Visualizing RNA structure ensembles by single-molecule correlated chemical probing

J. Winston Arney[1], Alain Laederach[2], and Kevin M. Weeks[1,*]
[1]Department of Chemistry and [2]Department of Biology, University of North Carolina, Chapel Hill, NC 27599-3290

* correspondence, weeks@unc.edu

Abstract

RNA molecules fold to form complex internal structures. Many of these RNA structures populate ensembles with rheostat-like properties, with each state having a distinct function. Until recently, analysis of RNA structures, especially within cells, was limited to modeling either a single averaged structure or computationally-modeled ensembles. These approaches obscure the intrinsic heterogeneity of many structured RNAs. Single-molecule correlated chemical probing (smCCP) strategies are now making it possible to measure and deconvolute RNA structure ensembles based on efficiently executed chemical probing experiments. Here, we provide an overview of fundamental single-molecule probing principles, review current ensemble deconvolution strategies, and discuss recent applications to diverse biological systems. smCCP is enabling a revolution in understanding how the plasticity of RNA structure is exploited in biological systems to respond to stimuli and alter gene function. The energetics of RNA ensembles are often subtle and a subset can likely be targeted to modulate disease-associated biological processes.

RNA molecules function through heterogenous structure ensembles

RNA molecules ubiquitously fold into complex higher-order structures with diverse biological functions. Many structured RNAs populate heterogeneous ensembles, and these structures can interconvert to regulate biological processes. In essence, RNA molecules can be thought of as functional rheostats in which distinct states in an RNA conformation enable or favor a given process or interaction to occur, or not. We note, and this review will show, that few RNAs function as true binary "switches". We prefer the term "rheostat" over switch because native ensembles rarely populate 100% of a given state. RNA structural dynamics have been implicated in nearly every facet of RNA biology [1–3]. For protein-coding RNAs, internal structures regulate splicing, translation initiation, ribosome frameshifting, and mRNA turnover. Among noncoding RNAs, dynamic structures regulate binding by protein, RNA, and DNA partners, yielding downstream

effects on ribonucleoprotein function and transcriptional regulation. It is likely that RNA structure ensembles perform numerous additional, yet-to-be-discovered, functions.

Our knowledge of RNA ensembles in their native cellular contexts has been limited by the difficulty of physically and directly characterizing their underlying states. Chemical probing provides one of the most direct and reliable high-throughput strategies for measuring nucleotide-resolution features of RNA structures in cells. Conventional chemical probing experiments interpret chemical probing data as a population-averaged RNA structure and thus obscure the underlying heterogeneity of structure ensembles. Single-molecule correlated chemical probing (smCCP) strategies are now addressing this limitation by enabling detection of multiple per-nucleotide events in the same RNA strand, offering new views of RNA ensembles in their native cellular contexts.

The physical mechanism of smCCP

Chemical probing is an established, straightforward method for probing RNA structure, both of purified RNA and of RNA in cells. In a chemical probing experiment, a biological sample is treated with a chemical probe that reacts to form covalent adducts with (conformationally dynamic, often single-stranded) RNA nucleotides (**Figure 1a**). After probing, multiple covalent adducts in a given RNA strand are encoded as mutations during relaxed fidelity reverse transcription, a process called mutational profiling (or MaP) [4]. These mutations in the resulting cDNA correspond to individual chemical reactivity events and are quantified by massively parallel sequencing [5,6]. In principle, each MaP sequencing read reports on the structure of a single RNA molecule. Conventional chemical probing-MaP experiments combine all sequencing reads of a particular RNA to generate a per-nucleotide reactivity profile that reflects the population average of RNA structures in a sample (**Figure 1b**). For RNAs that form multiple structures, a single reactivity profile is insufficient for accurate structure modeling. Critically, MaP captures multiple chemical adducts on a single molecule of RNA. The multiple chemical reactivity events in a single RNA will show distinct pattens of correlated reactivities as a function of both (i) the different states present in an ensemble and (ii) specific inter-nucleotide interactions including base pairing and tertiary interactions (**Figure 1c**).

The physical experiment required to perform a smCCP analysis is straightforward [4] and not much more difficult than a conventional RNA structure probing experiment. The key innovation are the use of MaP and the computational deconvolution of smCCP data in terms of the

underlying ensemble states.

Deconvolution of RNA structural ensembles by smCCP

To model RNA ensembles from experimental smCCP datasets, MaP reactivities must first be deconvoluted into multiple groups or clusters, each corresponding to an individual RNA state. Current chemical probing deconvolution strategies use two distinct approaches: strategies that cluster MaP reads directly prior to structure modeling, and those that model structures based on single MaP reads and then cluster by structure similarity. In the first approach, sequencing reads from a chemical probing-MaP experiment are clustered by correlated mutations (**Figure 2a**). Clustered reads are then used to compute individual reactivity profiles, which are used to model an ensemble of up to about five RNA structure states. In this cluster-directly category, DREEM (Detection of RNA folding ensembles using expectation-maximization) leverages sequencing reads from dimethyl sulfate mutational profiling (DMS-MaP) experiments, focusing on chemical adducts at A and C residues, which are deconvoluted using an expectation-maximization algorithm [7]. DRACO (Deconvolution of RNA alternative conformations) similarly uses two-nucleotide DMS-MaP data, but employs a spectral clustering algorithm for deconvolution and includes an integrated sliding window function for deconvoluting long RNAs [8]. DANCE-MaP (Deconvolution and annotation of ribonucleic conformational ensembles by mutational profiling) deconvolutes DMS-MaP data using an expectation-maximization clustering approach similar to DREEM [9]. DANCE-MaP implements additional features, including the abilities to measure chemical modifications at G and U residues and to directly measure through-space internucleotide interactions. After clustering, DANCE-MaP identifies pairwise internucleotide reactivity correlations within each ensemble state that directly measure specific base-pairing and tertiary interactions within RNA structures [9,10]. These MaP deconvolution methods have unique features and limitations [11], but all share a common framework and leverage the foundational DMS-MaP concept [4].

