Article

Single-Molecule Correlated Chemical Probing: A Revolution in RNA Structure Analysis

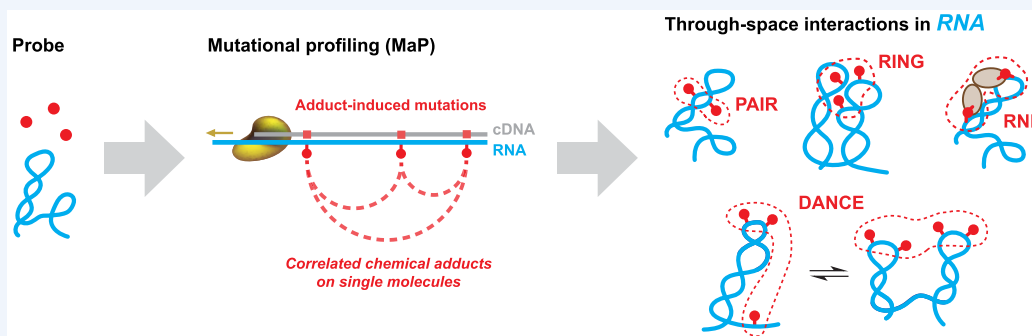Anthony M. Mustoe,[||] Chase A. Weidmann,[||] and Kevin M. Weeks*

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations

CONSPECTUS: RNA molecules convey biological information both in their linear sequence and in their base-paired secondary and tertiary structures. Chemical probing experiments, which involve treating an RNA with a reagent that modifies conformationally dynamic nucleotides, have broadly enabled examination of short- and long-range RNA structure in diverse contexts, including in living cells. For decades, chemical probing experiments have been interpreted in a per-nucleotide way, such that the reactivity measured at each nucleotide reports the average structure at a position over all RNA molecules within a sample. However, there are numerous important cases where per-nucleotide chemical probing falls short, including for RNAs that are bound by proteins, RNAs that form complex higher order structures, and RNAs that sample multiple conformations.

Recent experimental and computational innovations have started a revolution in RNA structure analysis by transforming chemical probing into a massively parallel, single-molecule experiment. Enabled by a specialized reverse transcription strategy called mutational profiling (MaP), multiple chemical modification events can be measured within individual RNA molecules. Nucleotides that communicate structurally through direct base pairing or large-scale folding—unfolding transitions will react with chemical probes in a correlated manner, thereby revealing structural complexity hidden to conventional approaches. These single-molecule correlated chemical probing (smCCP) experiments can be interpreted to directly identify nucleotides that base pair (the PAIR-MaP strategy) and to reveal long-range, through-space structural communication (RING-MaP). Correlated probing can also define the thermodynamic populations of complex RNA ensembles (DANCE-MaP). Complex RNA−protein networks can be interrogated by cross-linking proteins to RNA and measuring correlations between cross-linked positions (RNP-MaP).

smCCP thus visualizes RNA secondary and higher-order structure with unprecedented accuracy, defining novel structures, RNA−protein interaction networks, time-resolved dynamics, and allosteric structural switches. These strategies are not mutually exclusive; in favorable cases, multiple levels of RNA structure — base pairing, through-space structural communication, and equilibrium ensembles — can be resolved concurrently. The physical experimentation required for smCCP is profoundly simple, and experiments are readily performed in cells on RNAs of any size, including large noncoding RNAs and mRNAs. Single-molecule correlated chemical probing is paving the way for a new generation of biophysical studies on RNA in living systems.

■ KEY REFERENCES

- Homan, P. J.; Favorov, O. V.; Lavender, C. A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N. V.; Weeks, K. M. Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 13858−13863.[1] *The paper that started the smCCP field. We show for the first time that MaP can be used to detect correlated chemical modifications in RNA, introduce the RING*
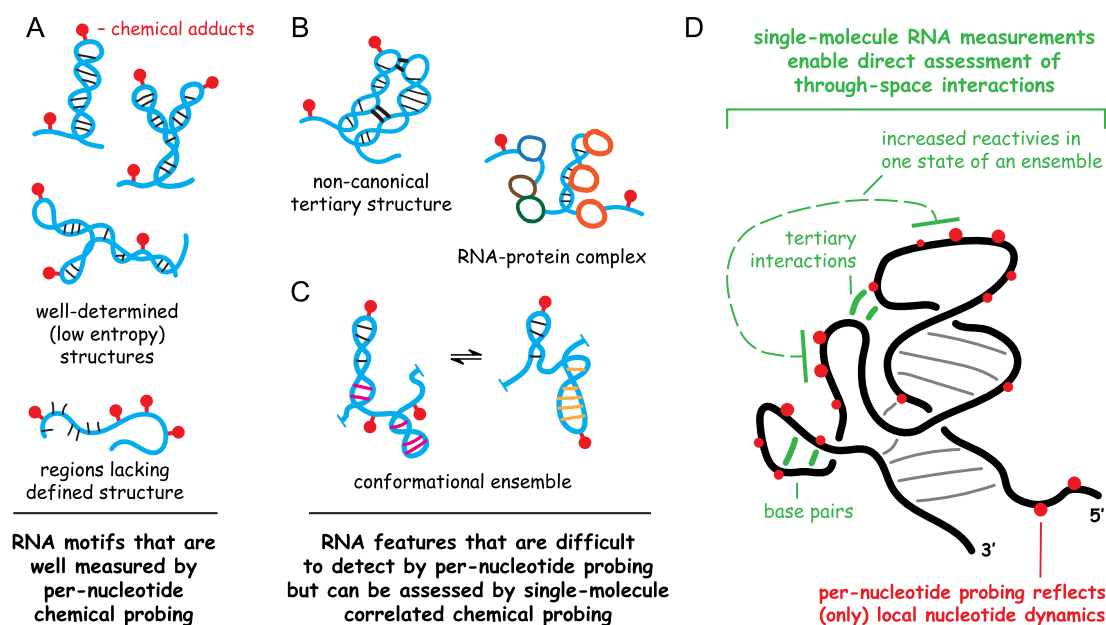
Figure 1. Comparison of per-nucleotide and single-molecule correlated chemical probing. (A) Structures that per-nucleotide probing characterizes well. (B, C) Classes of structures better probed by smCCP than by per-nucleotide probing. (D) Through-space interactions measurable by chemical probing.

framework, and use this insight to map conformational ensembles and through-space interactions in model RNAs.

- Mustoe, A. M.; Lama, N. N.; Irving, P. S.; Olson, S. W.; Weeks, K. M. RNA base pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 24574−24582.[2] *Recognition that smCCP experiments can be used to directly identify RNA base pairs. The method introduced, PAIR-MaP, makes possible highly accurate analysis of RNA secondary structure, in cells.*

- Weidman, C. A.; Mustoe, A. M.; Jariwala, P. B.; Calabrese, J. M.; Weeks, K. M. Analysis of RNA-protein networks with RNP-MaP defines functional hubs on RNA. *Nature Biotechnol.* **2021**, *39*, 347−356.[3] *Illustration that many experiments are better when "MaP-ed". RNA-protein interactions are captured by cross-linking. smCCP detects both protein binding sites and their interprotein networks, without prior knowledge of the proteins involved. RNP-MaP is used to define structural and functional communities in mouse Xist and human XIST lncRNAs.*

- Olson, S. W.; Turner, A.-M. W.; Arney, J. W.; Saleem, I.; Weidmann, C. A.; Margolis, D. M.; Weeks, K. M.; Mustoe, A. M. Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP. *Mol. Cell* **2022**, *82*, 1708−1723.[4] *First demonstration that smCCP can measure multiple layers of RNA structure, simultaneously. Introduces the DANCE framework for measuring RNA structural ensembles, including assignment of PAIR and RING correlations to distinct ensemble states. Used to define the conformational switch in the 7SK RNA that governs transcriptional activation by P-TEFb.*

THE OFT-HIDDEN COMPLEXITY OF RNA STRUCTURE

RNA molecules are inherently driven to fold back on themselves into complex secondary structures consisting of

A−U, G−C, and G−U base-pairing interactions and frequently into higher-order tertiary structures that are stabilized by a diverse array of noncanonical interactions.[5,6] However, our understanding of the functional roles of RNA structure has been limited by the difficulty of measuring RNA structure, particularly in cells. Chemical probing experiments are one of the oldest and most broadly accessible strategies for interrogating RNA structure and use chemical reagents to preferentially modify unpaired or conformationally dynamic RNA regions. These experiments have made it possible to experimentally monitor RNA structure at *per-nucleotide* resolution.[7−9] For RNAs that form well-determined structures, probing data can be used to guide accurate modeling of RNA structure (Figure 1A). However, there are many cases where per-nucleotide methods do not work well, and these challenging cases often correspond to RNAs that mediate important biology.

Per-nucleotide methods suffer from critical deficiencies. First, per-nucleotide chemical probing primarily measures local nucleotide conformational flexibility. That's it. It is often impossible to establish *why* a nucleotide is conformationally flexible or not. For example, per-nucleotide methods cannot distinguish whether a nucleotide is constrained due to base pairing versus tertiary interactions or because it is bound by an RNA-binding protein (Figure 1B). Second, conventional per-nucleotide probing is a time averaged ensemble measurement. If an RNA samples multiple conformations, or is partially bound by small-molecule or protein ligands, the per-nucleotide signal reflects a rough average over all states in the sample (Figure 1C). Thus, per-nucleotide probing often gives an incomplete picture of the most interesting and biologically impactful RNA structures.

