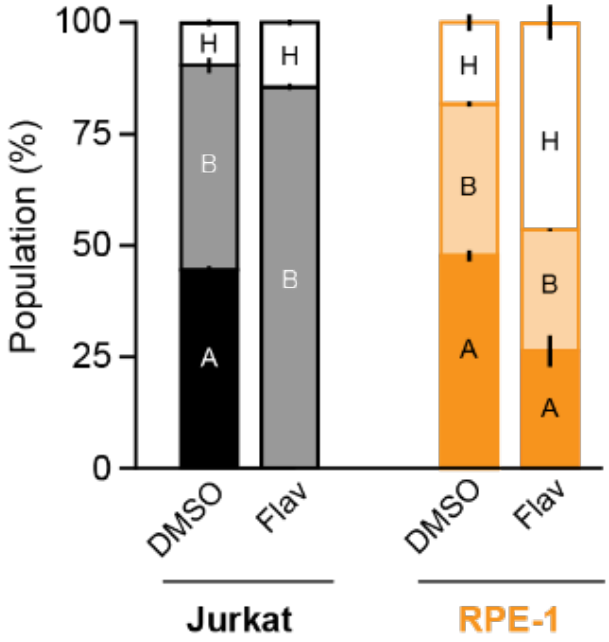

FIGURE 3.3: Shift in 7SK equilibrium upon flavopiridol treatment in Jurkat and RPE-1 cells. Cells were treated with vehicle (DMSO, 0.01%) or 1 μM flavopiridol for 1 hr (n = 2 for all experiments).

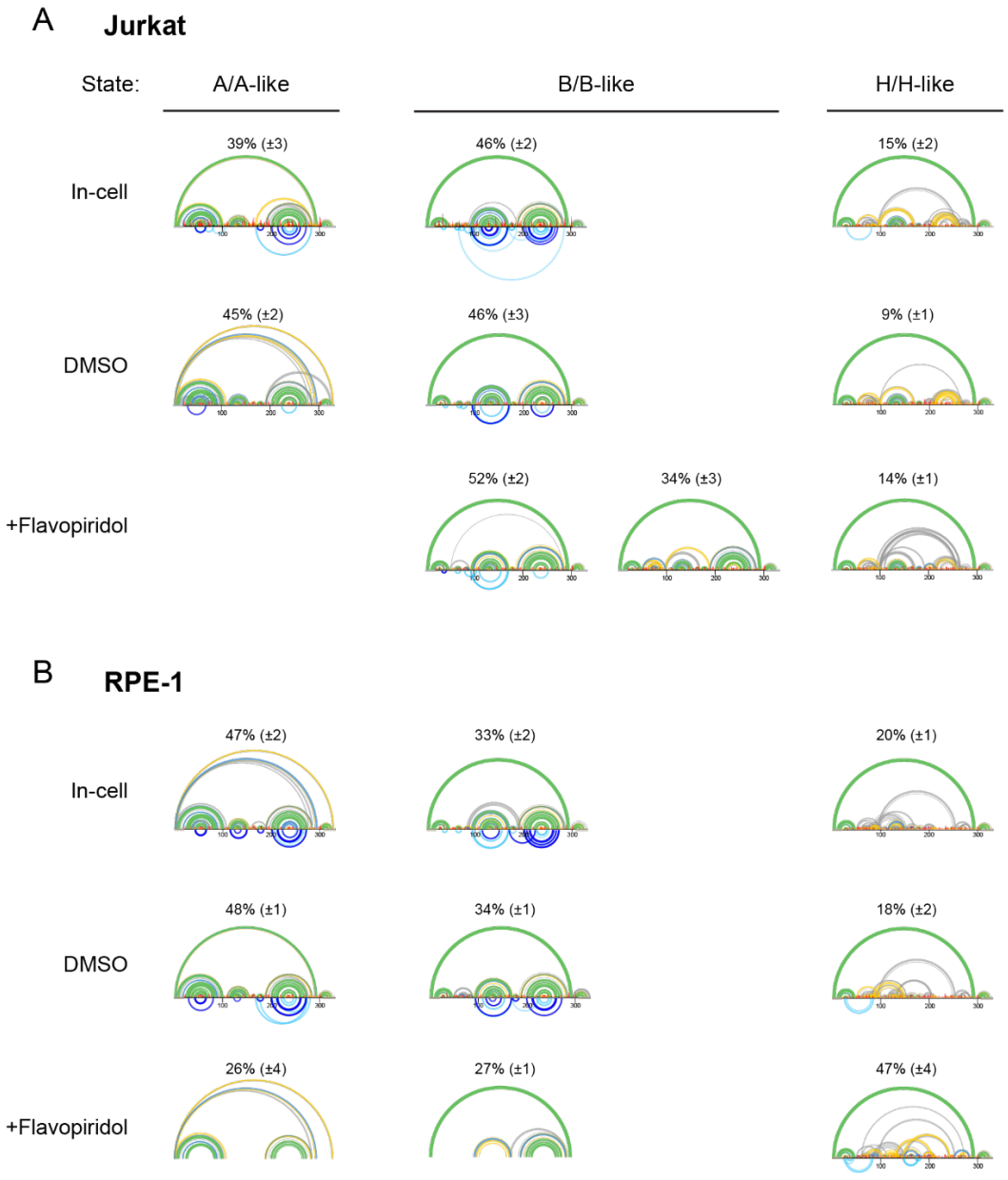

FIGURE 3.4: DANCE deconvolution of Flavopiridol treated Jurkat and RPE-1 cells. DANCE deconvolution of 7SK states in (**A**) Jurkat or (**B**) RPE-1 cells treated with vehicle (0.01% DMSO) or 1 μM flavopiridol. Structure models are representative of two independent replicates. Population errors represent standard deviations. In-cell Jurkat data from **TABLE 2.1** as reference.

specific RNA structure. The attenuated 7SK response to flavopiridol is further consistent with the greater tolerance of RPE-1 cells to transcription inhibition as compared Jurkat cells, which have strong transcription addictions (30). Collectively, these data establish that the 7SK conformational equilibrium is tunable, is cell type-specific, and remodels dynamically, coincident with P-TEFb release.

ASO stabilization of state B induces transcription in cells

This data emphasize that the 7SK structural switch is critically linked to P-TEFb release and transcription regulation, motivating us to explore the potential of targeting the 7SK ensemble as a strategy for controlling transcription. After screening multiple candidates, I identified a modified antisense oligonucleotide, ASO-B, that disrupts state A without impacting the major helices unique to state B (**FIGURE 3.5**). DANCE-MaP experiments on total cell-free RNA confirmed that ASO-B shifted the 7SK structural ensemble to exclusively B/B-like states (**TABLE 3.2**), whereas a control ASO containing five central mismatches (MM-B) had no significant impact on the 7SK ensemble (**FIGURE 3.6**). Thus, ASO-B constitutes a molecular tool for modulating the 7SK structural ensemble.