DaVinci (Determination of the variation of the RNA structure conformation through stochastic context-free grammar) [12] is distinct from the former strategies in that it first generates a structure model for each individual MaP read and then clusters these models (**Figure 2b**). Mutations within each read are used as forced single-stranded constraints in a stochastic context-free grammar algorithm that seeks to model the best-fit structures of individual RNA molecules, generating ~10^3 models. These structure models are subjected to dimensionality reduction and clustering, yielding a coarse-grained ensemble of up to four states, roughly analogous to direct

MaP deconvolution strategies.

Challenges in assessing the accuracy of ensemble deconvolution

At present, it is difficult to quantify which of the current deconvolution approaches yields RNA structural ensembles that best reflect biological reality, particularly across multiple biological contexts. Few ground-truth ensembles have been characterized, and none for an in-cell system. One informative test-tube model RNA is the adenine riboswitch encoded by the *V. vulnificus add* gene. This RNA folds into multiple conformations that re-equilibrate in response to adenine binding (**Figure 3a**) [13]. The widely expressed 7SK RNA is emerging as a powerful test system for comparing strategies for integrated in-cell and cell-free ensemble visualization [9,14–16]. The highly structured SARS-CoV-2 frameshifting element (FSE) likewise adopts multiple conformations [8,17–21].

Current deconvolution algorithms have yielded different ensembles when applied to the same RNA. DREEM, DRACO, and DANCE-MaP (**Figure 2a**) have each been used to deconvolute the adenine riboswitch ensemble [7–9]. Each method detected both principal riboswitch structures, but the relative populations of each state varied depending on the method used. Even when analyzing a common dataset, based on the SARS-CoV-2 FSE, deconvoluted ensembles showed significant differences depending on the deconvolution algorithm used [21]. Conformations proposed for the FSE structure ensembles additionally vary depending on flanking sequences and whether the RNA is probed in cells or *in vitro* [8,19,20]. The structure-first strategy used by DaVinci [12] (**Figure 2b**) has not yet been subjected to a comparative analysis with other methods.

Some of these differences between methods likely arise from variations in and quality of probing conditions, to which RNA ensemble measurements are often sensitive. Ensemble detection can also be affected by sequencing artifacts including PCR duplication and low read depths. Each analysis pipeline makes different assumptions during clustering, particularly in determining the number of states, which can be either user-defined or empirically determined by significance testing. Ensemble features determined by different approaches will likely become more similar as the field converges on best practices. Despite these challenges, each of the ensemble deconvolution strategies have detected alternative RNA structures in important biological systems. Many of these RNA ensembles have been shown to be functionally significant, yielding intriguing, suggestive insights about the importance of structure ensembles for RNA function.

RNA structure ensembles in viral genomes

Viral RNA genomes make extensive use of RNA structure to regulate their replication and interactions with the host cell. Several studies have used smCCP to analyze the structural heterogeneity of viral RNAs. smCCP, applied to the HIV-1 viral RNA genome expressed in infected $CD4^+$ T cells and analyzed using DREEM, confirmed expectations that structural heterogeneity is pervasive across the RNA genome [7]. Several local ensembles overlapped notable regulatory motifs. The Rev response element (RRE) has long been known to fold into multiple structure states. smCCP experiments demonstrated that the relative populations of RRE states are maintained in cells, inside purified virions, and for RNA transcribed *in vitro*, suggesting that the heterogeneity of the HIV-1 RRE is an intrinsic property of the RNA rather than a consequence of its environment. smCCP also identified a structure rheostat that either masks (by base pairing) or exposes the A3 splice acceptor site, which regulates expression of the transcription regulatory protein Tat (**Figure 3b**). Point mutations that stabilized the splice site-masking structure by adding 4-5 base pairs reduced splicing efficiency [7]. This work highlighted the pervasiveness of RNA structure heterogeneity and emphasized that perturbing alternative states has significant biological consequences.

Multiple studies applied smCCP to investigate the structural heterogeneity of the SARS-CoV-2 RNA genome. In the first study, DMS-MaP probing of the SARS-CoV-2 genome, extracted from cells, followed by windowed deconvolution using DRACO, revealed that approximately 15% of the genome resolved into local structural ensembles [8]. Alternative structures were enriched at open reading frame and protein boundaries, suggesting that RNA structures in these regions potentially fulfill functional rheostat-like roles. Two conformations were detected at the FSE, and an additional two-state ensemble was detected in the 3′-untranslated region. Both 3′-UTR structures are supported by phylogenetic analysis across coronavirus genomes, suggesting a functional role for these conformations.

In the second study, DMS-MaP probing of the SARS-CoV-2 genome in human Huh7 and Vero cells followed by windowed deconvolution with DREEM, suggested that half of the SARS-CoV-2 genome adopts multiple local RNA structures in at least one of these cell types [19]. The FSE was again observed to form a two-state ensemble, although the modeled structures differ from the DRACO-predicted ensemble. One of the DREEM-deconvoluted FSE structures contained long-range base pairs to other genomic regions. Extending the FSE sequence to include these

long-distance interactions increased frameshifting two-fold in a luciferase reporter system, suggesting alternative RNA structures that occur over long regions play regulatory roles in translation [19].