SINGLE MOLECULE CORRELATED CHEMICAL PROBING AND MUTATIONAL PROFILING

Many features of RNA structure that are missed by per-nucleotide probing methods can be monitored if reactions are
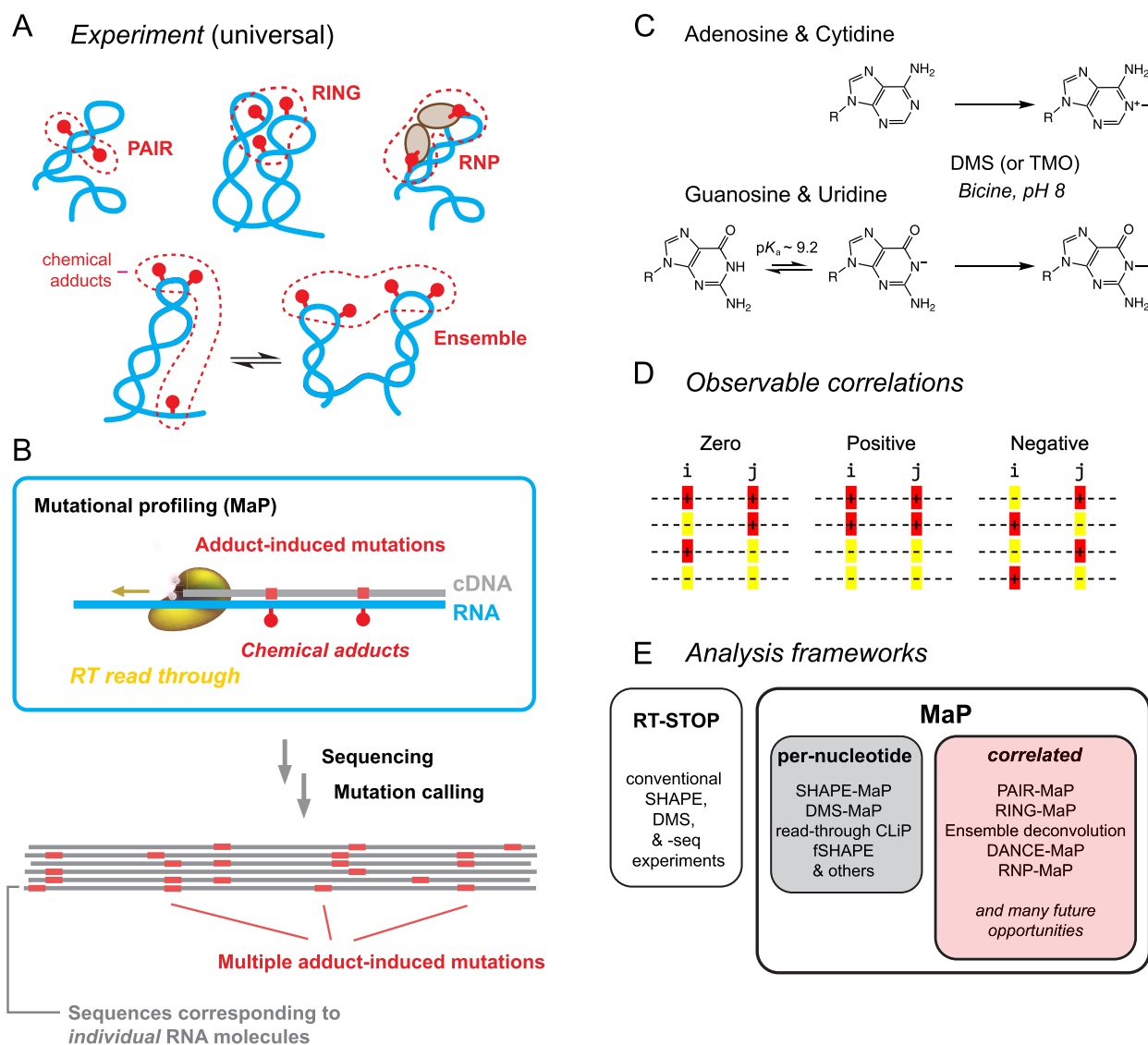
A   *Experiment* (universal)

B

C   Adenosine & Cytidine

DMS (or TMO)
Bicine, pH 8

Guanosine & Uridine

D   *Observable correlations*

E   *Analysis frameworks*

Figure 2. Conceptual framework for smCCP. (A) Filtering observed chemical adducts to reveal distinct tiers of RNA structure. (B) The mutational profiling (MaP) strategy, which leverages relaxed fidelity reverse transcription (RT).[1,13] (C) Schemes for methylation of all four ribonucleotides at their base pairing faces, based on carrying out reactions under well-buffered conditions.[2] (D) Illustration of correlated modifications at nucleotides *i* and *j*. (E) Comparison of per-nucleotide versus single-molecule probing methods.

measured simultaneously at *multiple* nucleotides in the same RNA strand (Figure 1D).[1] RNA structures link the molecular environments of sequence-distant nucleotides, and these interacting (or communicating) nucleotides will react with chemical probes in a *correlated* manner. These correlated reactivities in turn encode information on higher-order features of an RNA. The beauty of single-molecule correlated chemical probing (smCCP) is its experimental simplicity. smCCP experiments use essentially the same straightforward methods as conventional per-nucleotide experiments. The RNA is treated with a chemical reagent, which, over the course of the experiment, forms chemical adducts across the molecule. Structurally linked nucleotides are recurrently modified together or are modified in a mutually exclusive manner, and analysis frameworks then parse these relationships to identify individual base pairs,[2,10] through-space structural communication,[1,11,12] RNA-protein (RNP) interactions,[3] and conformational ensembles (Figure 2A).[4]

Mutational profiling (MaP) is the conceptual insight that makes smCCP possible. MaP was originally invented to accurately read out per-nucleotide probing information using massively parallel sequencing,[13,14] but we quickly recognized that MaP also enables measurement of through-space structural communication in RNA.[1] MaP exploits relaxed fidelity conditions that allow a reverse transcriptase enzyme to *read through* chemically modified nucleotides, leaving behind a mutation or small deletion at the sites of chemical adducts in the synthesized complementary DNA (cDNA; Figure 2B).[1,13,15] Modification information is read out by massively parallel sequencing, where each mutation-containing cDNA corresponds to a single modified RNA molecule.

To date, smCCP has been primarily performed using the classic chemical reagent dimethyl sulfate (DMS). DMS reacts efficiently at the canonical base pairing faces of adenine and cytosine, and recent innovations allow DMS to react with the pairing faces of guanosine and uracil as well, albeit with lower efficiencies (Figure 2C).[2] DMS is a favored reagent because it
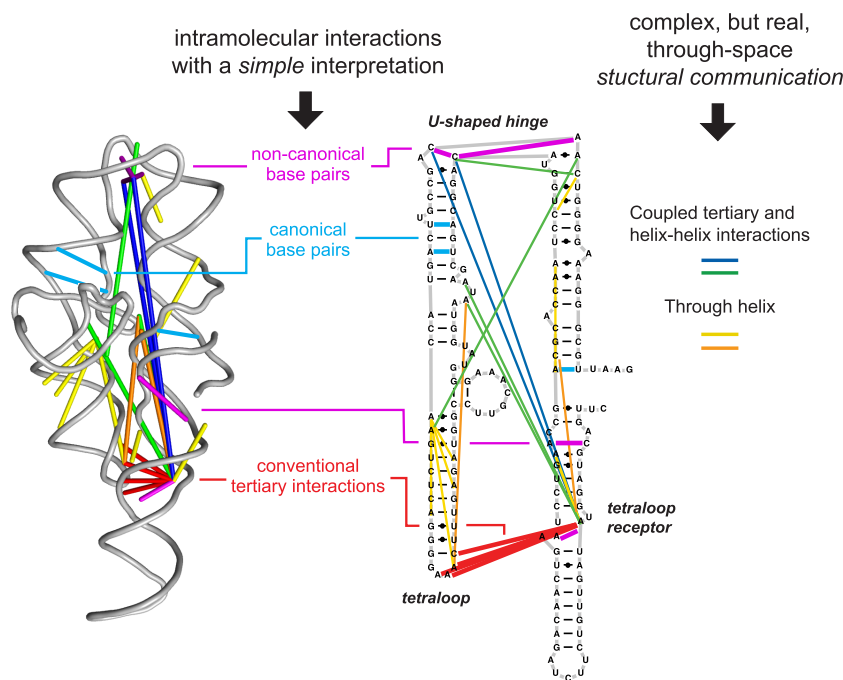
Figure 3. Visualization of the rich, but complex, classes of correlations observed using smCCP in the absence of structure-level filters and algorithms. Adapted with permission from ref 1. Copyright 2014, National Academy of Sciences.

is highly soluble, reactive, and cell permeable, supporting multiple modifications within single RNA molecules. Because DMS forms adducts on the base-pairing face of each nucleotide, nontemplated nucleotides are incorporated at a high rate during MaP. Probes other than DMS have varying mechanisms of action, but the procedural steps for probing are nearly the same for all smCCP experiments.