In collaboration with Dr. Anne-Marie Turner the ability of ASO-B to modulate transcription in cells was tested. First the delivery in HEK293 cells was tested using a gapmer oligonucleotide (GAP-B) targeting the same region of 7SK but designed to induce RNase H degradation. Both the cellular levels of 7SK and HEXIM1 RNAs were examined. HEXIM1 expression provides a sensitive measure of P-TEFb-mediated transcription activity (31, 32).Treatment with GAP-B reduced 7SK levels by 31 and 36% at 100 nM and 200 nM concentrations, respectively (**FIGURE 3.7A**). 7SK depletion increases the fraction of free

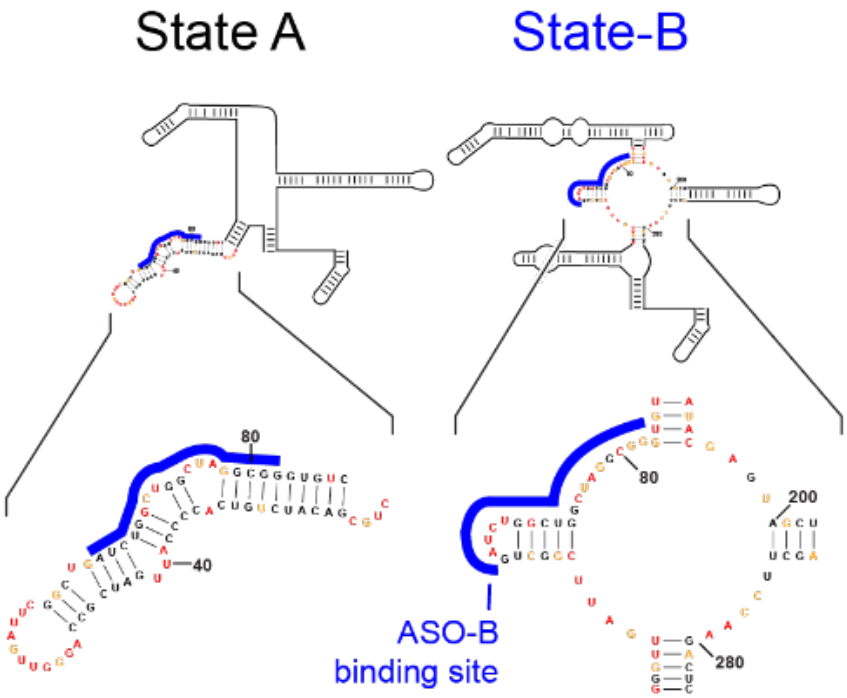

FIGURE 3.5: ASO-B design. ASO-B binding site shown superimposed on secondary structure models of 7SK states A and B.

Sample	A%	B/B-like %	M%	number of reads
ASO-B				
1		69, 24	7	191,463
2		66	34	374,261
3		60, 21	19	325,245
Mean (std)		80 (± 14)	20 (± 14)	
Consolidated		63, 17	20	898,739
MM-B				
1	63	37		215,674
2	45	55		408,439
3	29	48	23	322,121
Mean (std)	46 (± 17)	47 (± 9)	7 (± 13)	
Consolidated	36	30	34	952,241

TABLE 3.2: ASO-B treated cell-free replicates. Populations and standard deviation are based on 2 independent replicates.

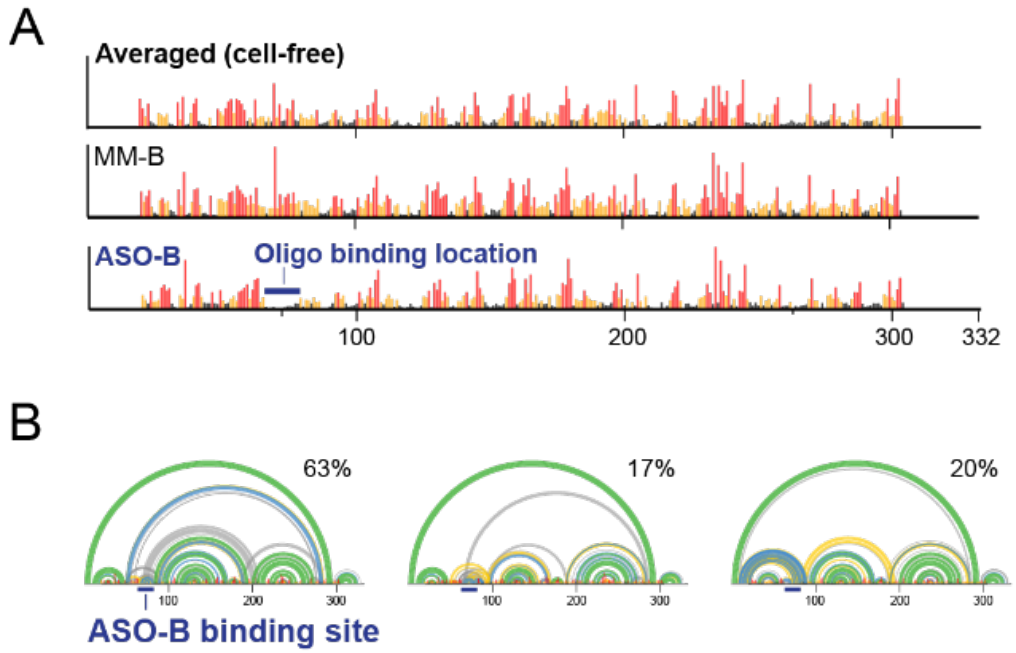

FIGURE 3.6: ASO-B engagement and stabilization of state B. (**A**) Engagement of ASO-B on the 7SK RNA observed as a complete reduction of per-nucleotide reactivity at the ASO binding site. (**B**) DANCE deconvolution reveals ASO-B-induced states all correspond to B/B-like states.

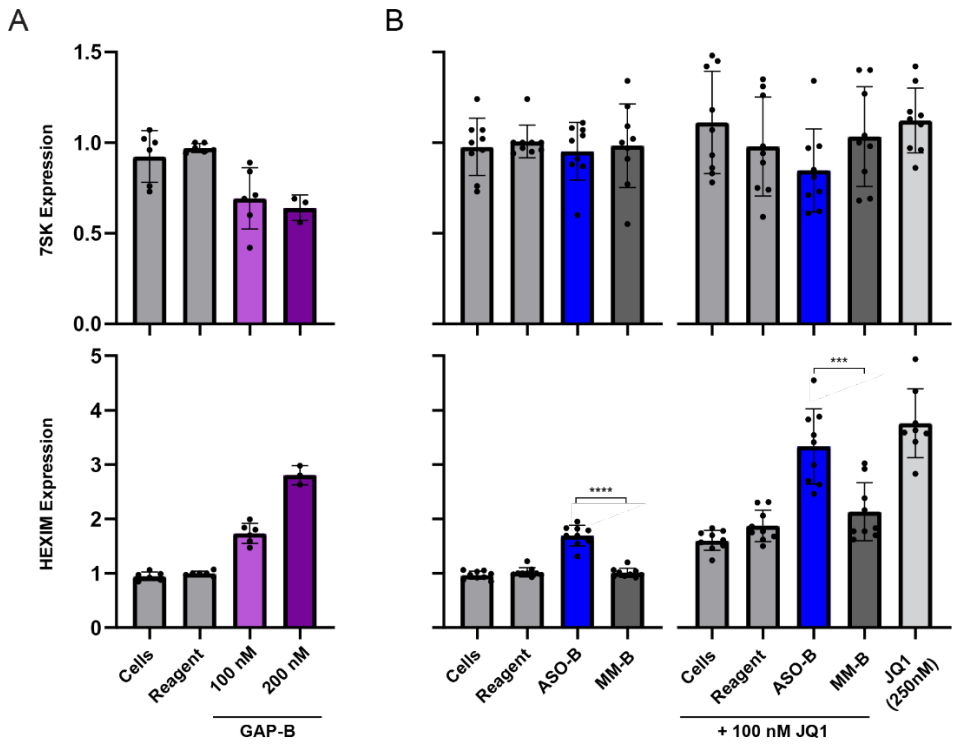

FIGURE 3.7: Stabilization of 7SK state B induces transcription. 7SK and HEXIM RNA levels measured by RT-qPCR. HEK293 cells were treated for 24 hours with oligos prior to total RNA extraction and quantification. Data normalized to the control gene TBP, in three triplicate experiments. (**A**) GAP-B induces a decrease in 7SK expression (top). ** p-value = 0.0097 and * p-value = 0.0108 for 100 nM and 200 nM Gapmer concentrations respectively compared to reagent alone. GAP-B induces a dose-dependent increase in 7SK expression (bottom). *** p-value = 0.001 and ** p-value = 0.0024 for 100 nM and 200 nM Gapmer concentrations respectively compared to reagent alone. (**B**) 7SK expression shows (*top*) no change when treated with ASO-B and MM-B oligos and JQ1. HEXIM shows (*bottom*) significant increase in expression for ASO-B compared to MM-B. **** p-value = <0.0001 and ASO-B (+100 nM JQ1) compared to MM-B (+100 nM JQ1), *** p-value = 0.0009; all significance was determined using a Welch's t-test.