In a third study, a set of *in vitro* transcribed SARS-CoV-2 FSE structures with varied flanking sequences were probed by DMS-MaP, and the data were deconvoluted with DREEM [20]. The results of this study emphasize that the FSE structure ensemble is affected by surrounding sequence context (**Figure 3c**). Changes in the sequences flanking the FSE had significant effects on the ensemble of FSE structures and associated frameshift efficiencies. Two-nucleotide point mutations that disrupted the canonical FSE pseudoknot reduced frameshifting efficiency from 26% to 1%, demonstrating that a pseudoknot conformation is essential for frameshifting. Disruption of the FSE pseudoknot with a targeted antisense oligonucleotide caused a 20% reduction in frameshifting efficiency in cells and dose-dependent elimination of frameshifting *in vitro*, demonstrating that perturbations that alter RNA structure can induce significant biological effects.

In sum, results from smCCP studies of two long viral RNA genomes, HIV-1 and SARS-CoV-2, emphasize, first, that RNA genomes form many alternative structures that persist in purified RNA, in virions, and in cells, strongly suggesting that heterogeneity is a widespread, intrinsic property of viral RNAs. Second, these conformational ensembles play notable, perturbable (via mutation or antisense oligonucleotide binding) roles in regulating splicing (in HIV) and ribosomal frameshifting (in SARS-CoV-2), processes that are critical for viral protein translation. Further smCCP studies with viral RNAs have the potential to uncover broad roles for RNA structure ensembles.

Dynamic structure rheostats in eukaryotic noncoding RNAs

The functions of long non-coding RNAs (lncRNAs) depend on complex sets of interactions involving protein, DNA, and other RNA partners that are modulated by the conformational state of the lncRNA. The 7SK lncRNA is the core of a ribonucleoprotein complex that is a key regulator of cellular transcriptional state. DMS probing of 7SK in human Jurkat cells, followed by DANCE-MaP deconvolution, revealed that 7SK populates a multi-state structure ensemble in cells [9]. This structure ensemble regulates transcription by rearranging to bind or release the transcription factor P-TEFb (**Figure 3d**). There is a strong relationship between the 7SK in-cell ensemble and overall transcription: 7SK shifts toward the transcriptionally inactive state during cell contact-induced quiescence and toward the transcriptionally active states to compensate for transcription

inhibition by flavopiridol treatment. An antisense oligonucleotide that binds and stabilizes the 7SK active state causes a two-fold increase in transcription of a 7SK-sensitive gene, demonstrating that the 7SK structural ensemble can be targeted to modulate transcription.

The human telomerase RNA (hTR) is an essential component of the telomerase complex, which appends repetitive telomere sequences to the ends of chromosomal DNA. DMS-MaP probing of hTR and DREEM deconvolution of the data revealed that a fraction of hTR transcripts fold into a conformation that differs from the canonical structure [22] (**Figure 3e**). Point mutations that stabilize this alternative conformation reduce assembly and activity of the telomerase complex. The observed alternative structure was interpreted as a nonfunctional, misfolded conformation, emphasizing that some alternative RNA structures disrupt biological function.

COOLAIR is a plant-specific lncRNA that epigenetically represses transcription of its antisense transcript, *FLC*, in response to cold exposure [23]. The *FLC* protein is a floral repressor, so cold-induced downregulation of *FLC* by *COOLAIR* induces flowering. *COOLAIR* was probed in *Arabidopsis* seedlings, and the data were deconvoluted using DaVinci, revealing three structural states whose populations changed in response to cold exposure [12]. Structural heterogeneity was observed in a region of *COOLAIR* complementary to the *FLC* transcription start site. Mutating four base-paired nucleotides within this region of the *COOLAIR* lncRNA resulted in enhanced association between *COOLAIR* and *FLC* chromatin, a two-fold reduction in *FLC* transcription, and an early-flowering phenotype (**Figure 3f**). This study suggests that RNA conformational states can respond to stimuli to regulate interactions with chromatin, inducing significant phenotypic changes.

In sum, smCCP studies of three distinct lncRNAs revealed broad principles about RNA ensembles in eukaryotes (**Figure 4**). For both 7SK and *COOLAIR*, observed structure ensembles were responsive to environmental stimuli including small-molecule perturbation, targeted antisense oligonucleotide treatment, cellular quiescence, and cold exposure. For all three lncRNAs, changes in the relative proportion of states in the RNA structure ensemble – induced by external stimuli or point mutagenesis – had significant functional consequences. Many RNA structure ensembles thus likely function as molecular rheostats that regulate cellular responses to stimuli.

Summary and outlook

Current smCCP studies reveal or reinforce four key insights about RNA structural ensembles and conformational heterogeneity.

First, RNAs have a broad capacity to fold into multiple conformations, both isolated in test tubes and within cells (**Figure 3**). This intrinsic heterogeneity suggests that alternative RNA structures are likely widespread in the transcriptome. Most ensembles examined to date consistently populate both (or all) of their underlying states, under all conditions measured, and thus perform roughly as functional rheostats.

Second, RNA structure ensembles are responsive to both environmental stimuli and targeted perturbations (**Figure 4a**). Energetic perturbations to RNA structures, such as point mutations or small molecule binding (equivalent to a few kcal/mol) can induce significant shifts in the populations of an RNA ensemble. This responsiveness enables RNAs to function as sensitive rheostat-like regulators of cellular functions.