Using MaP, sites of chemical adducts are detected across the length of an RNA strand and computational analysis then reveals correlations (Figure 2D). If two nucleotides are individually reactive and react together no more often than expected by their independent probabilities, their correlation is zero. By contrast, if two nucleotides in the same RNA strand tend to form chemical adducts jointly, then the two sites show positive correlation. Reaction at one nucleotide can also be correlated with the lack of reactivity at another, leading to a negative correlation. Until the invention of MaP, all chemical probing experiments performed with RNA were interpreted in a per-nucleotide way (Figure 2E, left). Now, multiple levels of information can be extracted from a smCCP experiment depending on the algorithm used to interpret the data. In the following sections, we describe how smCCP information can be used to define RNA structure at increasing levels of complexity (Figure 2E, right).

WHAT SINGLE-MOLECULE CORRELATED PROBING MEASURES

The first smCCP experiments emphasized the richness and complexity of smCCP data.[1] Pioneering smCCP analyses of the P546 domain of a group I intron, which forms a U-shaped structure,[16] produced dense information corresponding to many ways two nucleotides might show correlated reactivity, or not. First, we detected correlated reactivities between multiple canonical and noncanonical base pairs (Figure 3, left). Second, this RNA is stabilized by an RNA tetraloop that binds to a tetraloop receptor motif, and we detected strong

correlations reflective of the close-in-space tetraloop—receptor interaction. In addition, correlated nucleotides revealed other correct and specific—but very complex, and *not* close-in-space—coupled tertiary and helix—helix interactions (Figure 3, right). Some of these complex tertiary interactions likely reflect coordinated large-scale movements of the two arms of the U, indicative of structural breathing. Thus, in the absence of prior knowledge, teasing out correlation signals that are distinct to each level of RNA structure is nontrivial and has required ongoing development of sophisticated computational strategies.

DIRECT DETECTION OF BASE PAIRS

Base pairing generally protects RNA nucleobases from DMS modification. Nonetheless, analysis of our early MaP data sets revealed that even base-paired nucleotides are measurably reactive, enabling direct detection of base pairing from rare, correlated modification events.[2,10] Mechanistically, transient base pair opening permits DMS modification of a nucleobase within a pair.[2] Once one nucleobase is modified, the base pair is permanently destabilized, increasing the odds of modification at nearby opposing nucleobases and giving rise to a measurable correlation (Figure 4A). However, these base-pairing correlations are weak, requiring the development of strategies to reliably identify authentic pairing signals. Key breakthroughs enabling in-cell base pair detection included devising conditions that support DMS modification of uracil and guanine (Figure 2C), amplifying signals computationally by summing modification events over three-nucleotide windows (Figure 4B) and developing algorithms that specifically detect correlations arising from paired bases.

The resulting strategy identifies RNA duplexes from correlations that occur between complementary three-nucleotide-wide regions, which we call PAIRs (pairs ascertained from interacting RNA regions).[2] Nucleotide reactivity and correlation strength prioritize PAIRs as high versus medium
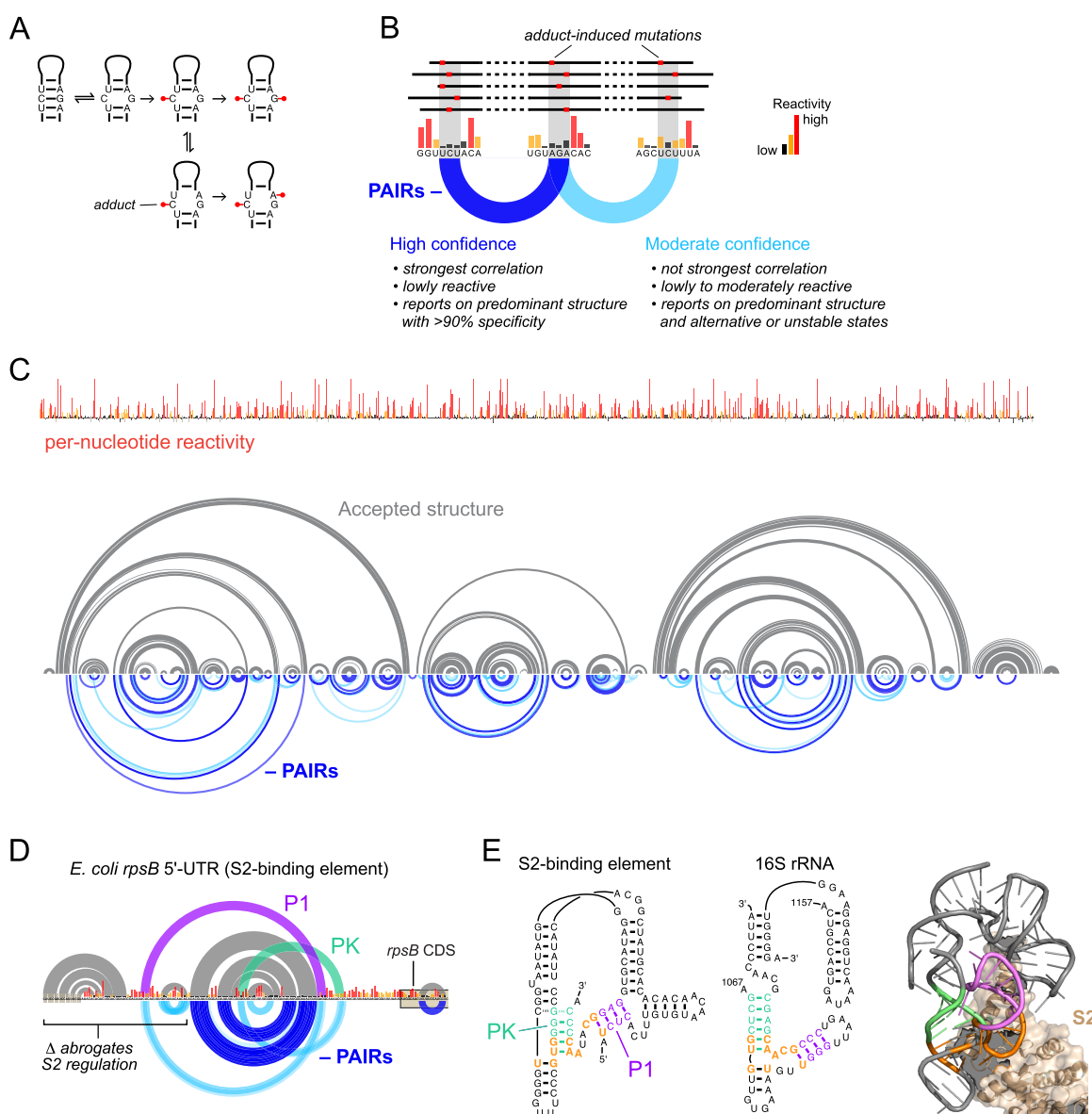
Figure 4. PAIR strategy for direct base pair detection. (A) Mechanism underlying detecting correlated modifications at base-paired nucleotides. (B) Strategy for identifying high and moderate confidence PAIRs. PAIR interactions are shown as arcs. (C) Per-nucleotide and PAIR data for the 1542 nt long *E. coli* 16S rRNA under (natively extracted) cell-free conditions. Per-nucleotide reactivities and PAIR correlations are obtained from the same experiment. (D) In-cell PAIRs for the S2 binding element located in the 5′-UTR of the *E. coli rpsB* gene. (E) Structural homology between the S2 mRNA binding element and S2 binding site in the 16S rRNA. Conserved nucleotides are shown in orange; P1 and PK helices are in color. Adapted with permission from ref 2. Copyright (2019) National Academy of Sciences.

confidence (Figure 4B). High-confidence PAIRs provide remarkably specific measures of a predominant RNA structure (positive predictive value > 95%), with ~45% of helices generating detectable PAIRs. For example, PAIR analysis applied to the *E. coli* 16S rRNA revealed numerous high-confidence PAIRs that precisely mirror the known structure, effectively defining its global architecture (Figure 4C). Medium confidence PAIRs report on less stable or alternatively paired states. These two types of PAIRs provide a direct visualization of RNA folding landscapes that, especially in cells, is more accurate and informative for modeling RNA secondary structure than provided by conventional per-nucleotide reactivities (Figure 4C). PAIRs also often provide direct evidence of long-range helices and pseudoknots,[2] which are particularly challenging to model confidently from reactivity data alone.