(active) P-TEFb, and simultaneously resulted in an 1.7- and 2.6-fold induction of HEXIM1 mRNA expression (**FIGURE 3.7A**). These data confirm in-cell engagement with 7SK by the ASO-B sequence. The ability of ASO-B to induce P-TEFb release via stabilization of 7SK state B was tested next. Treatment with 100 nM ASO-B but not MM-B yielded a 1.7-fold increase in HEXIM1 without impacting 7SK expression (**FIGURE 3.7B**), directly validating that 7SK switching induces P-TEFb release. Next, ASO-mediated structure-switching in tandem with transcriptional activation by the small molecule JQ1 was tested for complementarity. JQ1 is a bromodomain and extra-terminal (BET) inhibitor that induces P-TEFb release via a 7SK-independent mechanism (33, 34). Indeed, co-addition of 100 nM JQ1 increased HEXIM1 expression 3.3-fold, comparable to the upregulation observed upon treatment with 250 nM JQ1 alone (**FIGURE 3.7B**). All HEXIM analysis was done with three separate control genes (TBP, RPL13a, and GAPDH) yielding similar results (**FIGURE 3.7B, 3.8**). Together, these experiments establish a direct, causal relationship between 7SK structural switching and P-TEFb release and provide proof of principle for targeting the 7SK structural switch.

Discussion

Allostery couples a 7SK HEXIM1-P-TEFb aptamer domain to release factor binding sites

Regulated release of P-TEFb from the 7SK snRNP to phosphorylate Pol II is a critical control point in transcription (2). DANCE-MaP was used to show that the 7SK RNA intrinsically encodes a large-scale structural switch that modulates its P-TEFb binding ability (**FIGURE 3.9**). This study further shows that the 7SK structural equilibrium is actively controlled by the cell, with normal cells favoring a P-TEFb-sequestered state compared to transformed cells, and transcriptional stress (CDK9 inhibition) favoring a P-TEFb released state (**FIGURE 3.2**).

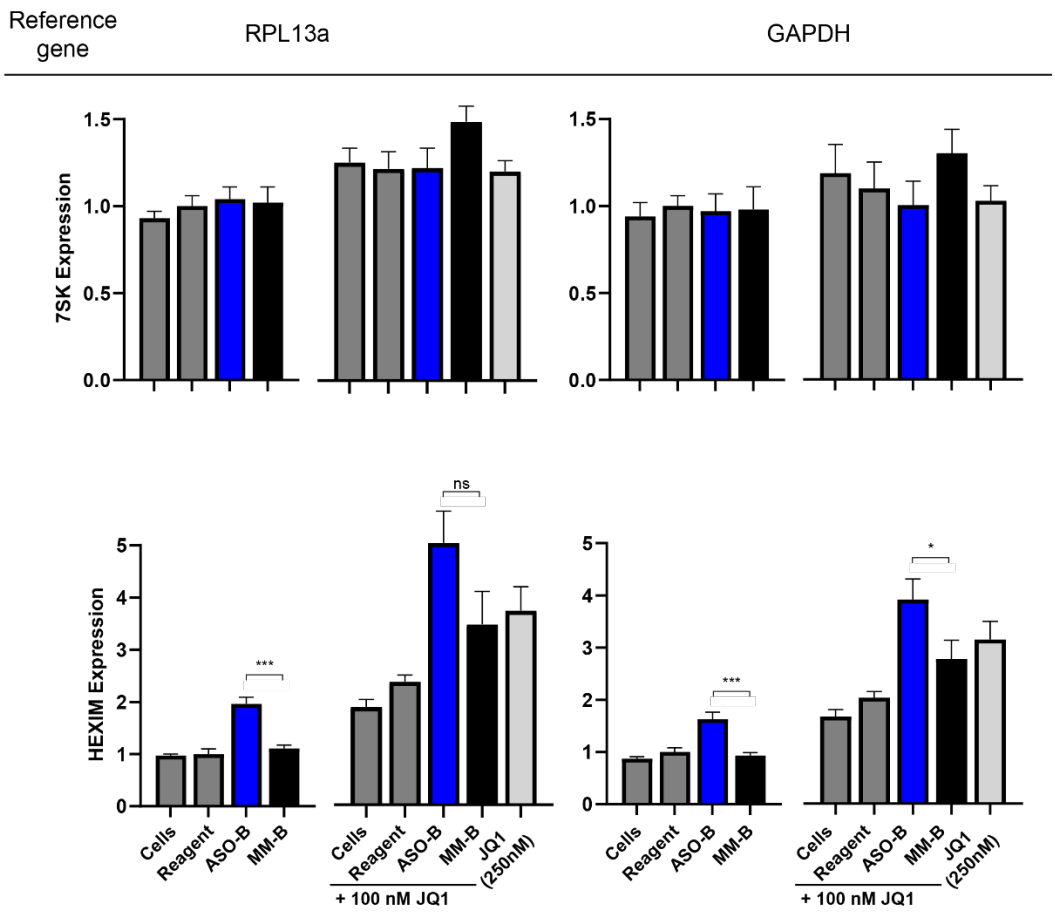

FIGURE 3.8: Stabilization of 7SK state B induces transcription with multiple reference genes. 7SK and HEXIM RNA levels measured by RT-qPCR. HEK293 cells were treated with ASOs for 24 hours with oligos prior to total RNA extraction and quantification, and normalized to the control genes RPL13a and or GAPDH, in three triplicate experiments. 7SK expression shows (*top*) no change when treated with ASO-B and MM-B oligos and JQ1. Using RPL13a as the control gene HEXIM shows (*bottom, left*) significant increase in expression for ASO-B compared to MM-B. *** p-value = 0.0001. Using GAPDH as the control gene HEXIM shows (*bottom, right*) significant increase in expression for ASO-B compared to MM-B. *** p-value = 0.0005 and ASO-B (+100 nM JQ1) compared to MM-B (+100 nM JQ1), * p-value = 0.0484; all significance was values were determined using a Welch's t-test.