Third, perturbations of RNA ensembles modulate diverse biological processes, including global or gene-specific transcription, alternative splicing, protein translation, and ribonucleoprotein assembly (**Figure 4b**). Together, these observations suggest that RNA structural ensembles can be intentionally targeted to affect a wide variety of RNA-centric processes.

Most broadly, given that dynamic RNA structures have now been observed in widely varied biological systems – from bacterial riboswitches, to viral RNA genomes, to eukaryotic lncRNAs – it is likely that we have obtained only a small picture of functional RNA ensembles. smCCP is now poised to fundamentally alter and expand our understanding of RNA structure by revealing numerous new functions for RNA structural ensembles. Ultimately, a subset of these RNA ensembles will likely prove targetable (with antisense oligonucleotides and small molecules) to manipulate biological processes and to create new therapeutic opportunities.

Acknowledgements

Work in our laboratories is supported by the US National Institutes of Health, the US National Science Foundation, and industry partners.

Disclosure

K.M.W. is an advisor to and holds equity in Ribometrix and A-Form Solutions. Other authors

declare no conflicts.

References

1. Mustoe AM, Corley M, Laederach A, Weeks KM: **Messenger RNA Structure Regulates Translation Initiation: A Mechanism Exploited from Bacteria to Humans**. *Biochemistry* 2018, **57**:3537–3539.
2. Ganser LR, Kelly ML, Herschlag D, Al-Hashimi HM: **The roles of structural dynamics in the cellular functions of RNAs**. *Nat Rev Mol Cell Biol* 2019, **20**:474–489.
3. Bose R, Saleem I, Mustoe AM: **Causes, functions, and therapeutic possibilities of RNA secondary structure ensembles and alternative states**. *Cell Chem Biol* 2024, **31**:17–35.
   • A thorough and accessible overview of the biological functions and mechanisms of RNA structure ensembles and current methods for ensemble detection by smCCP. The relevance of RNA structure rheostats to disease is also emphasized, providing a provocative perspective on the therapeutic potential of targeting RNA structural ensembles.
4. Homan PJ, Favorov O V., Lavender CA, Kursun O, Ge X, Busan S, Dokholyan N V., Weeks KM: **Single-molecule correlated chemical probing of RNA**. *Proc Natl Acad Sci U S A* 2014, **111**:13858–13863.
5. Weeks KM: **SHAPE Directed Discovery of New Functions in Large RNAs**. *Acc Chem Res* 2021, **54**:2502–2517.
6. Mustoe AM, Weidmann CA, Weeks KM: **Single-Molecule Correlated Chemical Probing: A Revolution in RNA Structure Analysis**. *Acc Chem Res* 2023, **56**:763–775.
7. Tomezsko PJ, Corbin VDA, Gupta P, Swaminathan H, Glasgow M, Persad S, Edwards MD, Mcintosh L, Papenfuss AT, Emery A, et al.: **Determination of RNA structural diversity and its role in HIV-1 RNA splicing**. *Nature* 2020, **582**:438–442.
   •• DREEM is a widely-used method for deconvoluting RNA structure ensembles from smCCP data. When applied to the HIV-1 RNA in cells, DREEM revealed abundant heterogeneous RNA structures. One RNA structure rheostat was found to regulate splicing of the Tat transcript.
8. Morandi E, Manfredonia I, Simon LM, Anselmi F, van Hemert MJ, Oliviero S, Incarnato D: **Genome-scale deconvolution of RNA structure ensembles**. *Nat Methods* 2021, **18**:249–252.
   • DRACO is an attractive strategy for deconvoluting structures in long RNAs, which are

automatically handled by a sliding window function. This feature may prove to be a significant advantage for investigating large viral genomes and lncRNAs where functional structure rheostats are potentially abundant.

9. Olson SW, Turner AMW, Arney JW, Saleem I, Weidmann CA, Margolis DM, Weeks KM, Mustoe AM: **Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP**. *Mol Cell* 2022, **82**:1708-1723.e10.
   •• DANCE-MaP introduces the assignment of individual reads to each ensemble, allowing direct detection of base pairing and other through-space interactions. DANCE-MaP resolved a structure rheostat in the human 7SK RNA that regulates cellular transcription. The 7SK RNA ensemble responds to multiple stimuli, including antisense oligonucleotide binding, demonstrating that RNA structure ensembles can be targeted to affect biological processes.
10. Mustoe AM, Lama NN, Irving PS, Olson SW, Weeks KM: **RNA base-pairing complexity in living cells visualized by correlated chemical probing**. *Proc Natl Acad Sci U S A* 2019, **116**:24574–24582.
11. Aviran S, Incarnato D: **Computational Approaches for RNA Structure Ensemble Deconvolution from Structure Probing Data: Deconvolution of RNA structure ensembles**. *J Mol Biol* 2022, **434**:167635.
12. Yang M, Zhu P, Cheema J, Bloomer R, Mikulski P, Liu Q, Zhang Y, Dean C, Ding Y: **In vivo single-molecule analysis reveals COOLAIR RNA structural diversity**. *Nature* 2022, **609**:394–399.
   •• DaVinci introduces a distinct method for deconvoluting RNA structure ensembles. When applied to structure probing data obtained on the *COOLAIR* lncRNA in plants, DaVinci revealed a temperature-dependent structure rheostat that represses *FLC* transcription to induce flowering. This work introduces probing of an RNA structure ensemble in a multicellular organism.
13. Reining A, Nozinovic S, Schlepckow K, Buhr F, Fürtig B, Schwalbe H: **Three-state mechanism couples ligand and temperature sensing in riboswitches**. *Nature* 2013, **499**:355–359.
14. Luo L, Chiu LY, Sugarman A, Gupta P, Rouskin S, Tolbert BS: **HnRNP A1/A2 Proteins Assemble onto 7SK snRNA via Context Dependent Interactions**. *J Mol Biol* 2021, **433**:166885.
15. Yang Y, Liu S, Egloff S, Eichhorn CD, Hadjian T, Zhen J, Kiss T, Zhou ZH, Feigon J: **Structural basis of RNA conformational switching in the transcriptional regulator**

7SK RNP. *Mol Cell* 2022, **82**:1724-1736.e7.