In one representative example, PAIR analysis provided critical insight into regulation of the *E. coli rpsB* gene by its encoded protein product, ribosomal protein S2 (Figure 4D).[2] The *rpsB* 5′-UTR encodes an element that binds excess S2 to inhibit translation and enable autoregulation of S2 protein levels. Ribosomal protein autoregulatory elements typically fold into structures that resemble three-dimensional motifs within rRNA, but prior analyses had failed to identify such homologies in the S2 element.[17] In-cell smCCP analysis detected numerous PAIRs that defined the structural architecture of the S2 element, including a metastable long-range helix (P1) and a pseudoknot (PK; Figure 4D), which possesses clear homology to the S2 binding site on the rRNA (Figure 4E). The PK helix was only observed in cells, emphasizing the value of in-cell probing.[2] In sum, PAIR analysis provides direct visualization of both stable and
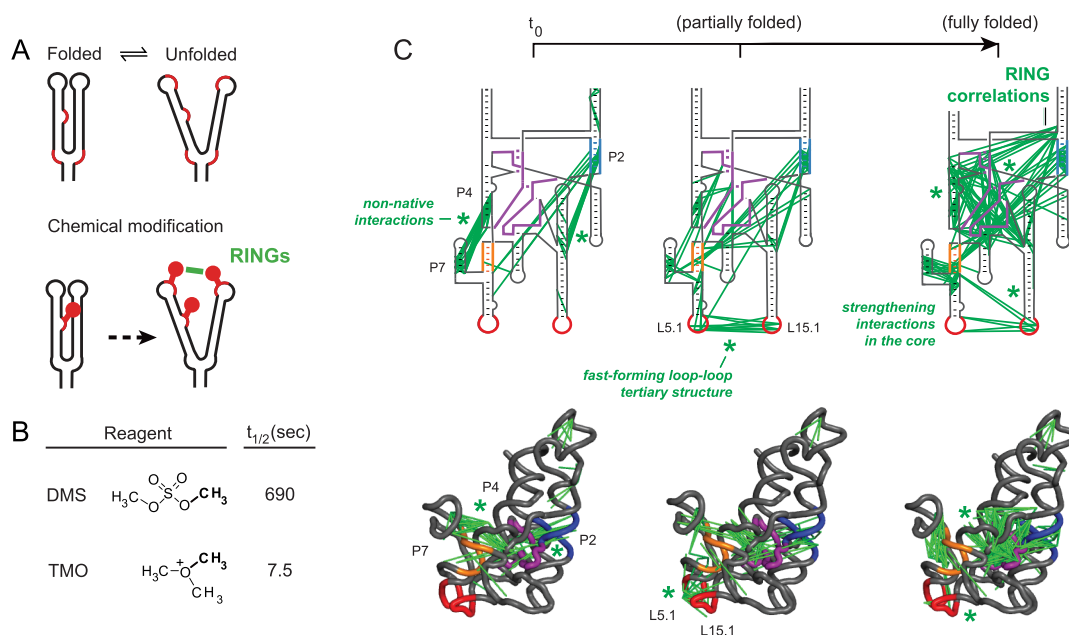
Figure 5. Time-resolved smCCP analysis of a complex RNA folding reaction. (A) Mechanism for detecting RING through-space tertiary interactions. (B) Comparison of DMS and TMO reagents. (C) Pairwise, through-space RINGs observed as a function of time for Mg^{2+}-induced folding of the *B. stearothermophilus* ribonuclease P catalytic domain. RINGs are shown as green lines superimposed on base pairing and three-dimensional models of the RNA. Mechanistically informative RINGs are emphasized with asterisks. Adapted with permission from ref 19. Copyright 2020, American Chemical Society.

alternative base pairing states and dramatically improves the ability to model complex RNA structures in cells.

THROUGH-SPACE STRUCTURAL COMMUNICATION

The original and most direct application of smCCP is to identify strong correlations between reactive nucleotides, called RNA interaction groups (RINGs), that report on (often complex) through-space structural communication networks (Figure 5A).[1] After filtering out correlations arising from base pairing,[11,18] RINGs provide a measure of long-range communication often reflective of tertiary structure.

Large RNAs fold via complex pathways that can be partially inferred from—but are not measured definitively by—per-nucleotide probing. We used smCCP to follow the Mg^{2+}-induced folding of the 265-nucleotide catalytic core of a bacterial RNase P enzyme.[19] DMS-based smCCP experiments typically require a ~5 min probing reaction, which prevents resolving fast dynamics. We thus introduced the chemical probe trimethyloxonium (TMO), which reacts 90-fold faster than DMS (Figure 5B).[19] In the absence of Mg^{2+} (the prefolded state), we observed multiple RINGs indicative of non-native interactions (Figure 5C, left). Upon adding Mg^{2+}, RINGs formed rapidly between nucleotides in two loops that form a long-range tertiary structure, followed by strengthening of RINGs in the core of the RNA (Figure 5C, middle and right). RINGs thus revealed that folding of the RNase P catalytic core proceeds in an unexpected and nonhierarchical way: The L5.1−L15.1 loop−loop tertiary structure forms early and guides formation of subsequent base pairs and tertiary interactions.[19] Even for an RNA whose structural biogenesis has been studied for decades, smCCP illuminated previously undetected, but critical, steps in the folding pathway.

The difference in insight provided by per-nucleotide and smCCP can be dramatic, as exemplified by in-cell smCCP studies of the bacterial ribosome.[12] Based on conventional per-nucleotide analysis, binding by the antibiotic spectinomycin alters DMS reactivity at only a single nucleotide in the 1542-nucleotide-long 16S RNA (Figure 6A). The protected nucleotide, C1192, forms hydrogen bonds with the antibiotic (Figure 6B).[20] By contrast, RING analysis of these same probing data revealed extensive changes in reactivity networks that extend throughout the 16S RNA (Figure 6C). Nucleotides that show dense correlations with other nucleotides can be grouped into four major networks, and these networks recapitulate the known domains of the ribosome (Figure 6D). In the absence of spectinomycin, these domain-specific networks are sparsely linked to each other, indicating that distinct domains move largely independently of one another (Figure 6C, top). By comparison, spectinomycin binding increases the strength and quantity of through-space structural correlations, both within each domain and between domains, with structural communication extending tens of angstroms from the spectinomycin binding site (Figure 6C, E). These increases in RING correlations show how a small antibiotic, interacting at a single site, functions as a jam that globally restricts the conformational freedom of the megadalton ribosome, preventing domains from moving independently as required during translation.[12,21]

COMPREHENSIVE RNA ENSEMBLE ANALYSIS

smCCP provides a solution to one of the most challenging problems in RNA structure analysis: measuring RNA structural ensembles. Many RNAs fold into ensembles of alternative structures, a subset of which are linked to RNA regulatory functions.[22] Conventional per-nucleotide analysis reports averages over all molecules in a sample and is fundamentally unable to capture such structural complexity. We realized early on that smCCP had the potential to resolve multiple coexisting states.[1] Each state within an RNA ensemble will have a unique
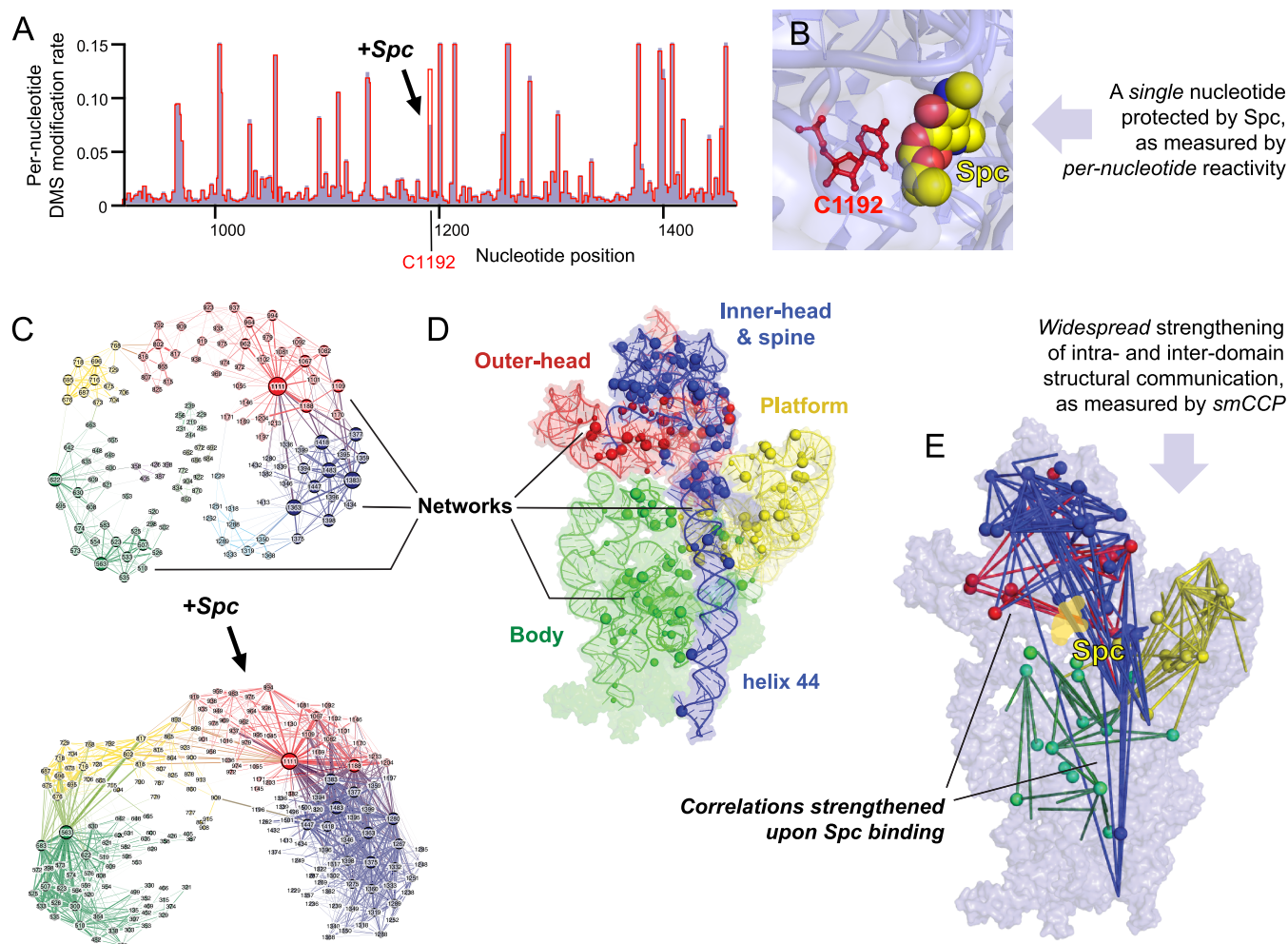
Figure 6. Per-nucleotide and smCCP of the *E. coli* 16S rRNA in fully assembled 30S subunits, in cells. (A) Per-nucleotide analysis indicates spectinomycin (Spc) binding alters reactivity of a single C nucleotide. (B) Spc binds near nucleotide C1192 (PDB: 4V56).[21] (C) smCCP data resolve four major structural networks (blue, red, yellow, green). Nucleotides are shown as nodes, correlation strength by line thickness. (D) Networks, defined *de novo* by smCCP, visualized on the three-dimensional structure of the 30S ribosome. Strongly correlated nucleotides are shown as spheres. Body and Platform networks correspond closely to conventional domain assignments, whereas two RING networks occur in the Head domain, which we designated outer-head (red) and inner-head and spine domains (blue). (E) Visualization of strengthened correlations in context of the three-dimensional structure of the 30S subunit. Image illustrates most significant correlations in the presence of Spc, minus those in the absence of an antibiotic. Spc binding site is shown in yellow. Adapted from ref 12.

reactivity profile, reflective of its underlying structure. Ensemble information is thus encoded as multivalent correlations between all modified positions in a molecule (Figures 1D, 2A, 7A).