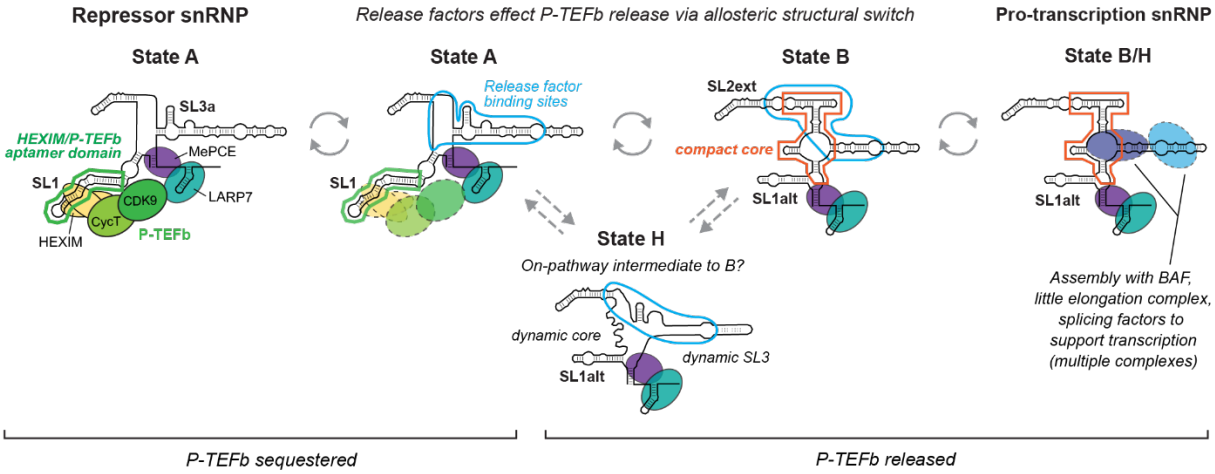

FIGURE 3.9: The in-cell 7SK ensemble as a dual-function signal integrator. States A, B, H are shown as schematic secondary structures. Compact core corresponds to region of dense RINGs. Annotations show binding sites for core (P-TEFb, HEXIM1/2, MePCE and LaRP7) and transient binding (release factors, helicase, chromatin-specific) factors.

Triggering the 7SK switch via an ASO that stabilizes State B induces transcription of P-TEFb target genes (**FIGURE 3.6**). Collectively, these data support that the 7SK structural switch functions as an important axis of transcriptional regulation.

P-TEFb binds to the SL1 hairpin of 7SK via the accessory protein dimer HEXIM1/2 (HEXIM/P-TEFb aptamer domain; Fig 3.9) (11, 13, 20, 35) . The SL1 aptamer adopts distinct structures depending on whether or not HEXIM/P-TEFb are bound (8, 19, 36), but the broader structural context of this conformational change and its mechanistic relationship to P-TEFb release remained unclear. The data directly reveal, in-cells, that the SL1 aptamer domain exists in dynamic equilibrium with an alternative structure, SL1alt, that lacks HEXIM/P-TEFb binding ability (9, 21). This study further links dissolution of SL1 to formation of the novel SL2ext structure and 7SK compact core, hundreds of nucleotides away (**FIGURE 3.9**). Remarkably, the SL2ext and compact core structures directly overlap or are immediately adjacent to binding sites of known P-TEFb release factors (**FIGURE 3.9**), a group that includes both helicases (37–40) and general RNA binding proteins (39–42). How such factors induce release of P-TEFb despite binding distally to SL1 had been unclear. This data now support a model wherein these release factors effect P-TEFb release by allosterically remodeling SL1 by binding to and remodeling the central core (**FIGURE 3.9**). This model further rationalizes the involvement of multiple RNA helicases in P-TEFb release.

The allosteric mechanism of P-TEFb release unifies multiple prior observations, as allostery: (*i*) enables 7SK to maintain a specialized “release domain” that can integrate cellular signals unencumbered by bound P-TEFb; and (*ii*) prevents P-TEFb reassociation once release is triggered. The importance of role (*ii*) is specifically supported by the observation that HIV-1 Tat, despite directly abstracting P-TEFb without binding the 7SK release domain, also induces

conformational changes in 7SK (8), which can now be interpreted as switching from state A to B. Less is known regarding how 7SK resequesters P-TEFb. One possible mechanism is that helicases stimulate disassociation of hnRNPs and remodel 7SK to state A (the SL1-containing form), enabling HEXIM/P-TEFb to rebind. Given that 7SK may be involved in transcription termination (43), this process may be linked to Pol II recycling. In this model, distinct sets of helicases and other RNA binding proteins catalyze 7SK switching between structural states to either sequester or release P-TEFb.

The allosteric switching model also rationalizes the extreme sequence conservation of the first ~100 7SK nucleotides across vertebrates and invertebrates (44), which must preserve both HEXIM/P-TEFb binding and the dual constraints of forming the distinct SL1 and SL1alt structures. By comparison, the compact 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, variable core in 7SK, compact tertiary structure in one of two states, recalls classic riboswitches, where conserved aptamer domains are often integrated into diverse expression domain architectures (45).

This data support a model whereby release factors function to catalyze 7SK structural switching and thereby allosterically effect P-TEFb release (**FIGURE 3.9**). Significantly, release factor binding sites directly overlap or are immediately adjacent to the SL2ext and compact core structures in state B, ideally positioning them to effect structure switching. Given that the unstructured regions of state H also overlap these release factor binding sites, state H may represent an intermediate along the A to B pathway.

This allosteric model of P-TEFb release unifies multiple prior observations, as allostery:

(*i*) enables 7SK to maintain a specialized "release domain" that can integrate cellular signals unencumbered by bound P-TEFb; and (*ii*) prevents P-TEFb reassociation once release is triggered. The importance of role (*ii*) is supported by the observation that HIV-1 Tat, despite directly abstracting P-TEFb without binding the 7SK release domain, also induces conformational changes in 7SK (8), which can now be interpreted as switching from state A to B. Less is known regarding how 7SK resequesters P-TEFb. One possible mechanism is that helicases stimulate disassociation of hnRNPs and remodel 7SK to state A (the SL1-containing form), enabling HEXIM/P-TEFb to rebind. Given that 7SK may be involved in transcription termination (43), this process may be linked to Pol II recycling. In this model, distinct sets of helicases and other RNA binding proteins catalyze 7SK switching between structural states to either sequester or release P-TEFb.

7SK structural switch links P-TEFb release to pro-transcription functions

7SK is canonically considered a transcriptional repressor due to its P-TEFb sequestering function. However, the 7SK snRNP also has pro-transcription functions, including blocking convergent transcription via association with the BAF complex (7) and facilitating spliceosome production (4, 5). These pro-transcription functions are specific to 7SK snRNPs lacking P-TEFb. Notably, in light of this study, prior chemical probing data obtained for BAF-associated 7SK (7) can now clearly be interpreted as corresponding to state B or H. Together, these data support an overarching model in which the 7SK structural switch integrates P-TEFb release with conversion of 7SK into a pro-transcription snRNP that can scaffold assembly of elongation-supporting factors (**FIGURE 3.9, *right***). Significantly, in this model, 7SK switching would enable spatial and temporal coupling between Pol II pause release and BAF-mediated inhibition of convergent

transcription (7). The dual-function model also rationalizes observations that 7SK is inessential for basal P-TEFb regulation (46), but that 7SK depletion perturbs global chromatin structure (47) and compromises stress-induced transcriptional reprogramming (46).

Overall, the new 7SK model (**FIGURE 3.9**) emphasizes how structural switching enables the 7SK snRNP to integrate diverse signals to cooperatively inactivate or activate transcription in response to cellular demand. RNAs are unique among biomolecules in their ability to encode large but precise changes in structure (48, 49) and form internal aptamer domains, making RNAs optimally suited to serve as molecular integrators. Likely a similar switching mechanisms broadly underlie non-coding RNA regulatory function.

7SK switch constitutes a novel therapeutic target for modulating transcription

This study shows that the 7SK structural equilibrium is regulated in response to changing transcription needs. In collaboration with Dr. Anne-Marie Turner we further show using proof-of-principle ASO studies that switching the 7SK state can induce transcription of P-TEFb–sensitive targets. Developing small molecules or improved ASOs that stabilize state B, thereby release P-TEFb, and activate transcription represents a component of a promising strategy to eradicate persistent HIV infection, by inducing the expression of latent provirus (16, 17). Conversely, P-TEFb is commonly dysregulated in cancer and there is intense interest in developing pharmacological inhibitors of P-TEFb as a cancer therapeutic (50) . Disruption of the 7SK/P-TEFb regulatory axis has been linked to tumorigenesis and cancer progression (51–53) consistent with a model in which dysregulation of the 7SK structural equilibrium supports elevated transcription in cancer cells. Designing small molecules or ASOs that reduce the cellular availability of P-TEFb by selectively stabilizing 7SK state A represents a compelling

therapeutic hypothesis for targeting transcription in cancer.