16. Camara MB, Lange B, Yesselman JD, Eichhorn CD: **Visualizing a two-state conformational ensemble in stem–loop 3 of the transcriptional regulator 7SK RNA**. *Nucleic Acids Res* 2024, **52**:940–952.
17. Krueger BJ, Varzavand K, Cooper JJ, Price DH: **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* 2010, **5**:e12335.
18. Schlick T, Zhu Q, Dey A, Jain S, Yan S, Laederach A: **To Knot or Not to Knot: Multiple Conformations of the SARS-CoV-2 Frameshifting RNA Element**. *J Am Chem Soc* 2021, **143**:11404–11422.
19. Lan TCT, Allan MF, Malsick LE, Woo JZ, Zhu C, Zhang F, Khandwala S, Nyeo SSY, Sun Y, Guo JU, et al.: **Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells**. *Nat Commun* 2022, **13**:1–14.
20. Pekarek L, Zimmer MM, Gribling-Burrer AS, Buck S, Smyth R, Caliskan N: **Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function**. *Nucleic Acids Res* 2023, **51**:728–743.
    • Three orthogonal methods are used to interrogate the SARS-CoV-2 FSE ensemble: smCCP, optical tweezers, and frameshift reporter assays. The FSE structure exhibits notable heterogeneity depending on flanking sequences. Targeting the FSE structure by point mutations and antisense oligonucleotide binding significantly reduces frameshifting.
21. Dey A, Yan S, Schlick T, Laederach A: **Abolished frameshifting for predicted structure-stabilizing SARS-CoV-2 mutants: Implications to alternative conformations and their statistical structural analyses**. *bioRxiv* 2024, doi:10.1101/2024.03.28.586935.
22. Forino NM, Woo JZ, Zaug AJ, Jimenez AG, Cech TR, Rouskin S, Stone MD: **Dissecting telomerase RNA structural heterogeneity in living human cells with DMS-MaPseq**. *bioRxiv* 2023, doi:10.1101/2023.10.04.560962.
23. Swiezewski S, Liu F, Magusin A, Dean C: **Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target**. *Nature* 2009, **462**:799–802.

Figure Legends

Figure 1

Illustration of the smCCP experiment. (**a**) Chemical probing induces multiple covalent adducts in individual RNA strands at conformationally flexible nucleotides. These multiple adducts are recorded as mutations in the cDNAs generated during the reverse transcription step of MaP. (**b**) Conventional per-nucleotide analysis of chemical probing data yields a population average of structures for a given RNA. (**c**) In smCCP, correlated mutation events can be analyzed to resolve multiple RNA structure states and inter-nucleotide interactions including base pairing and tertiary contacts.

Figure 2

Strategies for detection of RNA structural ensembles by smCCP. (**a**) In the strategy used by DREEM, DRACO, and DANCE, MaP reads are clustered by correlated mutations, and these clustered MaP profiles are then used to model each state of the ensemble. (**b**) In the strategy used by DaVinci, a structure model is generated for each individual MaP read, and these are then clustered by structure similarity.

Figure 3

Biological functions of RNA structural ensembles. (**a**) The *V. vulnificus* adenine riboswitch undergoes a conformational change upon adenine binding, which regulates translation of the *add* transcript by masking or exposing the Shine-Dalgarno (SD) sequence [13]. (**b**) An RNA structure rheostat masks the HIV-1 A3 splice site, reducing splice site usage and Tat transcript abundance [7]. (**c**) Alternative structures at the SARS-CoV-2 frameshift element modulate frameshift efficiency [19,20]. The ensemble can be altered by binding of an antisense oligonucleotide (ASO). (**d**) The human 7SK lncRNA is a transcription regulatory rheostat that responds to cell contact-induced quiescence, transcriptional stress by flavopiridol treatment, and a structure-targeted antisense oligonucleotide [9,15]. (**e**) Human telomerase RNA can misfold into an alternative structure state, resulting in reduced assembly of the telomerase ribonucleoprotein (RNP) complex [22]. (**f**) *Arabidopsis COOLAIR* lncRNA undergoes a change in preferred structure after cold exposure that represses *FLC* transcription and induces flowering [12].

Figure 4

Perturbations to RNA ensembles affect diverse biological processes. (**a**) RNA ensembles

respond to numerous influences, ranging from environmental stimuli to targeted perturbations. (**b**) Examples of functions influenced by RNA structural dynamics in cells.

