2-17-2022

Chemical reversible crosslinking enables measurement of RNA 3D distances and alternative conformations in cells

Ryan Van Damme

Kongpan Li

Minjie Zhang

Jianhui Bai

Wilson H. Lee

See next page for additional authors

Authors

Ryan Van Damme, Kongpan Li, Minjie Zhang, Jianhui Bai, Wilson H. Lee, Joseph D. Yesselman, Zhipeng Lu, and Willem A. Velema

Chemical reversible crosslinking enables measurement of RNA 3D distances and alternative conformations in cells

Ryan Van Damme[1,4], Kongpan Li[1,4], Minjie Zhang[1,4], Jianhui Bai[1], Wilson H. Lee [1], Joseph D. Yesselman[2], Zhipeng Lu [1✉] & Willem A. Velema [3✉]

Three-dimensional (3D) structures dictate the functions of RNA molecules in a wide variety of biological processes. However, direct determination of RNA 3D structures in vivo is difficult due to their large sizes, conformational heterogeneity, and dynamics. Here we present a method, Spatial 2′-Hydroxyl Acylation Reversible Crosslinking (SHARC), which uses chemical crosslinkers of defined lengths to measure distances between nucleotides in cellular RNA. Integrating crosslinking, exonuclease (exo) trimming, proximity ligation, and high throughput sequencing, SHARC enables transcriptome-wide tertiary structure contact maps at high accuracy and precision, revealing heterogeneous RNA structures and interactions. SHARC data provide constraints that improves Rosetta-based RNA 3D structure modeling at near-nanometer resolution. Integrating SHARC-exo with other crosslinking-based methods, we discover compact folding of the 7SK RNA, a critical regulator of transcriptional elongation. These results establish a strategy for measuring RNA 3D distances and alternative conformations in their native cellular context.

[1] Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA 90033, USA. [2] Department of Chemistry, University of Nebraska-Lincoln, 832A Hamilton Hall, Lincoln, NE 68588, USA. [3] Institute for Molecules and Materials, Radboud University Nijmegen, Nijmegen, The Netherlands. [4] These authors contributed equally: Ryan Van Damme, Kongpan Li, Minjie Zhang. ✉email: zhipengl@usc.edu; willem.velema@ru.nl

RNA plays critical roles throughout the cell, ranging from carrying genetic information to regulation and catalysis[1]. To perform these tasks, RNA must fold into complex three-dimensional (3D) structures that undergo intricate conformational transitions[2–7]. Physical methods can be applied to elucidate RNA structure, such as nuclear magnetic resonance (NMR), cryo-EM, and crystallography. These approaches have helped characterize RNA structures, often at atomic resolution, but require well-behaved and purified samples, whereas cellular RNA structures can be highly dynamic and heterogenous[2]. Alternatively, numerous low-resolution approaches, such as chemical mapping and crosslinking, are high-throughput and can be applied in vivo. These low-resolution methods can be coupled with ever-improving computational tools to build 3D models[8].

Chemical probing, such as selective 2′-hydroxyl acylation (SHAPE) and dimethyl sulfate (DMS) alkylation, report various aspects of nucleotide flexibility and have been used to constrain local secondary structure predictions[9–13]. Correlated chemical probing methods such as multiplexed •OH cleavage analysis (MOHCA), mutate-and-map (M²), and RNA interacting group mutational profiling (RING-MaP) infer spatial proximity of nucleotides but provides fuzzy distances to constrain 3D modeling[14–18]. While these methods are improvements over 1D DMS chemical mapping, they are often limited to smaller RNAs as they require the two correlated nucleotides on the same sequencing read, and the sequencing coverage scaled exponentially with RNA length. Furthermore, MOHCA and M² are only applicable to in vitro synthetic RNAs, while RING-MaP is limited by the noisy background and low correlation levels[14].