Methods

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 (LifeTech) at 37 °C and 5% CO_2. . In collaboration with Dr. Anthony Mustoe 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. In collaboration with Dr. Anne-Marie Turner HEK293T/17 cells were obtained from ATCC (CRL-11268) and maintained in DMEM (LifeTech) supplemented with 10% FBS (Millipore) and 100 U/mL Pen/Strep at 37 °C and 5% CO_2.

DMS probing of 7SK RNA in cells

Jurkat 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, Omega BioTek; 1.8× bead:volume ratio) and quantified by UV absorbance (Nanodrop).

. In collaboration with Dr. Anthony Mustoe using RPE-1 cells, 1.5×10^6 cells were seeded into a 10 cm dish 48 hr 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 then 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, and RNA was extracted (RNeasy mini columns; Qiagen) and quantified by UV absorbance (Nanodrop).

DMS probing of flavipiridol treated cells

Jurkat cells (3 million cells in 10 mL fresh media) were seeded 23 hours prior to treatment and were then treated with either vehicle (0.01% DMSO) or with 1 µM flavopiridol (in DMSO) for 1 hour. In collaboration with Dr. Anthony Mustoe RPE-1 cells were seeded 23 hr prior to be 70% confluent on the day of experiment and were treated with 0.01% DMSO or 1 µM flavopiridol for 1 hour. 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 exactly as described (54, 55). Primers used are in **TABLE 2.5** as reference. 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 (Mag-Bind TotalPure NGS beads, Omega

BioTek; 1.8× ratio) or G-50 Sephadex columns (Cytiva).

Sequencing library construction

Sequencing libraries were generated using the two-step PCR approach (56). 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.

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 7SK, *DanceMapper* was run allowing a maximum of 3 clusters (--maxc=3). PAIR and RING analyses were performed via *DanceMapper* using default options.

Structure modeling

Structure modeling was performed using *RNAstructure* (v6.2) (57). The *partition* module was modified to enable DMS-guided pairing probably calculations using nucleotide-specific DMS reactivity restraint functions (54); this modified code is available from the authors 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 3.1, 3.4**, and **3.6** 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*, the script *foldClusters*.py is provided that automates structure modeling and visualization for all states of a deconvoluted ensemble.

ASO experiments

The ASO-B antisense oligonucleotide was designed to bind 7SK nts 64-82 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 with the following sequences:

ASO-B:

mC*mC*mG*mC*mC*mU*mA*mG*mC*mC*mA*mG*mC*mC*mA*mG*mA*mU*mC

MM-B:

mC*mC*mG*mC*mC*mU*mA*mC*mG*mG*mU*mC*mC*mC*mA*mG*mA*mU*mC

GAP-B: mC*mU*A*G*C*C*A*G*C*C*A*G*A*mU*mC

(m: 2'-O-Methyl RNA base; *: phosphorothioate backbone)

ASO engagement with 7SK was confirmed by 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 $MgCl_2$] (54) 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.

In collaboration with Dr. Anne-Marie Turner HEK293T cells were seeded at 30,000 cells/well in a 96-well flat bottom plate 24 hrs 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 hrs post transfection for a total incubation time of 20 hrs. 250 nM JQ1 was added for 24 hrs. After 24 hrs, 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 hrs prior to RNA isolation.

Gene expression analysis

In collaboration with Anne-Marie Turner total RNA was isolated (Quick RNA 96-well; Zymo) and cDNA was generated (Maxima First Strand cDNA Synthesis Kit for RT-qPCR; with dsDNase, ThermoFisher). 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 3.3**. 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 3.4**) designed to mimic the target amplicon. Expression was standardized to indicated control genes using the Pfaffl method (58). Data in **FIGURE 3.7A** corresponds to three biological replicates from 2 independent experiment (n=6), except the 200 nM GAP-B sample

qPCR Primer	Sequence
7SK-F	CCTGCTAGAACCTCCAAACAA
7SK-R	GGAGTCTTGGAAGCTTGACTAC
HEXIM1-F	CCGAGGCCAGTAAGTTGGG
HEXIM1-R	GACGGGCGTCTCCTATGTTT
TBP-F	GAGAGTTCTGGGATTGTACCG
TBP-R	ATCCTCATGATTACCGCAGC
RPL13a-F	GCCTACAAGAAAGTTTGCCTATC
RPL13a-R	TGGCTTTCTCTTTCCTCTTCTC
GAPDH-F	GTCAACGGATTTGGTCGTATTG
GAPDH-R	TGTAGTTGAGGTCAATGAAGGG

TABLE 3.3: qPCR primer sequences.

gBlock	Sequence
7SK-qPCR	GGATGTGAGGGCGATCTGGCTGCGACATCTGTCACCCCATTGATCGCCAGGG TTGATTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTTCCTCCCTCACCGC TCCATGTGCGTCCCTCCCGAAGCTGCGCGCTCGGTCGAAGAGGACGACCATC CCCGATAGAGGAGGACCGGTCTTCGGTCAAGGGTATACGAGTAGCTGCGCTC CCCTGCTAGAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTA GGGTAGTCAAGCTTCCAAGACTCCAGACACATCCAAATGAGGCGCTGCATGT GGCAGTCTGCCTTTCTTTT
HEXIM-qPCR	AGCCTTGTCATGACTCCGAGGCCAGTAAGTTGGGGGCTCCTGCCGCAGGGGG CGAAGAGGAGTGGGGACAGCAGCAGAGACAGCTGGGGAAGAAAAAACATA GGAGACGCCCGTCCAAGAAGAAGC
TBP-qPCR	GCCAGCTTCGGAGAGTTCTGGGATTGTACCGCAGCTGCAAAATATTGTATCC ACAGTGAATCTTGGTTGTAAACTTGACCTAAAGACCATTGCACTTCGTGCCCG AAACGCCGAATATAATCC CAAGCGGTTTGCTGCGGTAATCATGAGGATAAGAGAGCCA
RPL13a-qPCR	ATCCCACCGCCCTACGACAAGAAAAAGCGGATGGTGGTTCCTGCTGCCCTCA AGGTCGTGCGTCTGAAGCCTACAAGAAAGTTTGCCTATCTGGGGCGCCTGGC TCACGAGGTTGGCTGGAAGTACCAGGCAGTGACAGCCACCCTGGAGGAGAA GAGGAAAGAGAAAGCCAAGATCCACTACCGGAAGAAGAAACAGCTCATGAG GCTACGGAAACAGGCCGAGAAGAACGTGGAGAAGAAAATTGACAAATACAC AGAGGTCCTCAAGACCCACGGACTCCTGGTC
GAPDH-qPCR	GAAGGTCGGAGTCAACGGATTTGGTCGTATTGGGCGCCTGGTCACCAGGGCT GCTTTTAACTCTGGTAAAGTGGATATTGTTGCCATCAATGACCCCTTCATTGA CCTCAACTACATGGTTTACATGTTCCAATATGATTCCACCCATGGCAAATTCC ATGGCACCGTCAAGGCTGAGAACGGGAAGCTTGTCATCAATGGAAAT

TABLE 3.4: gBlock sequences for qPCR standardization.

which corresponds to three biological replicates from one independent experiment (n=3). Data in **FIGURE 3.7B** correspond to three biological replicates from 3 independent experiments (n=9).