We devised a maximum likelihood algorithm to determine whether smCCP data are consistent with a single per-nucleotide reactivity profile or multiple coexisting profiles.[4] For RNAs that populate ensembles, this strategy determines how many reactivity patterns contribute to the overall profile and yields per-nucleotide reactivity profiles and relative populations for each ensemble member with populations ≥ 10% (Figure 7A). Significantly, once the reactivity profiles have been deconvoluted, it is then possible to assign each smCCP read, corresponding to individual RNA molecules, to each ensemble state (Figure 7B). Consequently, the strategies discussed above can be applied to measure PAIRs and RINGs for *each* state in an ensemble, creating an unprecedently rich view of RNA structure. We call this integrated ensemble determination strategy DANCE, or deconvolution and annotation of RNA conformational ensembles.[4]

We validated DANCE using the *add* adenine riboswitch system, which populates a two-state equilibrium consisting of translation ON and OFF states.[4,23] The ON state features an adenine-binding aptamer domain and accessible ribosome binding site, whereas the OFF state masks the ribosome binding site (SD) via base pairing (Figure 7C). DANCE visualizes this equilibrium in remarkable detail, with reactivities and PAIRs mapping the known secondary structure in each state, and RINGs identifying the tertiary structure unique to the ON state (Figure 7C). Populations measured by DANCE are thermodynamically accurate: we could directly measure the K_d for adenine binding from the ON state population. Thus, DANCE provides biophysical level insight, in a (simple) chemical probing experiment that can be performed in cells.

We used DANCE to examine the structure of the human 7SK RNP complex in cells (Figure 8).[4] The 7SK RNA regulates global cellular transcription by binding and inhibiting the transcription factor P-TEFb.[24] Multiple groups, employing diverse approaches, had developed conflicting models for the 7SK RNA structure, in part, because 7SK functions as an
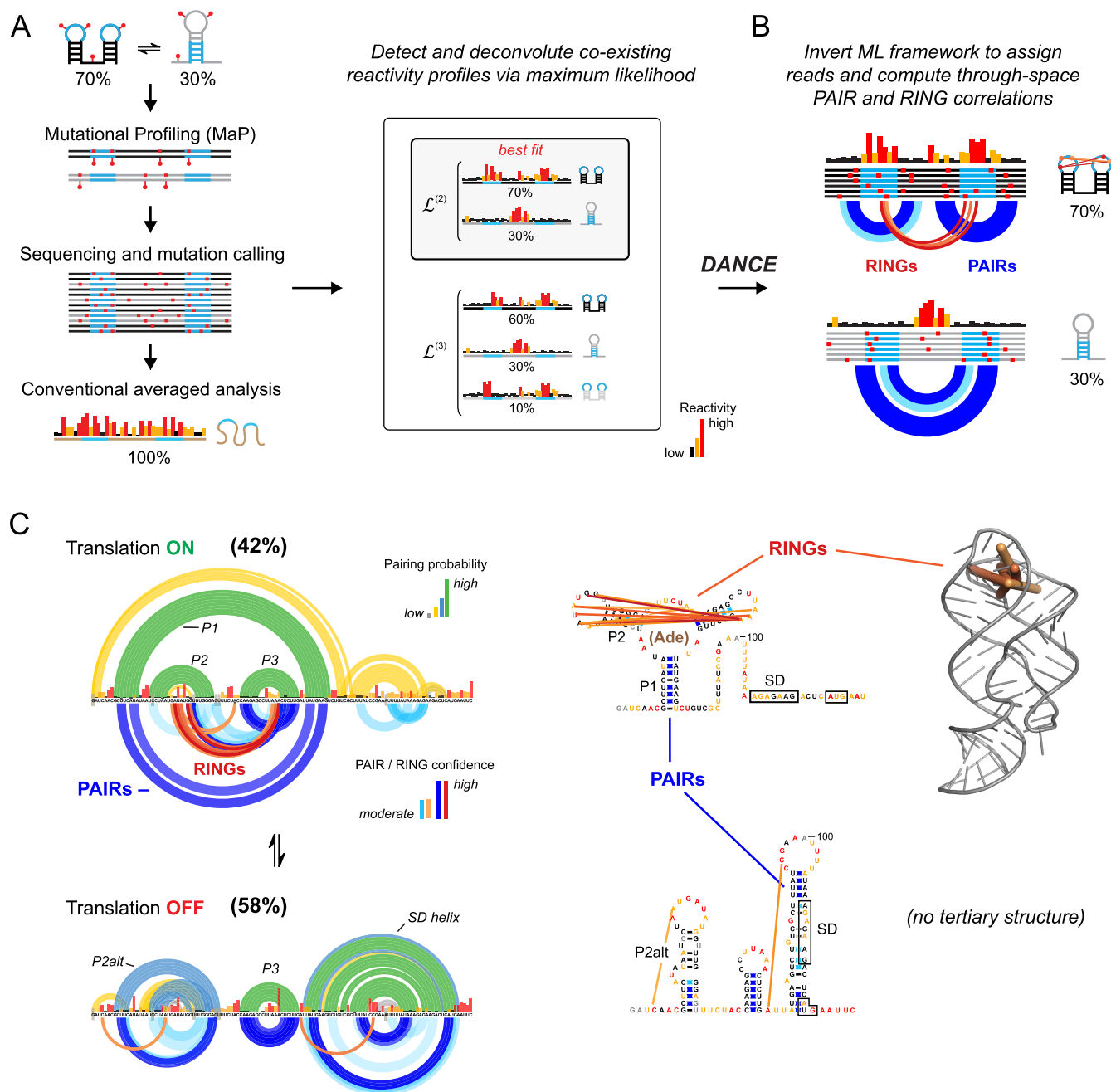
Figure 7. DANCE strategy for comprehensive characterization of RNA ensembles. (A) Illustration of deconvolution of an RNA ensemble, based on smCCP. (B) The DANCE algorithm. (C) DANCE applied to the *add* adenine riboswitch, which folds into translation ON and OFF states in an adenine-dependent manner. (Left) DANCE resolved reactivities, PAIRs, RINGs, and state-specific pairing probabilities computed for each state. Through-space interactions are shown with arcs. (Middle) The same data shown using conventional secondary structure diagrams. (Right) ON state RINGs superimposed on the crystal structure of the aptamer domain (PDB: 4TZX).[36] Adapted with permission from ref 4. Copyright 2022, Elsevier.

ensemble and does not exist as a single state. In-cell ensemble deconvolution resolved these discrepancies and revealed that 7SK populates three structural states, comprising a P-TEFb-bound and two P-TEFb-released states (Figure 8A). DANCE analysis further revealed PAIR and RING correlations critical for defining the structures of each state, including potential tertiary interactions, unique to the P-TEFb-released state (Figure 8B). Functionally, the 7SK ensemble links structural remodeling of a "P-TEFb aptamer domain" to formation of new structures in a "release factor binding domain", creating an

allosteric mechanism for regulating P-TEFb binding and release (Figure 8C).