Crosslinking and proximity ligation represent an alternative strategy to capture spatial distances among nucleotides, overcoming the limitations of correlated chemical probing[19]. Recently developed psoralen-crosslinking-based methods, such as PARIS, LIGR-seq, SPLASH, and COMRADES directly capture base pairs either within or between different RNA molecules in high throughput[20–24]. Psoralen crosslinks staggering pyrimidines in opposite strands through [2 + 2] photocycloadditions. At the cost of low efficiency, this reaction offers high specificity, is challenging to reverse, and is limited to uridines in helical regions. Even though the gapped reads from such methods can go down to 15 nucleotides on each arm, unambiguous identification of base pairs remains challenging. Recently reported bifunctional acylating crosslinkers, BINARI, react with the 2′-OH on all four nucleotides and offer an approach to capturing nucleotide pairs in spatial proximity crosslinking capacity to 3D space[25]. However, the nine-step synthesis, large molecular size, and complex reversal mechanism rendered the BINARI compounds unsuitable for cellular application to measure RNA tertiary contacts on a transcriptome-wide level[25].

This study develops highly efficient and accessible 2′-hydroxyl acylation chemistry for crosslink-formation and -reversal in living cells (SHARC), overcoming the technical challenges in the preparation and application of BINARI reagents. We develop an exonuclease (exo) trimming approach to pinpoint crosslinked nucleotides, improving the precision of distance measurements to the crosslinked atoms (2′-O in ribose). The integration of SHARC crosslinking, exo trimming, proximity ligation, and high throughput sequencing (SHARC-exo) enables transcriptome-wide analysis of spatial distances between nucleotides at nanometer resolution in cells, without sequence length limitations. We rigorously benchmarked the distance measurement and structure capture using complex, yet well-studied models in cells, such as the ribosome, spliceosome, 7SL, and RNase P, revealing both static structures, interactions, and alternative conformations.

The incorporation of distance measurements into Rosetta-based 3D modeling dramatically improved structure resolution. We combined SHARC-exo with established methods, such as PARIS and CLIP, to discover compact folding of the 7SK RNA, a critical regulator of transcriptional elongation in higher eukaryotes. These experiments demonstrate the power of integrating multiple orthogonal approaches to capture proximity constraints in complex RNAs to study their structures. Together, we developed cheap and easily synthesized compounds that dramatically outperform known crosslinking tools, providing the community with a strategy for understanding RNA 3D structures and alternative conformations in cells.

Results

Quantitative RNA crosslinking with bifunctional 2′-hydroxyl acylation. Unlike proteins, RNA's overall structure is governed by sparse tertiary contacts (Fig. 1a). Highly structured RNAs as large as 500 nucleotides may only contain a few critical tertiary contacts[6,26–28]. Knowledge of these tertiary contacts significantly improves the modeling of complex RNAs[15,18,29]. To determine these constraints for RNAs in cells, we sought to develop a set of bifunctional and reversible 2′-hydroxyl acylation reagents with flexible linkers (Fig. 1b). To improve accessibility and facilitate optimizations, we focused on a modular design using simple dicarboxylic acids, the length of which can be easily adjusted. Subsequent reversal allows facile analysis of the crosslinked sequences. To synthesize such reagents, we activated simple dicarboxylic acids using 1,1′-carbonyldiimidazole (CDI) in a one-step reaction (Fig. 1c, see Methods in Supplementary Information). We tested crosslinking efficiency on a model self-complementary duplex RNA **1** in vitro, where acylations are expected to occur on un-constrained nucleotides (Fig. 1d).