REFERECES

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

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

3. M. Barboric, *et al.*, 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7798–7803 (2009).

4. S. Egloff, *et al.*, The 7SK snRNP associates with the little elongation complex to promote snRNA gene expression. *EMBO J.* **36**, 934–948 (2017).

5. C. Ji, *et al.*, Interaction of 7SK with the Smn complex modulates snRNP production. *Nat. Commun.* **12** (2021).

6. S. Eilebrecht, *et al.*, 7SK small nuclear RNA directly affects HMGA1 function in transcription regulation. *Nucleic Acids Res.* **39**, 2057–2072 (2011).

7. R. A. Flynn, *et al.*, 7SK-BAF axis controls pervasive transcription at enhancers. *Nat. Struct. Mol. Biol.* **23**, 231–238 (2016).

8. B. J. Krueger, K. Varzavand, J. J. Cooper, D. H. Price, 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** (2010).

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

10. B. M. Peterlin, D. H. Price, Controlling the Elongation Phase of Transcription with P-TEFb. *Mol. Cell* **23**, 297–305 (2006).

11. D. Martinez-Zapien, *et al.*, The crystal structure of the 5' functional domain of the transcription riboregulator 7SK. *Nucleic Acids Res.* **45**, 3568–3579 (2016).

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

13. B. J. Krueger, *et al.*, 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 (2008).

14. C. D. Eichhorn, Y. Yang, L. Repeta, J. Feigon, Structural basis for recognition of human 7SK long noncoding RNA by the La-related protein Larp7 [Biophysics and

Computational Biology]. *Pnas*, 1–10 (2018).

15. C. M. Olson, *et al.*, Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. *Nat. Chem. Biol.* **14**, 163–170 (2018).

16. D. D. Richman, *et al.*, The challenge of finding a cure for HIV infection. *Science (80-.).* **323**, 1304–1307 (2009).

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

18. I. Lebars, *et al.*, HEXIM1 targets a repeated GAUC motif in the riboregulator of transcription 7SK and promotes base pair rearrangements. *Nucleic Acids Res.* **38**, 7749–7763 (2010).

19. J. E. Brogie, D. H. Price, Reconstitution of a functional 7SK snRNP. *Nucleic Acids Res.*, 1–17 (2017).

20. S. Egloff, E. Van Herreweghe, T. Kiss, Regulation of Polymerase II Transcription by 7SK snRNA: Two Distinct RNA Elements Direct P-TEFb and HEXIM1 Binding. *Mol. Cell. Biol.* **26**, 630–642 (2006).

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

22. M. Tyagi, R. J. Pearson, J. Karn, Establishment of HIV Latency in Primary CD4 + Cells Is due to Epigenetic Transcriptional Silencing and P-TEFb Restriction . *J. Virol.* **84**, 6425–6437 (2010).

23. Z. Lin, G. C. Fillmore, T. H. Um, K. S. J. Elenitoba-Johnson, M. S. Lim, Comparative microarray analysis of gene expression during activation of human peripheral blood T cells and leukemic Jurkat T cells. *Lab. Investig.* **83**, 765–776 (2003).

24. A. G. Bodnar, *et al.*, Extension of Life-Span by Introduction of Telomerase into Normal Human Cells. *Science (80-.).* **279**, 349–352 (1998).

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

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

27. S. H. Chao, *et al.*, Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. *J. Biol. Chem.* **275**, 28345–28348 (2000).

28. S. Biglione, *et al.*, 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**, 1–12 (2007).

29. L. M. Shaw, *et al.*, DirtyGenes: testing for significant changes in gene or bacterial population compositions from a small number of samples. *Sci. Rep.* **9**, 1–10 (2019).

30. N. Kwiatkowski, *et al.*, Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).

31. P. Liu, *et al.*, 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 (2014).

32. N. He, A. C. Pezda, Q. Zhou, Modulation of a P-TEFb Functional Equilibrium for the Global Control of Cell Growth and Differentiation. *Mol. Cell. Biol.* **26**, 7068–7076 (2006).

33. A. Chaidos, *et al.*, Potentantimyeloma activity of the novel bromodomain inhibitors I-BET151 and I-BET762. *Blood* **123**, 697–705 (2014).

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

35. K. Roder, G. Stirnemann, J. Wales, S. Pasquali, Structural transitions in the RNA 7SK 5 hairpin and their effect on HEXIM binding. *Nucleic Acids Res.* **48**, 1–17 (2019).

36. L. Luo, *et al.*, HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions. *J. Mol. Biol.*, 166885 (2021).

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

38. N. Sithole, C. A. Williams, T. E. M. Abbink, A. M. L. Lever, DDX5 potentiates HIV-1 transcription as a co-factor of Tat. *Retrovirology* **17**, 1–16 (2020).

39. E. Van Herreweghe, *et al.*, Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *EMBO J.* **26**, 3570–3580 (2007).

40. E. Calo, *et al.*, RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature* **518**, 249–253 (2015).

41. X. Ji, *et al.*, SR Proteins Collaborate with 7SK and Promoter-Associated Nascent RNA to Release Paused Polymerase. *Cell* **153**, 855–868 (2013).

42. A. Bugai, *et al.*, P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress. *Mol. Cell* **74**, 254-267.e10 (2019).

43. G. Castelo-Branco, *et al.*, The non-coding snRNA 7SK controls transcriptional

termination, poising, and bidirectionality in embryonic stem cells. *Genome Biol.* **14** (2013).

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

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

46. C. Studniarek, *et al.*, The 7SK/P-TEFb snRNP controls ultraviolet radiation-induced transcriptional reprogramming. *Cell Rep.* **35** (2021).

47. K. V. Prasanth, *et al.*, Nuclear organization and dynamics of 7SK RNA in regulating gene expression. *Mol. Biol. Cell* **21**, 4184–4196 (2010).

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

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

50. W. Yang, *et al.*, Discovery of a High Affinity, Orally Bioavailable Macrocyclic FXIa Inhibitor with Antithrombotic Activity in Preclinical Species. *J. Med. Chem.* **63**, 7226–7242 (2020).

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

52. Y. Cheng, *et al.*, LARP7 is a potential tumor suppressor gene in gastric cancer. *Lab. Investig.* **92**, 1013–1019 (2012).

53. J. L. Tan, *et al.*, Stress from Nucleotide Depletion Activates the Transcriptional Regulator HEXIM1 to Suppress Melanoma. *Mol. Cell* **62**, 34–46 (2016).

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

55. A. Sengupta, G. M. Rice, K. M. Weeks, 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**, 1–19 (2019).

56. M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, 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 (2015).

57. J. S. Reuter, D. H. Mathews, RNAstructure: Web servers for RNA secondary structure prediction and analysis. *BMC Bioinformatics* (2010) https:/doi.org/10.1093/nar/gkt290.

58. Michael W. Pfaffl, A new mathematical model for relative quantification in real-time RT–PCR. *Mon. Not. R. Astron. Soc.* **29**, 2002–2007 (2001).

APPENDIX: DANCE-MAP PROTOCOL

Introduction

Many RNAs can adopt more than one structural state, and thus form conformational ensembles, but an assumption of most structure-prediction methods is that the RNA is in a single homogeneous state. This single-state assumption can result in a poorly defined averaged ensemble that is not representative of any actual, individual structural state. DANCE-MaP is a single-molecule correlated chemical probing experiment that can be performed in cells or on extracted total RNA. The current version of the DANCE experiment uses the classic reagent dimethyl sulfate (DMS). The DANCE strategy deconvolutes the reactivities of each conformational state to reveal the populations and individual reactivity profiles of each ensemble state. Correlated reactivities are then used to directly determine through-space base paring (PAIR) and tertiary interactions (RING). For a single RNA, the DANCE-MaP protocol requires 2 to 3 hours of hands-on time for all steps of chemical probing of the RNA and library preparation performed over 2 days. After sequencing and computational analysis results can be obtained in around a week.