Figure 1

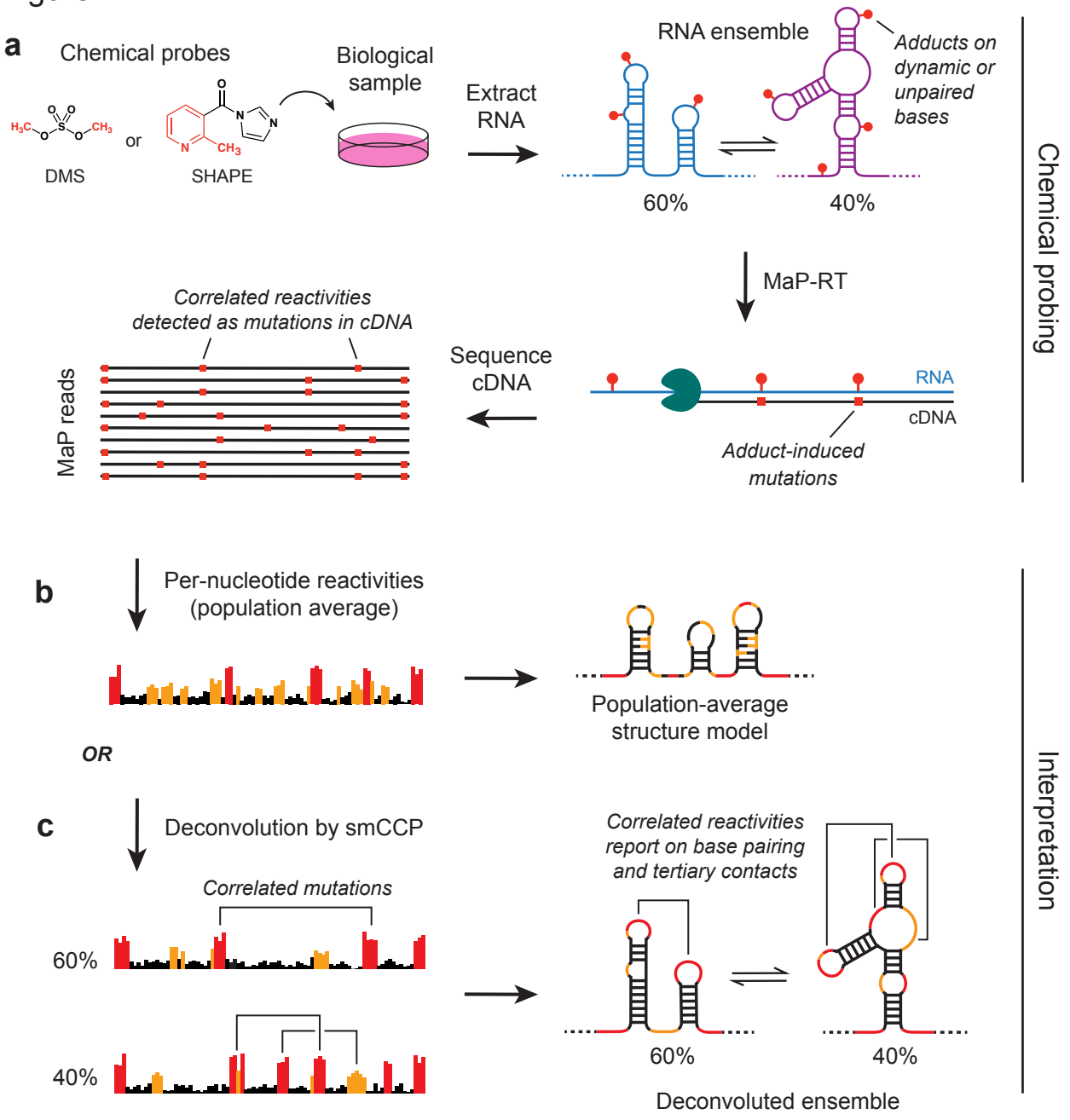

a
Chemical probes
Biological sample
DMS
or
SHAPE
Extract RNA
RNA ensemble
Adducts on dynamic or unpaired bases
60%
40%
MaP-RT
RNA
cDNA
Adduct-induced mutations
Sequence cDNA
Correlated reactivities detected as mutations in cDNA
MaP reads
Chemical probing
b
Per-nucleotide reactivities (population average)
Population-average structure model
OR
c
Deconvolution by smCCP
Correlated mutations
60%
40%
Correlated reactivities report on base pairing and tertiary contacts
60%
40%
Deconvoluted ensemble
Interpretation