DANCE also allowed us to measure how the 7SK structural ensemble relates to global transcriptional activity.[4] For example, in fast-growing and more transcriptionally active Jurkat cells, 7SK binds less P-TEFb than in slower growing RPE-1 cells (Figure 8D). Transcription is further down-regulated in quiescent RPE-1 cells, resulting in an additional shift in 7SK toward the P-TEFb bound state. The transcriptional activator flavopiridol does the opposite, pushing the 7SK population toward P-TEFb-released states (Figure 8D). We
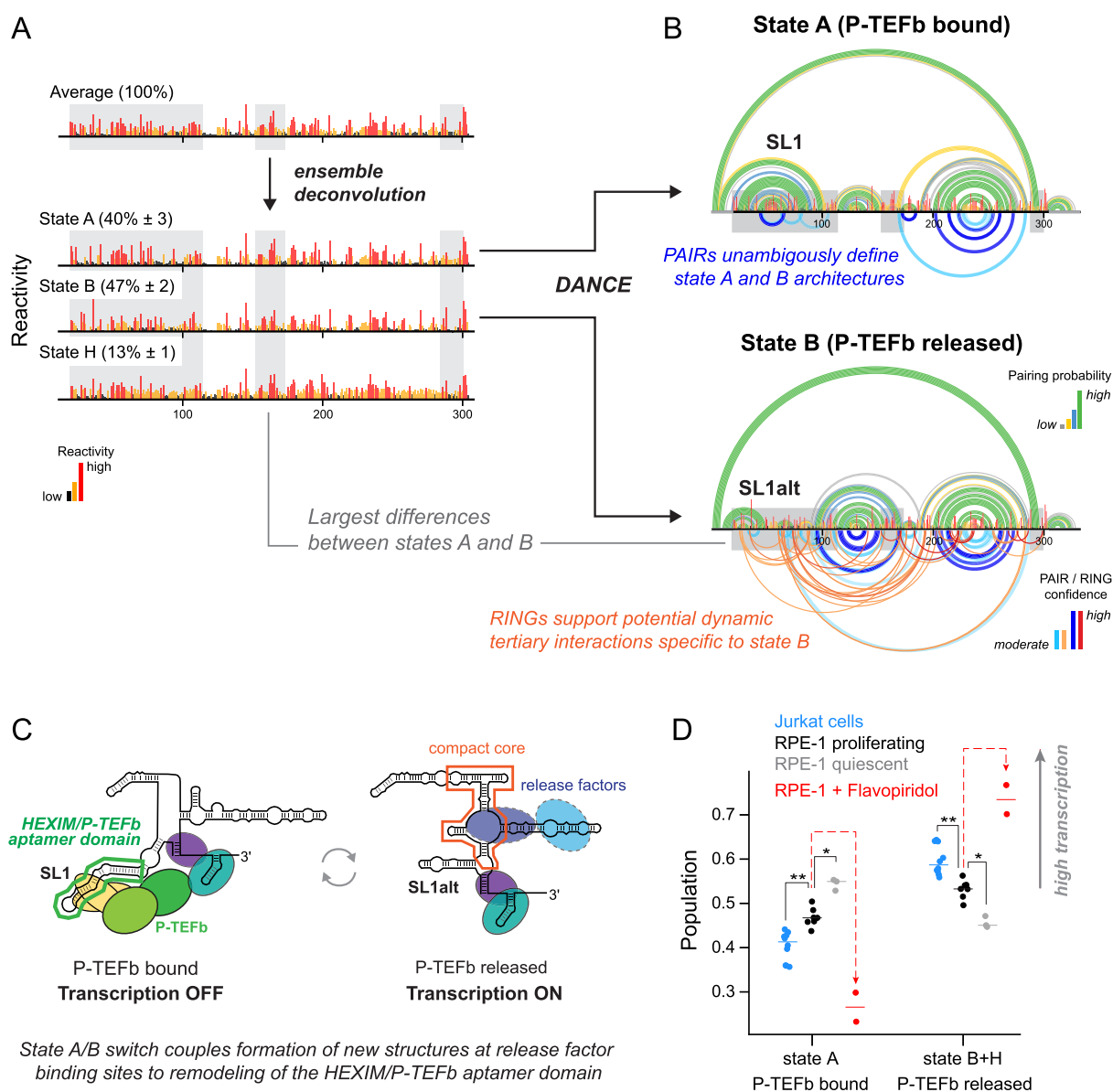
Figure 8. Discovery of a large-scale structural switch in the human 7SK RNA. (A) In-cell DANCE experiments resolve three states from what is otherwise a heterogeneous averaged reactivity profile. (B) PAIR-supported structural models for states A and B, corresponding to P-TEFb bound and released states. State B features a dense network of RINGs, suggestive of tertiary structure within the central junction. State H (not shown) has a heterogeneous structure consistent with an alternative P-TEFb-released state. (C) Model for how the 7SK ensemble functions as an allosteric switch to sequester and release P-TEFb. (D) Population of 7SK states changes across cell types and dynamically remodels in response to cell growth state and transcriptional stress (Flavopiridol). Adapted with permission from ref 4. Copyright 2022, Elsevier.

could also intentionally upregulate transcription using antisense oligos that stabilize P-TEFb-released states of 7SK, establishing the 7SK ensemble as a potential drug target.

Recent studies by other groups provide additional examples where smCCP enabled discovery of functional RNA ensembles in viral RNAs and plant lncRNAs.[25−27] Collectively, these studies reveal how RNA structural ensembles integrate cellular signals to control biology. smCCP specifically makes it possible to "see" these once-invisible RNA mechanisms.

■ PROBING RNA−PROTEIN INTERACTION NETWORKS

Many RNA structure probing experiments can be improved by inventing a MaP version. In one compelling example, we devised a smCCP strategy to measure higher-order RNA-

protein interactions (RNP-MaP).[3] RNP-MaP uses the bifunctional chemical probe SDA (succinimidyl 4,4′-azipentanoate) to cross-link proteins bound to RNA. The NHS (N-hydroxysuccinimide) ester moiety of SDA reacts with protein lysine residues, and the diazirine moiety is photoactivated with UV light to trigger reaction with proximal RNA nucleotides. The cross-linked proteins are then digested, leaving behind small adducts that are read out by MaP (Figure 9A). As lysine residues are highly enriched in RNA-binding protein domains,[28] this strategy captures most RNA−protein interactions. RNP-MaP yields two distinct classes of information (Figure 9B). Protein cross-links on RNA generate *mutation* signals during MaP and measure RNA−protein interactions at nucleotide resolution. *Correlations* between cross-links then
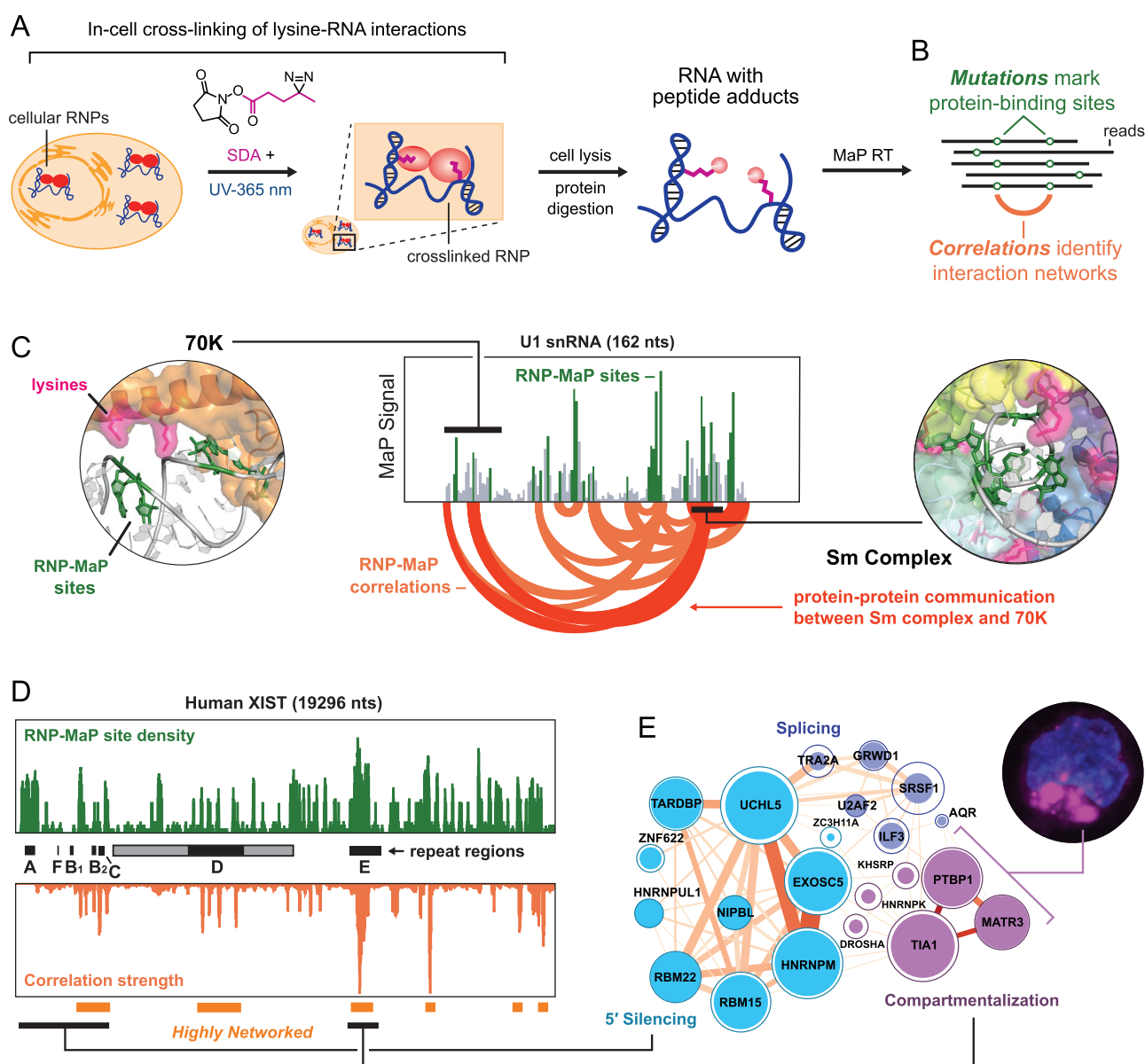
Figure 9. Design and application of RNP-MaP. (A) In-cell RNP-MaP probing, where an NHS-diazirine (SDA) cross-links RNA-protein complexes (RNPs). Cross-link sites are read out by MaP reverse transcription. (B) Data reveal cross-link-induced mutations and correlations, indicative of protein-binding sites and interaction networks, respectively. (C) RNP-MaP specifies protein binding sites (green) from the background signal (gray) and identifies protein−protein communication (orange) on the U1 small nuclear RNA. Highlighted interactions are projected onto high-resolution structures (PDB: 3CW1).[37] Communication between 70K and the Sm protein ring is emphasized with red arcs. (D) RNP-MaP site density (top) and summed strengths of nucleotide correlations (bottom) over 51 nucleotide windows, visualized across the human XIST RNA. Locations of functionally conserved core (black) and expanded (gray) tandem repeat arrays and of highly networked regions are emphasized. (E) Community analysis of XIST-binding proteins yields three functional groups. Two communities (5′ silencing and compartmentalization) correspond to proteins that bind highly networked regions of XIST. Micrograph of RNP granules formed by an RNA including XIST region E (purple), visualized in the context of the nucleus (DAPI stain, blue). Adapted with permission from ref 3. Copyright 2021, Nature Portfolio.

reveal *networks* of protein−protein communication on RNA molecules in cells.