In-cell DMS modification protocol

(1) Grow 10 mL of Jurkat (E6-1) cells in a T25 flask in complete RPMI [RMPI 1640 (Gibco) supplemented with 10% FBS (Millipore), 100 U/mL Pen/Strep (LifeTech)] at 37 °C and 5% CO_2 to a concentration of 1-3 $\times 10^6$ cells per mL.

(2) Centrifuge the 10 mL culture for 3 min at 150-250 RCF, remove spent media, wash once with PBS, and centrifuge again. Resuspend in 700 µL of fresh complete RPMI media and add 200 µL of 1 M bicine (pH 8 at 37 ˚C). Note: Bicine pH is important and should be pH 8.0 at 37 °C; if the 1 M bicine is prepared at room temperature the pH should be at 8.3.

(3) Mix 10 µL of 99.9% DMS with 50 µL ethanol to prepare a solution of 170 mM DMS. Steps 3-16 should be done in the chemical hood.

(4) Add 50 µL of the DMS solution (+ sample) or neat ethanol (– sample) to a 1.6-mL microcentrifuge tube.

(5) Add 450 µL of bicine-buffered cells to the tubes containing DMS solution or ethanol. Mix by pipetting up and down and incubate at 37 °C for 6 min.

(6) Mix 200 µL of 2-mercaptoethanol (BME) with 800 µL of water to make a 20% solution. Store BME solution on ice until use.

(7) Add 500 µL of the BME solution to the (+) and (–) samples, mix by pipetting, and place on ice for 3 min.

(8) Centrifuge samples for 3 min at 150-250 RCF.

(9) Discard the supernatant and resuspend the pellet in 1 mL Trizol reagent (Invitrogen). Mix and incubate at room temperature for 5 min.

(10) Add 200 µL of 100% chloroform and shake to mix, incubate at room temperature for 3 min.

(11) Centrifuge at 12,000 RCF for 15 min at 4 °C.

(12) Transfer the top aqueous layer to a clean 1.6-mL tube, taking care not to disrupt the white interphase layer. This should yield 400 to 600 µL of sample.

(13) Add 500 µL of 100% isopropanol to each sample and mix by shaking, incubate at room temperature for 10 min.

(14) Centrifuge at 12,000 RCF for 10 min at 4 °C.

(15) Remove the supernatant and add 500 µL of 75% ethanol to the pellet.

(16) Centrifuge at 7500 RCF for 5 min at 4 °C.

(17) Remove the supernatant and dry the pellet. Resuspend the pellet in 44 µL of nuclease-free water.

(18) Add 5 µL of 10× Turbo DNase buffer and 1 µL of Turbo DNase. Incubate for 30 min at 37 °C.

(19) (optional) Add 1 µL of Turbo DNase (Ambion) and incubate an additional 30 min. Note: This is a more stringent clean-up to guard against accidental DNA contamination from the RNA extraction.

(20) Add 90 µL of SPRI beads (MagBind TotalPure NGS, Omega BioTek, 1.8× bead:volume ratio) to each sample, mix, and incubate at room temperature for 5 min. Note: Rather than SPRI beads, an RNeasy column (Qiagen) or isopropanol precipitation can also be used to isolate the RNA.

(21) Place on a magnetic stand and allow 5 min for the beads to separate.

(22) Remove the liquid, and add 150 µL of 70% ethanol to each tube, taking care not to disturb the beads. Let stand for 30 s before removing the ethanol. Repeat the ethanol wash two additional times.

(23) Remove ethanol after the final wash, and dry the beads for 10 min at room temperature.

(24) Resuspend the beads in 20 µL of nuclease-free water and incubate at room temperature for 2 min.

(25) Place samples back on the magnetic stand and allow 1 min of separation before removing the supernatant containing the RNA.

(26) Quantify the amount of RNA (Nanodrop, Thermo). Note: There should be approximately 5 μg of RNA in each sample.

Cell-free DMA modification protocol

(1) Grow 10 mL of Jurkat (E6-1) cells in a T25 flask in complete RPMI media at 37 °C and 5% CO_2 to a concentration of 1-3 × 10^6 cells per mL.

(2) Centrifuge cells for 3 min at 150-250 RCF. Remove spent media, and resuspend in 1 mL of TRIzol reagent (Invitrogen). Follow steps 9-26 of in-cell treatment to isolate and quantify RNA.

(3) Dilute 5 μg of total RNA into a total volume of 50 μL, incubate at 98 °C for 1 min, and place at 4 °C for 1 min.

(4) Add 50 μL of 2× bicine folding buffer [400 mM bicine (pH 8.0), 400 mM potassium acetate (pH 8.0), and 10 mM $MgCl_2$] and incubate at 37 °C for 20 min.

(5) Mix 2 μL of 99.9% DMS with 10 μL ethanol to prepare a solution of 170 mM DMS. Steps 5-12 should be done in the chemical hood.

(6) Add 5 μL of the DMS solution (+ sample) or neat ethanol (– sample) to 1.6-mL microcentrifuge tubes.

(7) Pipet 45 μL of the RNA sample into the tubes containing DMS solution or ethanol. Mix by pipetting up and down, and incubate at 37 °C for 6 min.

(8) Mix 20 μL of BME with 80 μL of water to make a 20% solution. Store on ice.

(9) Add 50 μL of the BME solution to the (+) and (–) samples, mix by pipetting, and place on ice for 3 min.

(10) Add 4 μL of 5 M NaCl and 1 μL of glycogen to each tube. The glycogen is optional; it helps visualize the pellet.

(11) Add 100 μL 100% isopropanol, mix by shaking, and incubate at room temp for 10 min.

(12) Centrifuge at 10,000 RCF for 10 min at 4 °C, remove supernatant.

(13) Wash pellet with 200 μL of 75% ethanol, centrifuge at 10,000 RCF for 5 min at 4 °C, remove supernatant, and dry the pellet.

(14) Resuspend in 20 μL nuclease-free water and quantify RNA (Nanodrop, Thermo). Note: There should be approximately 2 μg of RNA in each sample.

MaP-RT protocol

(1) Design a 15 to 25 nucleotide-long RT primer that is the reverse complement of the 3′ end of the RNA of interest.

(2) Mix 2 µL of 10 mM dNTPs, 1 µL of 2 µM RT primer, and 9.5 µL nuclease-free water containing 1-2 µg total RNA.

(3) Incubate at 68 °C for 5 min and at 4 °C for 2 min.

(4) Prepare 10× NTP(–) buffer: Mix 50 µL 1 M Tris (pH 8), 37.5 µL 2 M KCl, 2.5 µL nuclease-free water, and 10 µL 1 M DTT (DTT should be added last). This buffer needs to be freshly prepared for each experiment. The 1 M DTT stock should be prepared and aliquoted into single-use tubes stored at -20 °C until use.

(5) Prepare the RT master mix solution: For each sample, mix 4 µL 5 M betaine, 2 µL 10X NTP(–) buffer, and immediately before adding master mix solution add 0.5 µL of 240 mM $MnCl_2$. $MnCl_2$ should be prepared fresh from a concentrated stock before each use as it can oxidize.

(6) Add 6.5 µL RT master mix solution to each sample and incubate at 25 °C for 2 min.