We activated a set of eight dicarboxylic acids with diverse linker lengths and chemical properties (Fig. 1e, see characterization in Supplementary Figs. 1 and 2a). The efficiency of crosslinking RNA **1** was measured by polyacrylamide electrophoresis (Fig. 1e, Supplementary Fig. 2b). Activated oxalic and succinic acids showed low to modest crosslinking of 1–24% (Fig. 1e), possibly due to the short linker lengths that might be insufficient to bridge 2′-OH groups on opposing strands (see estimated crosslinker lengths in Supplementary Fig. 2a). Activated glutaric acid showed 94% crosslinking, while diglycolic acid, which is similar in size, exhibited significantly lower crosslinking of 27%. The differences between these aliphatic compounds can potentially be explained by the inductive effect of the beta oxygen that substantially increases the reactivity of the activated ester[30], making it more susceptible to hydrolysis. The activated aromatic compounds, terephthalic, isocinchomeronic, and dipicolinic acids all exhibited excellent crosslinking efficiencies between 97–99%, likely due to optimal spacing and conformation of the linkers to bridge opposing 2′-OH groups and favorable reactivity towards 2′-OH moieties as previously demonstrated by aromatic SHAPE reagents[13]. The activated bipyridine compound 6,6′-binicotinic acid showed an apparent crosslinking efficiency of 31%, though solubility in aqueous solution was limited, hampering the exact determination of its crosslinking performance. We selected dipicolinic acid imidazolide (DPI) as a candidate to test further based on these results. To characterize the reaction kinetics of DPI at physiological pH 7.4 at room temperature, we measured its hydrolysis with NMR and found that 50% was hydrolyzed after 5 min (Supplementary Fig. 2c–e). This reaction is significantly faster than the structurally related cell-permeable SHAPE reagent NAI (half-life ~30 min, HEPES buffer pH = 8.0) due to the additional electron-withdrawing group[31], suggesting its potential for rapid RNA crosslinking in cells.
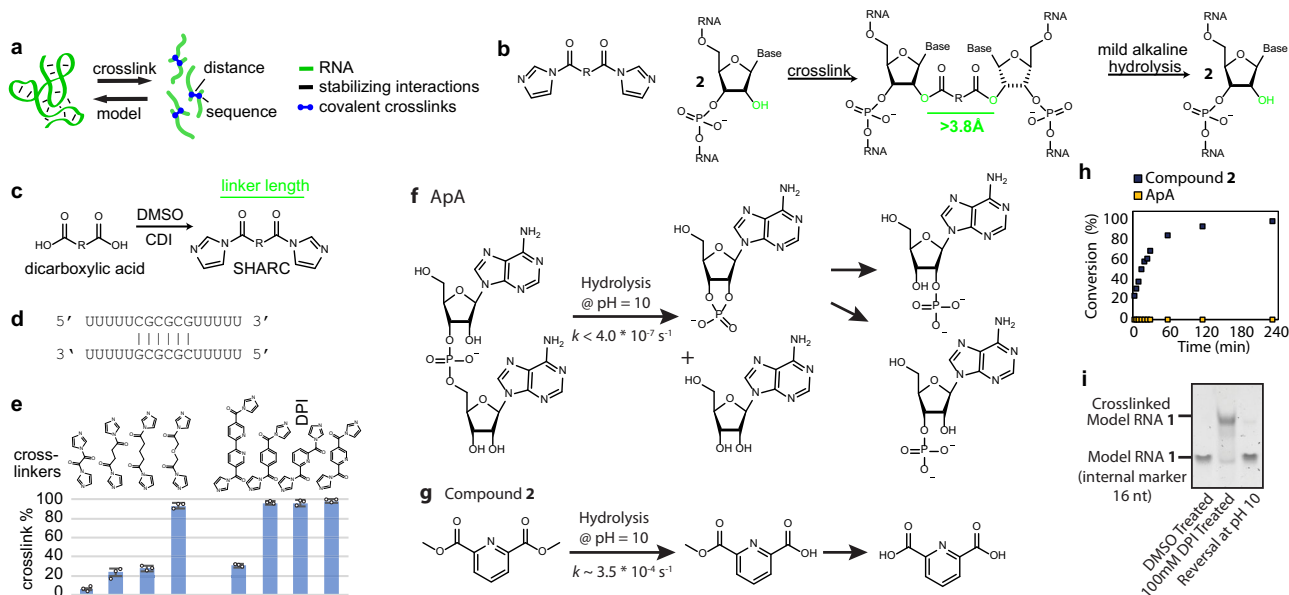
Fig. 1 Efficient RNA crosslinking and reversal using SHARC reagents. a Principle of using crosslinking to determine RNA 3D structures. Spatial distances among nucleotides in an RNA should be sufficient to rebuild its 3D structure. **b** SHARC crosslinking and reversal of ribose 2′ hydroxyls. **c** One-step activation of dicarboxylic acids to produce SHARC reagents. CDI: carbonyldiimidazole. **d** Model RNA 1 homodimer duplex for testing SHARC crosslinking and reversal. **e** Crosslinking efficiency of a series of SHARC reagents on the model RNA 1 duplex. Common names for the 9 dicarboxylic acids to prepare the crosslinkers: oxalic, succinic, diglycolic, glutaric, 6,6′-binicotinic, terephthalic, dipicolinic, and isocinchomeronic. Crosslinking condition: MOPS buffer (pH 7.5 0.1 M KCl, 6 mM MgCl2), 4 h at room temperature. Data are mean ± s.d.; $n = 3$, technical replicates. **f** Hydrolysis kinetics of RNA phosphodiester bonds in a model dinucleotide ApA. **g** Hydrolysis of a model SHARC crosslinking product, compound 2. **h** Measurements of hydrolysis rates for the ApA and model compound 2 (5 mM starting concentration). **i** Example SHARC crosslinking and reversal of model RNA 1 duplex on a 20% urea-denatured TBE polyacrylamide gel. Source data are provided as a Source Data file.