RNP-MaP accurately defined protein networks in diverse RNPs of known structure.[3] For example, the U1 small nuclear ribonucleoprotein (snRNP) is stabilized by assembly of the ring-shaped Sm protein complex with U1 RNA.[29] RNP-MaP detected multiple, strong correlations between nucleotides at the Sm ring binding site and other proteins in the U1 snRNP, including a long distance interaction with the 70K protein, revealing the protein-network architecture of this complex (Figure 9C).

In larger RNAs, RNP-MaP identifies analogous highly networked "hubs", where multiple strong correlations converge (Figure 9D). Such hub analysis enabled us to identify conserved functional regions across the 20 kb XIST RNA, even in the absence of significant linear sequence homology between species.[3] For example, sequences in a motif called region E are divergent between human XIST and mouse Xist RNAs. Nonetheless, region E contains the highest density of protein-binding nucleotides and of protein−protein communication networks in both human XIST (Figure 9D) and

mouse Xist RNAs,[3] emphasizing their likely functional importance.

We further used RNP-MaP network strengths to classify proteins with known XIST binding sites[30] into functional "communities". We identified a highly networked community of region E-binding proteins and showed that the network of region E RNA and E-binding proteins is important for assembly of XIST into an RNP compartment (Figure 9E). In sum, by inventing a smCCP version of an RNA−protein cross-linking experiment, we defined *network intensity* as a new property of RNA−protein interactions in cells and used this insight to identify and characterize novel functional elements in very large RNAs.

■ NUANCES AND LIMITATIONS

Most fundamentally, smCCP is a *single molecule* strategy and therefore does not measure structural communication between different RNAs (or other) molecules. Nonetheless, with careful design, it is clearly possible to measure *inter*molecular interactions, as exemplified by measuring ligand binding to riboswitches or the ribosome,[4,12] or protein binding via the RNP-MaP strategy (Figures 6, 7, and 9).[3] We look forward to seeing many future innovative examples.

Experimental details matter for successful smCCP experiments. First, it is important to confirm that each sequencing read corresponds to an individual molecule: each analyzed sequence must be unique and not an artifact of library preparation. Second, smCCP involves multiple-hit, reagent-induced weakening of RNA structure, and accumulation of chemical adducts has the potential to push RNAs toward non-native conformations. Data to date suggest that misfolding artifacts are rare,[1,2,4,25−27] but care is required to ensure that RNA integrity is maintained. Strategies and reagents that employ shorter probing times are likely to yield more accurate structural data. Third, smCCP experiments provide superior information when performed under conditions that monitor all four ribonucleotides. DMS and other alkylating reagents are strongly acidifying, and thus rigorous buffering is essential.[1,2] Fourth, smCCP data are best interpreted using high read depths (typically corresponding to 0.3−1 million unique RNA molecules). New reagents, improved MaP strategies, and alternative algorithms will likely lower sequencing depth requirements for some smCCP applications (as has been reported for ensemble deconvolution[26]). Further reductions in sequencing costs are needed for smCCP to be practical on whole-transcriptome scales.

smCCP data are physically and structurally rich, but complex. The spatial resolution of detectable interactions is enormous, extending from individual base pairs, angstroms apart (Figure 4),[2] to long-range tertiary interactions spanning many nanometers (Figures 5 and 6).[12,19] Distinguishing short versus long-range interactions, and direct versus indirect, remains imperfect. All strategies for deconvoluting smCCP data involve assumptions and simplifications, and care is required not to misinterpret a complex RNA system. We think the 7SK RNA, which is highly expressed and contains significant but not extreme structural heterogeneity, represents a strong emerging reference system for smCCP analysis in cells and to understand differences among deconvolution algorithms. Going forward, we anticipate many advances as increasingly sophisticated and well-validated algorithms are developed to interpret and deconvolute rich smCCP data.

■ FUTURE OPPORTUNITIES

Any experiment that creates chemical adducts on RNA can, in principle, be performed as a smCCP experiment. In coming years, we envision smCCP being used to monitor in-cell, time-dependent assembly of RNA−protein and RNA−RNA complexes and to measure RNA−small molecule interactions. MaP can also detect many types of epigenetic RNA modifications, and smCCP is well-suited for resolving how epigenetic modifications influence higher-order RNA structure and ternary complex assembly.

A major opportunity exists in melding smCCP with SHAPE probing, which would yield two major advantages relative to using DMS. SHAPE chemistry exploits hydroxyl-selective electrophiles that react with the ribose 2′-OH group, with the consequence that many SHAPE reagents react evenly with all four ribonucleotides regardless of base identity.[8,9,14,31] SHAPE also provides a holistic, biophysically rigorous measurement of local nucleotide flexibility.[14] Recent studies have shown potential success using single-molecule SHAPE probing for ensemble deconvolution.[27] It is likely that both (*i*) new highly reactive, short half-life SHAPE reagents that destabilize local RNA structure and (*ii*) new reverse transcription enzymes and strategies, capable of detecting these adducts with high efficiency, need to be identified to realize the full potential of smCCP applications. The future clearly lies in a universal experiment that measures both nucleotide-level RNA biophysics and internucleotide structural communication in cells.

There are also significant opportunities to harness smCCP data to model complex RNA and RNA−protein three-dimensional structures. RING measurements clearly convey abundant information about RNA tertiary structure,[1,4,11,12,19] but are often challenging to interpret. New algorithms, such as machine and deep learning approaches, should advance analysis of three-dimensional RNA structures, especially in cells. There may also be opportunities to combine smCCP with complementary mutation scanning strategies[32] to obtain even higher-dimensional data sets.

New sequencing technologies will eventually extend the sequence length limits for smCCP to enable visualization of much longer-distance communication events in RNA. Innovations-in-waiting include harnessing highly processive reverse transcriptases[33] and long-read methods to sequence full-length cDNAs or performing direct RNA nanopore sequencing,[34,35] and are the likely future of smCCP.

In sum, we hope that the profound simplicity and experimental concision of MaP-based chemical probing, the established and emerging algorithmic frameworks for data analysis, and the expansive room for technology growth inspire RNA biologists to pick up their pipettes and try MaP and smCCP. The field will continue to benefit and mature rapidly from these collective discoveries.

■ AUTHOR INFORMATION

Corresponding Author

Kevin M. Weeks − *Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290, United States;* ⦿ orcid.org/0000-0002-6748-9985; Email: weeks@unc.edu

Authors

Anthony M. Mustoe − *Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Department of Molecular and Human Genetics, and Therapeutic Innovation Center (THINC), Baylor College of Medicine, Houston, Texas 77030, United States*

Chase A. Weidmann − *Department of Biological Chemistry, Center for RNA Biomedicine, University of Michigan Medical School, Ann Arbor, Michigan 48109, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.accounts.2c00782

Author Contributions

‖Contributed equally. CRediT: All authors collaborated to write and edit the manuscript. CRediT: **Anthony Mustoe** writing-original draft (equal), writing-review & editing (equal); **Chase Weidmann** writing-original draft (equal), writing-review & editing (equal); **Kevin M. Weeks** writing-original draft (equal), writing-review & editing (equal).

Notes

The authors declare the following competing financial interest(s): K.M.W. is an advisor to and holds equity in Ribometrix.

Biographies

Anthony Mustoe is an Assistant Professor of Biochemistry at Baylor College of Medicine. His laboratory combines chemical biology and computation to understand how RNA sequences program cellular function. The Mustoe lab is committed to training the next generation of scientists at the experimental-computational interface.

Chase Weidmann is an Assistant Professor of Biological Chemistry and Faculty Scholar of the Center for RNA Biomedicine at the University of Michigan. His laboratory is focused on the roles of noncoding RNA−protein complexes in biology and human disease. Chase has a strong interest in mentoring and hopes his research program will contribute to a diverse scientific future.

Kevin Weeks is Kenan Distinguished Professor of Chemistry at the University of North Carolina. His laboratory works at the interface of chemistry, RNA biology, and genomics. The overarching vision of the Weeks laboratory is to prepare team members for fulfilling careers as long-term scientific leaders.