(7) Add 1 µL of Superscript II (Invitrogen) to each tube and place in a thermocycler. Incubate at 25 °C for 10 min followed by 42 °C for 90 min, then perform the following cycle 10 times: 50

°C for 2 min then 42 °C for 2 min. Heat at 70 °C for 10 min to inactivate the Superscript II reverse transcriptase, then cool to 4 °C. The samples can be kept at 4 °C overnight.

(8) Prepare G-50 microspin columns (Cytiva): Vortex the column, remove the bottom stopper and unscrew the lid slightly, and spin at 750 RCF for 1 min to remove storage buffer.

(9) Place the column on a 1.6-mL tube, and load 20 µL of sample onto the column. Spin 2 min at 750 RCF. This procedure should yield 26-28 µL of cDNA.

Gene-specific two-step PCR protocol

(1) Design step 1 PCR primers to RNA of interest, where the red nucleotides are the Illumina specific adaptor sequence, the blue are the random nucleotides, and gene-specific sequence is located at the 3′ end of each primer.

Forward primer:

GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNN–[sequence sense to the 5′ end of the RNA of interest]

Reverse primer:

CCCTACACGACGCTCTTCCGATCTNNNNN–[reverse complement of 3′ end of target]

(2) Add step 1 PCR components shown below to a PCR tube. Make a master mix of all components except cDNA template to reduce pipetting errors.

Component	Amount (µL)	Final Concentration
Q5 reaction buffer (5×)	10	1×
dNTPs (10 mM each)	1	0.2 mM each
Step 1 Primers (25 µM each)	1	0.5 µM each
Q5 hot start DNA polymerase	0.5	0.02 U/µL
cDNA template	5	
Nuclease-free water	32.5	
Final	50	

(3) Perform PCR with the following program:

Step	Denature	Anneal	Extend
1	98 ˚C, 30 s		
2 (10 to 20 cycles)	98 ˚C, 8 s	63-68 ˚C, 20 s	72 °C, 20 s
3			72 °C, 2 min

The number of cycles used is based on the prevalence of the RNA in cell. If the RNA is abundant use 10 cycles and if the RNA is lowly abundant 20 cycles may be needed. The annealing temperature should be chosen based on the melting temperature of the primers determined using the NEB calculator; 63-68 ˚C is ideal.

(4) Add 50 µL of the sample to 40 µL SPRI beads (MagBind TotalPure NGS, Omega BioTek, 0.8× bead:volume ratio). Bead concentration is based on length of the product and can be adjusted to select the expected product and to remove primer dimers.

(5) Incubate at room temperature for 5 min, and then place on a magnetic strip for 2-5 min. Remove the liquid and then wash with 100 µL of 80% ethanol two times with 30 s of incubation after each ethanol addition.

(6) After the final ethanol wash, remove the ethanol and dry the beads for 10 min at room temperature.

(7) Elute with 20 µL of nuclease-free water and quantify (High Sensitivity dsDNA Qubit Assay, Thermo). The procedure should yield 0.1 ng/µL to 2 ng/µL of PCR product for most RNA.

(8) Obtain step 2 universal primer and primers with TruSeq adaptors. A different index primer should be used for each sample.

Universal primer:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG:

Indexed Primer:

CAAGCAGAAGACGGCATACGAGAT - [barcode] – GTGACTGGAGTTCAGAC

(9) Place step 2 PCR components shown below into a PCR tube. Make a master mix of all components except the PCR product and the Step 2 primers to reduce pipetting errors.

Component	Amount (µL)	Final Concentration
Q5 reaction buffer (5×)	10	1×
dNTPs (10 mM each)	1	0.2 mM each
Step 2 Primers (10 µM each)	5	0.5 µM each
Q5 hot start DNA polymerase	0.5	0.02 U/µL
PCR product from step 1 (0.1 ng/µL)	10	
Nuclease-free water	32.5	
Final	50	

(10) Perform PCR with the following program:

Note: combination of cycles for step 1 and 2 should not exceed 30 cycles.

Step	Denature	Anneal	Extend
1	98 ˚C, 30 s		
2 (10-20 cycles)	98 ˚C, 8 s	65 ˚C, 20 s	72 °C, 20 s
22			72 °C, 2 min

(11) Purify the PCR product using SPRI beads (40 µL, MagBind TotalPure NGS, Omega BioTek, 0.8× bead:volume ratio). Bead concentration is based on length of the product and can be adjusted to select the expected product and to remove primer dimers.

(12) Elute with 20 µL of nuclease-free water and quantify (High Sensitivity dsDNA Qubit Assay). This procedure should yield 10 ng/µL to 100 ng/µL of DNA for most RNA.

(13) Use a Bioanalyzer to evaluate the length of amplicons. The sequencing adaptors will add ~140 nucleotides to the target RNA region length.

(14) Dilute samples to 2 or 4 nM and sequence on an Illumina MiSeq.

DANCE-MAP data analysis

(1) Obtain fastq.gz files from the sequencing run. Place into folders for DMS-treated (+) and ethanol-treated (–) samples. Prepare a reference fasta file with primer-binding regions in lowercase. Subsequent steps will consume computer time and memory.

Example folder construction to run ShapeMapper

|run-folder/

|----{target-RNA}.fa

|----plus/

|--------{DMS-treated-sample}-R1.fastq.gz

|--------{DMS-treated-sample}-R2.fastq.gz

|----minus

|--------{untreated-sample}-R1.fastq.gz

|--------{untreated-sample}-R2.fastq.gz

Example: {target-RNA}.fa

>target RNA

ggatgtgagggcgatctgGC--[rest of sequence]—CGctgcatgtggcagtctgcctttctttt

(2) Download ShapeMapper2, DanceMapper, RingMapper, and RNAtools from Weeks lab github (https://github.com/weeks-unc). Download RNAstructure from the Mathews lab website (https://rna.urmc.rochester.edu/RNAstructure.html). Install the programs and add them to local path. For installation instructions refer to the readme files after download.

(3) Run the following command in ShapeMapper (Note: run with at least 5 GB of memory and allow for a 6 hour run, should be done in 30 min to 2 h depending on sequence depth achieved):

```
shapemapper \
--output-classified \
--target {target-rna}.fa \
--name "sample" \
--overwrite \
--amplicon \
--modified --folder plus \
--untreated --folder minus \
'
```

(4) When ShapeMapper is finished running, find the folder shapemapper_out, and run the following command (Note: run with at least 50 GB of memory and run for 7 days, should be completed in 6 hours to 4 days depending on sequence depth achieved and how many clusters are found):

```
python DanceMap.py \
--fit \
--pairmap \
--ring \
--profile sample_XX_profile.txt \
--outputprefix sample \
--modified_parsed sample_Modified_XX_parsed.mut \
--untreated_parsed sample_Untreated_XX_parsed.mut \
```

Red: Optional, will result in automatic creation of correlation data for each individual cluster.

(5) To fold and plot secondary structure, use arcPlots for each state. Download arcPlot from the Weeks lab github (https://github.com/weeks-unc), install and add to your path, then run the following command (Note: this can be done locally and should only take a few seconds to minutes):

```
python foldClusters.py \
sample-reactivity.txt \
sample_out \
Optional inputs:
--bp sample \        # only use if pairmap was run
--prob \             # output probability-based structure instead of minimum free energy
--pk \               # use ShapeKnots instead of Fold to determine structures
```

(6) For quick comparison of per nucleotide reactivity between samples and replicates, run the following command:

```
python PlotClusters.py \
--bm1 sample.bm \
--bm2 sample2.bm \
--out out \
--align
```

Conclusion

DANCE-MaP can be used to separate multiple structural states of an RNA transcript. DANCE allows rapid determination of whether a functional RNA adopts more than one conformation and yields PAIR and RING correlations enabling very robust RNA structure modeling.