Reversing 2′-hydroxyl crosslinking under mild alkaline conditions without RNA damage. Reversal of the crosslinks is necessary for subsequent sequence analysis. We hypothesized that the lower stability of the 2′-acylation products relative to the phosphodiester bonds could allow selective crosslink reversal without causing RNA chain breaks. To test this, we first analyzed the rate of phosphodiester cleavage in a model RNA dinucleotide ApA (Fig. 1f). We compared it to the methyl ester of dipicolinic acid **2** (Fig. 1g), a simple ester derivative of the SHARC reagent DPI. The two compounds were incubated in a 3:1 mixture of 100 mM borate buffer and DMSO at pH 10.0, and the stability was monitored over time by ^{1}H NMR (Fig. 1h and Supplementary Fig. 2f, g). No degradation of ApA was observed even after 48 h, and the rate constant was estimated to be below $4.0 \times 10^{-7}\,s^{-1}$ (Fig. 1f). In contrast, compound **2** was fully hydrolyzed after ~120 min, with a rate constant of $3.5 \times 10^{-4}\,s^{-1}$ ($R^2 = 0.99$) (Fig. 1g). From this, we concluded that the ~1000-fold difference in rate constant should provide sufficient opportunity to selectively reverse the crosslinks under mild alkaline conditions without RNA damage.

To investigate if the alkaline conditions can be successfully applied to reverse SHARC crosslinks in longer RNA, the model RNA **1** was crosslinked with DPI, purified, and 10 μM of crosslinked RNA was incubated in 100 mM Borate buffer pH 10.0 for 2 h at 37 °C. The crosslinked RNA was nearly fully reversed without apparent degradation (Fig. 1i). Increasing pH to 11.0 did not result in noticeable degradation, suggesting a broad window for robust reversal of crosslinks (Supplementary Fig. 2h, i). Together, we showed that 2′-OH acylation could be easily reversed at moderately alkaline pH without significant RNA damage, opening the possibility for subsequent sequence analysis in various applications. For larger RNAs, degradation may be unavoidable. However, fragmentation is an inherent step in sequencing library preparation, so the residual RNA degradation does not affect subsequent sequence analysis.

Exonuclease trimming: a strategy to determine crosslinking sites at near nucleotide resolution. Having demonstrated efficient SHARC crosslinking and reversal, next we developed a strategy, exonuclease (exo) trimming, to measure inter-nucleotide distances, based on our previously established PARIS method[21,32] (Fig. 2a). Crosslinked RNA samples are first digested with RNase III, which fragments both single and double-stranded RNA into short pieces[21]. RNA fragments are fractionated on a denatured-denatured 2-dimension (DD2D) gel[24], where the second dimension is denser than the first (e.g., 16% vs. 8%). The differential gel densities enable the separation of crosslinked from non-crosslinked fragments. The crosslinked fragments migrate as a smear above the diagonal (Fig. 2a), therefore achieving near 100% purity without the contamination of RNA with mono adducts of the crosslinker. The purified cross-linked fragments are then trimmed by an exonuclease, e.g., RNase R, which removes nucleotides from the 3′ end until it is blocked by the crosslink sites[33]. The trimmed fragments are ligated so that the two arms are joined to form a continuous RNA molecule. After mild alkaline crosslink reversal, the bipartite RNA molecules are reverse transcribed for cDNA library preparation and sequenced. The gapped reads are clustered into duplex groups (DGs, similar to our previous definition[21,34], but include all gapped reads from secondary and tertiary structures). Each group corresponds to one specific pair of nucleotides that are close to each other. The gapped reads should reveal trimmed 3′ ends at a fixed distance from the actual crosslinking sites (~5 nts, see details below). The spatial distances between the crosslinked nucleotides (the 2′-OH groups, to be precise) are determined by the length of the linkers and the flexibility of the RNA structure.