Figure 2

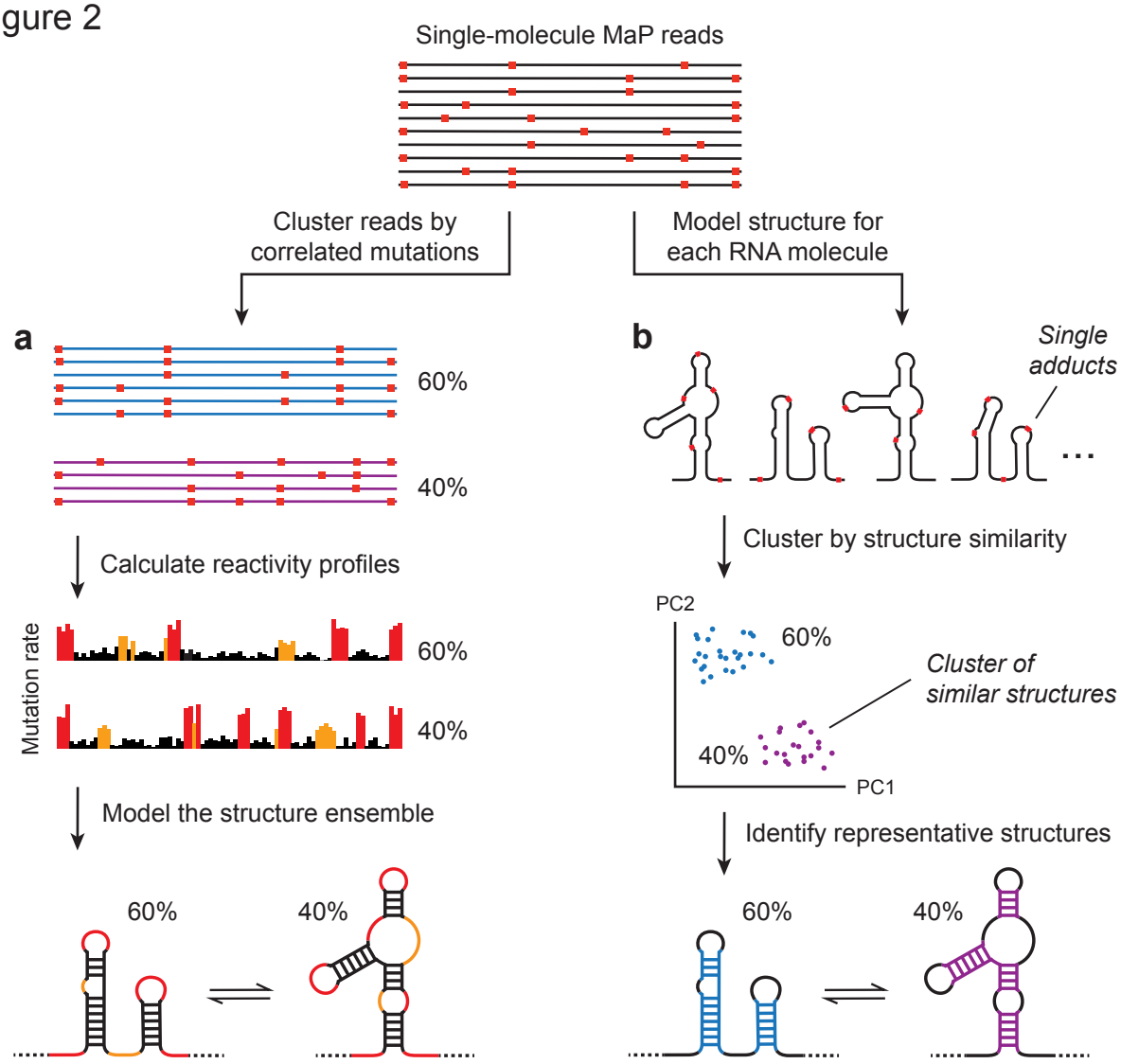

Single-molecule MaP reads
Cluster reads by correlated mutations
Model structure for each RNA molecule
a
60%
40%
Calculate reactivity profiles
Mutation rate
60%
40%
Model the structure ensemble
60%
40%
b
Single adducts
...
Cluster by structure similarity
PC2
60%
Cluster of similar structures
40%
PC1
Identify representative structures
60%
40%