■ REFERENCES

(1) Homan, P. J.; Favorov, O. V.; Lavender, C. A.; Kursun, O.; Ge, X.; Busan, S.; Dokholyan, N. V.; Weeks, K. M. Single-Molecule Correlated Chemical Probing of RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 13858−13863.

(2) Mustoe, A. M.; Lama, N. N.; Irving, P. S.; Olson, S. W.; Weeks, K. M. RNA Base-Pairing Complexity in Living Cells Visualized by Correlated Chemical Probing. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 24574−24582.

(3) Weidmann, C. A.; Mustoe, A. M.; Jariwala, P. B.; Calabrese, J. M.; Weeks, K. M. Analysis of RNA-Protein Networks with RNP-MaP Defines Functional Hubs on RNA. *Nat. Biotechnol.* **2021**, *39*, 347−356.

(4) Olson, S. W.; Turner, A.-M. W.; Arney, J. W.; Saleem, I.; Weidmann, C. A.; Margolis, D. M.; Weeks, K. M.; Mustoe, A. M. Discovery of a Large-Scale, Cell-State-Responsive Allosteric Switch in the 7SK RNA Using DANCE-MaP. *Mol. Cell* **2022**, *82*, 1708−1723.

(5) Brion, P.; Westhof, E. Hierarchy and Dynamics of RNA Folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113−137.

(6) Vicens, Q.; Kieft, J. S. Thoughts on How to Think (and Talk) about RNA Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, e2112677119.

(7) Weeks, K. M. Advances in RNA Structure Analysis by Chemical Probing. *Curr. Opin. Struct. Biol.* **2010**, *20*, 295−304.

(8) Strobel, E. J.; Yu, A. M.; Lucks, J. B. High-Throughput Determination of RNA Structures. *Nat. Rev. Genet.* **2018**, *19*, 615−634.

(9) Mitchell, D.; Assmann, S. M.; Bevilacqua, P. C. Probing RNA Structure in Vivo. *Curr. Opin. Struct. Biol.* **2019**, *59*, 151−158.

(10) Krokhotin, A.; Mustoe, A. M.; Weeks, K. M.; Dokholyan, N. V. Direct Identification of Base-Paired RNA Nucleotides by Correlated Chemical Probing. *RNA* **2017**, *23*, 6−13.

(11) Dethoff, E. A.; Boerneke, M. A.; Gokhale, N. S.; Muhire, B. M.; Martin, D. P.; Sacco, M. T.; McFadden, M. J.; Weinstein, J. B.; Messer, W. B.; Horner, S. M.; Weeks, K. M. Pervasive Tertiary Structure in the Dengue Virus RNA Genome. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 11513−11518.

(12) Sengupta, A.; Rice, G. M.; Weeks, K. M. 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.* **2019**, *17*, e3000393.

(13) Siegfried, N. A.; Busan, S.; Rice, G. M.; Nelson, J. A. E.; Weeks, K. M. RNA Motif Discovery by SHAPE and Mutational Profiling (SHAPE-MaP). *Nat. Methods* **2014**, *11*, 959−965.

(14) Weeks, K. M. SHAPE Directed Discovery of New Functions in Large RNAs. *Acc. Chem. Res.* **2021**, *54*, 2502−2517.

(15) Smola, M. J.; Rice, G. M.; Busan, S.; Siegfried, N. A.; Weeks, K. M. Selective 2′-Hydroxyl Acylation Analyzed by Primer Extension and Mutational Profiling (SHAPE-MaP) for Direct, Versatile and Accurate RNA Structure Analysis. *Nat. Protoc.* **2015**, *10*, 1643−1669.

(16) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **1996**, *273*, 1678−1685.

(17) Fu, Y.; Deiorio-Haggar, K.; Anthony, J.; Meyer, M. M. Most RNAs Regulating Ribosomal Protein Biosynthesis in Escherichia Coli Are Narrowly Distributed to Gammaproteobacteria. *Nucleic Acids Res.* **2013**, *41*, 3491−3503.

(18) Hajdin, C. E.; Bellaousov, S.; Huggins, W.; Leonard, C. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-Directed RNA Secondary Structure Modeling, Including Pseudoknots. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5498−5503.

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

(20) Moazed, D.; Noller, H. F. Interaction of Antibiotics with Functional Sites in 16S Ribosomal RNA. *Nature* **1987**, *327*, 389−394.

(21) Borovinskaya, M. A.; Shoji, S.; Holton, J. M.; Fredrick, K.; Cate, J. H. D. A Steric Block in Translation Caused by the Antibiotic Spectinomycin. *ACS Chem. Biol.* **2007**, *2*, 545−552.

(22) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The Roles of Structural Dynamics in the Cellular Functions of RNAs. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 474−489.

(23) 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.

(24) Peterlin, B. M.; Brogie, J. E.; Price, D. H. 7SK SnRNA: A Noncoding RNA That Plays a Major Role in Regulating Eukaryotic Transcription. *Wiley Interdiscip. Rev. RNA* **2012**, *3*, 92−103.

(25) Tomezsko, P. J.; Corbin, V. D. A.; Gupta, P.; Swaminathan, H.; Glasgow, M.; Persad, S.; Edwards, M. D.; Mcintosh, L.; Papenfuss, A. T.; Emery, A.; Swanstrom, R.; Zang, T.; Lan, T. C. T.; Bieniasz, P.; Kuritzkes, D. R.; Tsibris, A.; Rouskin, S. Determination of RNA Structural Diversity and Its Role in HIV-1 RNA Splicing. *Nature* **2020**, *582*, 438−442.

(26) Morandi, E.; Manfredonia, I.; Simon, L. M.; Anselmi, F.; van Hemert, M. J.; Oliviero, S.; Incarnato, D. Genome-Scale Deconvolution of RNA Structure Ensembles. *Nat. Methods* **2021**, *18*, 249−252.

(27) 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.

(28) Krüger, D. M.; Neubacher, S.; Grossmann, T. N. Protein-RNA Interactions: Structural Characteristics and Hotspot Amino Acids. *RNA* **2018**, *24*, 1457−1465.

(29) So, B. R.; Wan, L.; Zhang, Z.; Li, P.; Babiash, E.; Duan, J.; Younis, I.; Dreyfuss, G. A U1 SnRNP-Specific Assembly Pathway Reveals the SMN Complex as a Versatile Hub for RNP Exchange. *Nat. Struct. Mol. Biol.* **2016**, *23*, 225−230.

(30) Van Nostrand, E. L.; Freese, P.; Pratt, G. A.; Wang, X.; Wei, X.; Xiao, R.; Blue, S. M.; Chen, J.-Y.; Cody, N. A. L.; Dominguez, D.; Olson, S.; Sundararaman, B.; Zhan, L.; Bazile, C.; Bouvrette, L. P. B.; Bergalet, J.; Duff, M. O.; Garcia, K. E.; Gelboin-Burkhart, C.; Hochman, M.; Lambert, N. J.; Li, H.; McGurk, M. P.; Nguyen, T. B.; Palden, T.; Rabano, I.; Sathe, S.; Stanton, R.; Su, A.; Wang, R.; Yee, B. A.; Zhou, B.; Louie, A. L.; Aigner, S.; Fu, X.-D.; Lécuyer, E.; Burge, C. B.; Graveley, B. R.; Yeo, G. W. A Large-Scale Binding and Functional Map of Human RNA-Binding Proteins. *Nature* **2020**, *583*, 711−719.

(31) Kubota, M.; Tran, C.; Spitale, R. C. Progress and Challenges for Chemical Probing of RNA Structure inside Living Cells. *Nat. Chem. Biol.* **2015**, *11*, 933−941.

(32) Cheng, C. Y.; Kladwang, W.; Yesselman, J. D.; Das, R. RNA Structure Inference through Chemical Mapping after Accidental or Intentional Mutations. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 9876−9881.

(33) Guo, L.-T.; Adams, R. L.; Wan, H.; Huston, N. C.; Potapova, O.; Olson, S.; Gallardo, C. M.; Graveley, B. R.; Torbett, B. E.; Pyle, A. M. Sequencing and Structure Probing of Long RNAs Using MarathonRT: A Next-Generation Reverse Transcriptase. *J. Mol. Biol.* **2020**, *432*, 3338−3352.

(34) Stephenson, W.; Razaghi, R.; Busan, S.; Weeks, K. M.; Timp, W.; Smibert, P. Direct Detection of RNA Modifications and Structure Using Single-Molecule Nanopore Sequencing. *Cell Genomics* **2022**, *2*, 100097.

(35) Aw, J. G. A.; Lim, S. W.; Wang, J. X.; Lambert, F. R. P.; Tan, W. T.; Shen, Y.; Zhang, Y.; Kaewsapsak, P.; Li, C.; Ng, S. B.; Vardy, L. A.; Tan, M. H.; Nagarajan, N.; Wan, Y. Determination of Isoform-Specific RNA Structure with Nanopore Long Reads. *Nat. Biotechnol.* **2021**, *39*, 336−346.

(36) Zhang, J.; Ferré-D'Amaré, A. R. Dramatic Improvement of Crystals of Large RNAs by Cation Replacement and Dehydration. *Structure* **2014**, *22*, 1363−1371.

(37) Pomeranz Krummel, D. A.; Oubridge, C.; Leung, A. K. W.; Li, J.; Nagai, K. Crystal Structure of Human Spliceosomal U1 SnRNP at 5.5 A Resolution. *Nature* **2009**, *458*, 475−480.