To validate the exo trimming approach, we first applied it to PARIS experiments, where the well-established crosslinking preference of psoralen enables rigorous testing of the trimming efficiency. After RNase R treatment, the reads are significantly
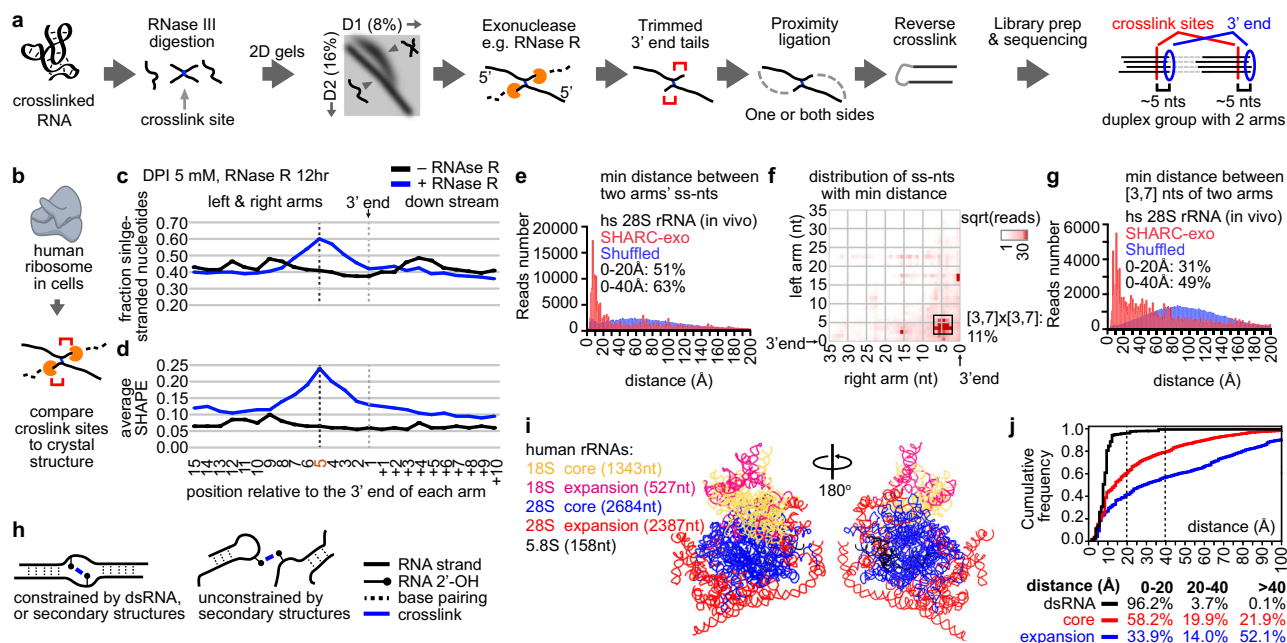
Fig. 2 Exonuclease trimming measures spatial distance at nucleotide resolution in cells. a RNA samples, either in vitro or in cells, are crosslinked, digested to short fragments using RNase III, and the crosslinked fragments isolated using a DD2D gel. Isolated cross-linked fragments are then trimmed with exonucleases, proximally ligated, decrosslinked, ligated with adapters for cDNA library preparation. The putative crosslinking sites are determined based on the trimmed 3′ ends. **b** Benchmarking SHARC-exo against the human ribosome in cells. **c, d** Identification of SHARC crosslinking sites in the human ribosome from cross-linked cells. SHARC-exo condition: 5 mM DPI and 12 h RNase R trimming. Error bars represent standard deviations from two biological replicates. **c** Fraction of single-stranded nucleotides was calculated along each arm of the gapped reads. **d** Average icSHAPE signal was calculated along each arm of the gapped reads. **e** Distribution of minimal distances between single-stranded nucleotides (ss-nts) in the two arms of each gapped-read for the human 28S ribosomal RNA. The blue histogram is the distribution for randomly shuffled reads. **f** Positions of single-stranded nucleotides that are closest to each other on the two arms of each read. The square root of the reads was plotted. All reads were aligned relative to the 3′ end. Positions 3–7 for each arm were boxed, which represent 11% of total coverage. **g** Distribution of minimal distances between nucleotides in the 3–7 range on both arms. The blue histogram is the distribution for randomly shuffled reads. **h** Two types of SHARC crosslinked spatial proximity. **i** Illustration of the core and expansion segments (ESs) in three human cytoplasmic rRNAs. **j** Cumulative distribution of the minimal distances between the two arms of each read for reads mapped to different types of structures. Only non-dsRNA, or tertiary contact, reads are used for the core and expansion segments. Note, the color-coding in panels i and j have different meanings. Source data are provided as a Source Data file.