Figure 3

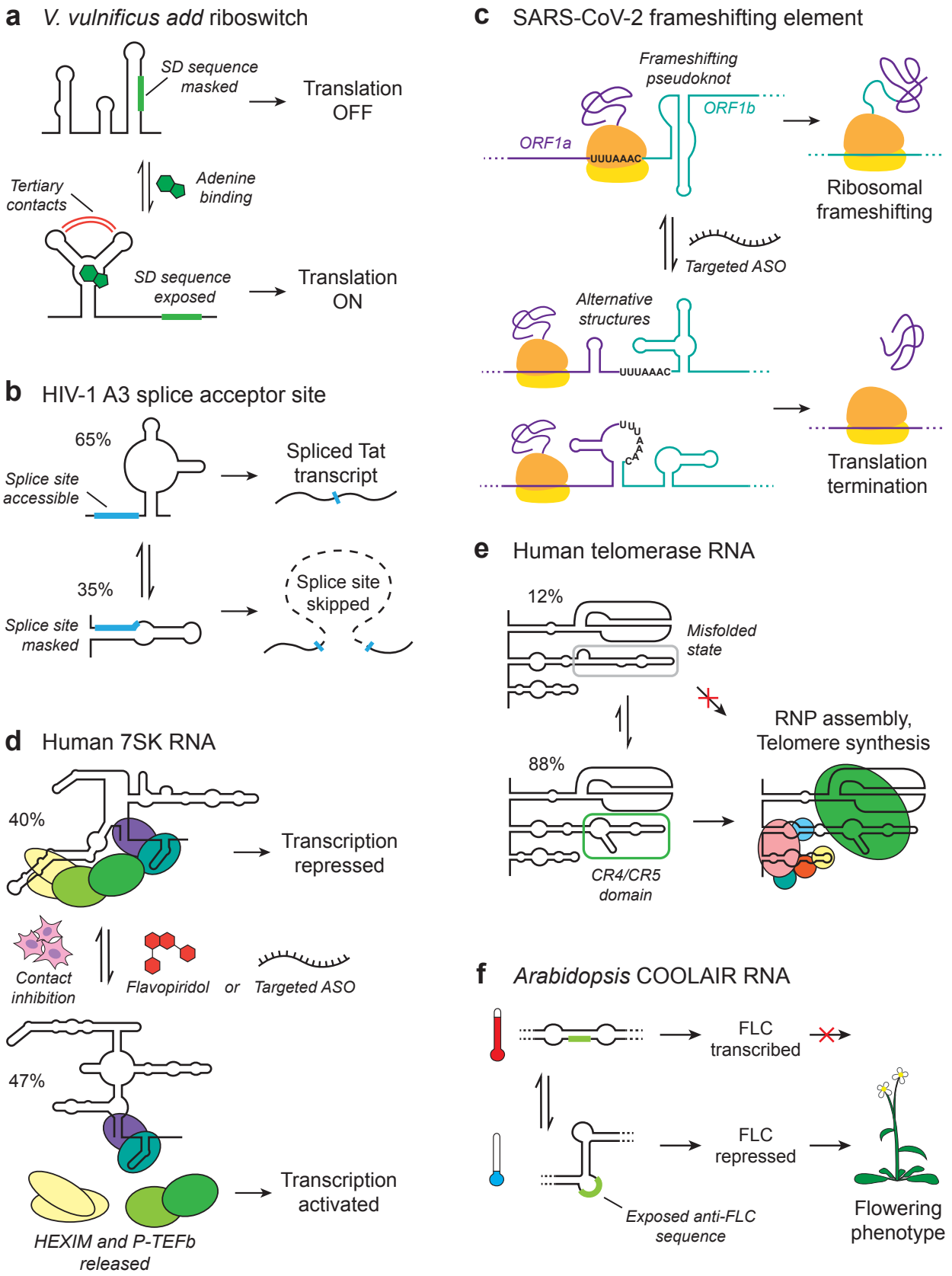
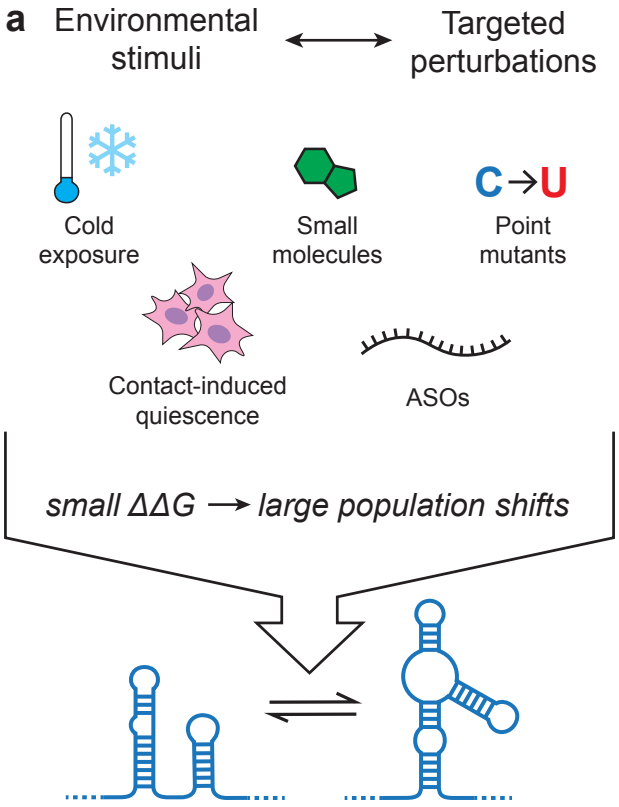
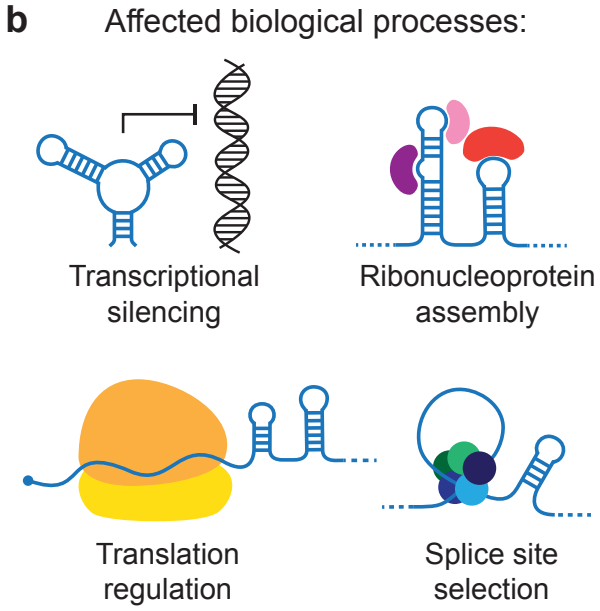

a V. vulnificus add riboswitch
SD sequence masked
Translation OFF
Tertiary contacts
Adenine binding
SD sequence exposed
Translation ON
b HIV-1 A3 splice acceptor site
65%
Splice site accessible
Spliced Tat transcript
35%
Splice site masked
Splice site skipped
c SARS-CoV-2 frameshifting element
Frameshifting pseudoknot
ORF1a
ORF1b
UUUAAAC
Ribosomal frameshifting
Targeted ASO
Alternative structures
UUUAAAC
Translation termination
d Human 7SK RNA
40%
Transcription repressed
Contact inhibition
Flavopiridol
or
Targeted ASO
47%
Transcription activated
HEXIM and P-TEFb released
e Human telomerase RNA
12%
Misfolded state
RNP assembly, Telomere synthesis
88%
CR4/CR5 domain
f Arabidopsis COOLAIR RNA
FLC transcribed
FLC repressed
Exposed anti-FLC sequence
Flowering phenotype

Figure 4


a
Environmental stimuli
Targeted perturbations
Cold exposure
Small molecules
C→U
Point mutants
Contact-induced quiescence
ASOs
small ΔΔG → large population shifts


b
Affected biological processes:
Transcriptional silencing
Ribonucleoprotein assembly
Translation regulation
Splice site selection