shorter (Supplementary Fig. 3a). Counting from the 3′ end of each arm of the gapped reads, we observed a strong enrichment of uridines at the third to sixth position, peaking at the fifth nucleotide, suggesting that the psoralen crosslinking to uridine blocked RNase R trimming, leaving ~5 nts at the 3′ end (Supplementary Fig. 3b). In contrast, no enrichment of uridine was observed at the exact location without trimming. Therefore, the exo trimming strategy allows us to pinpoint the crosslinking sites with high precision. As an example, we showed the identification of crosslinking sites in helical regions of the 28S rRNA (Supplementary Fig. 3c–e).

SHARC-exo accurately measures static and dynamic spatial distances in RNA in cells. To test SHARC-exo, we first cross-linked HEK293T cells with DPI, extracted RNA using proteinase K (PK) and the TNA method, fragmented RNA with RNase III, and isolated the crosslinked fragments using the DD2D gel method. We recovered 1.01%, 1.31%, and 1.89% RNA fragments as crosslinked using 5, 12.5, or 25 mM DPI, respectively (Supplementary Fig. 3f–h, comparable to psoralens used in the PARIS method)[24]. PK digests proteins to <6 amino acids long[35], and efficiently removed the vast majority of proteins (Supplementary Fig. 3i, TNA method), therefore the detected interactions are unlikely to be mediated by proteins. We sequenced the SHARC-exo libraries and observed 3.3–14.5% of the reads are gapped, similar to PARIS (Supplementary Table 1)[20,21]. Crosslinked reads are highly reproducible at different DPI concentrations and

RNase R trimming conditions (Supplementary Fig. 3j). The two arms of each gapped-read span a wide range of distances, for example, up to the entire length of rRNAs (1869 and 5070 nucleotides, respectively, Supplementary Fig. 3k). The crosslinked together, these results demonstrated efficient and robust SHARC crosslinking of RNA in cells.

To test the ability of SHARC-exo in measuring spatial distances, we focused on the ribosome due to its high abundance, complex structures, and intermolecular interactions (Fig. 2b)[36,37]. Given that RNA homodimers are rare due to the transient nature of most intermolecular interactions[38], all subsequent analyses of structures in individual RNA species were performed under the assumption that the gapped reads were derived from the same RNA, instead of two identical RNA molecules. We first calculated the fraction of single-stranded nucleotides close to the 3′ end of each arm of the gapped reads, based on the ribosome cryo-EM structure[36] (Fig. 2c, d, and Supplementary Fig. 4a–d). Counting from the 3′ end, trimmed samples exhibited a dramatic increase in the fraction of single-stranded nucleotides between the 1st and 8th nucleotides, with a peak at the 5th (Fig. 2c, ~1.3-fold over non-trimmed). We observed a similar trend when using experimentally determined icSHAPE reactivities for the ribosome[21] (Fig. 2d). The stronger enrichment of icSHAPE signal (Fig. 2d, ~3.7-fold) compared to the counts of single-stranded nucleotides further confirmed the selective crosslinking of unconstrained nucleotides by DPI and the efficient trimming. A/U nucleotides are slightly enriched near the crosslink sites,

likely reflecting their lower base-pairing potential (Supplementary Fig. 4e).

To determine the range and precision of distance measurements by SHARC-exo, we calculated spatial distances between the two arms of each gapped-read in the ribosome cryo-EM model[36,37]. The minimal distance has a narrow distribution with a long tail, where 51% are within 20 Å, with a mode of ~8 Å, close to the physical length of the crosslinker (~7 Å) (Fig. 2e, Supplementary Fig. 2a). In contrast, the distances for randomly shuffled reads have a much broader distribution (Wilcoxon ranksum (WRS) test, $p < 10^{-300}$). To determine whether trimming precisely reveals the crosslinked nucleotides, we searched for nucleotides along each arm closest between the two arms (Fig. 2f). Not surprisingly, the nearest point is the fifth nt, consistent with the highest SHAPE reactivity (not seen on the non-trimmed SHARC data, Supplementary Fig. 5). The distance between 5th ± 2 nts on the two arms follow a narrow distribution, with a mode distance of 8 Å, and 31% reads less than 20 Å, and 49% less than 40 Å (WRS test $p < 10^{-300}$, compared to shuffled reads, Fig. 2g).

The ribosome is a highly dynamic and flexible macromolecular machine. SHARC-exo captures spatial distances of the ribosome in its entire life cycle in cells that include both intra-ribosome dynamics and inter-ribosome contacts. To understand the long tails in distributions (in Fig. 2e, g), we separated distances into ones constrained by extensive base pairs, which are more stable, and those simply in spatial proximity (tertiary motifs), which are more dynamic (Fig. 2h)[39]. We further split tertiary contacts to the stable core and the more flexible Expansion Segments (ESs), many of which are not resolved with cryo-EM (Fig. 2i). As expected, distances constrained by secondary structures are predominantly within 20 Å (dsRNA, 96.2%, Fig. 2j), whereas the core and ES tertiary distances have increasingly broader distributions (58.2% and 33.9% within 20 Å, respectively). Together, this analysis further demonstrated SHARC-exo's high accuracy and the ability to capture heterogeneous conformations in cells.

To test the robustness of SHARC-exo, we compared multiple DPI concentrations and trimming conditions. Regardless of DPI concentration, SHARC-exo produced consistent enrichment of single-stranded nucleotides near the 5th nucleotide (Supplementary Fig. 4a–d). The minimum distances between the two arms are primarily within 20 Å (Supplementary Fig. 5a). However, higher DPI concentrations reduced trimming efficiency, which can potentially be explained due to disruption of the endogenous structure of mono adducts that block trimming. The mono adducts may reduce the resolution in distance measurements. At the same DPI concentration, heavier trimming increased the resolution of spatially proximal nucleotides (Supplementary Fig. 5b–d).

SHARC-exo analysis of RNA structures and interactions in vivo. To test the ability of SHARC-exo in capturing known structures, we extracted spatial distances within 20 Å in the ribosome (Fig. 3a, b, left panels). SHARC-exo measurements (upper right triangles) are highly consistent with distances between icSHAPE-reactive nucleotides in the cryo-EM model for both the 18S and 28S rRNAs (lower left triangles). Shuffling of SHARC-exo reads resulted in random distributions (Fig. 3a, b, right panels). The zoom-in views of the two regions showed both highly consistent distance measurements and ones missed by SHARC-exo (blue and red boxes in areas 1–2, Fig. 3c,d). The missed spatial proximities likely represent tight ribosome regions inaccessible to DPI or nucleotides with steric hindrance.

SHARC-exo captured spatial distances both constrained by secondary structures or simply in spatial proximity (Fig. 3e–m, see more examples in Supplementary Fig. 6). In one instance in the 28S rRNA, RNase R trimming resulted in significantly shortened 3′ ends for both arms of the DG (Fig. 3e). Tracing back to the 5th nucleotides from the 3′ ends, where the crosslinks are expected, we obtained a pair of nucleotides with a spatial distance of 11.1 Å between the 2′ oxygens in the ribosome cryo-EM model (Fig. 3f, g). SHARC-exo also captured intermolecular interactions. For example, SHARC-exo precisely mapped a tertiary contact between the loop on 5.8 S helix 9 ES 3 (H9ES3) and an internal bulge on 28S H54 (Fig. 3h). This interaction is stabilized by base stacking between 5.8 S U126 and 28 S G2544 (Fig. 3i–k). These two stretches of single-stranded nucleotides have significantly higher icSHAPE reactivity than the surrounding helical regions (Fig. 3l). The spatial distances among the most reactive two nucleotides on each side range from 7.1 to 12.8 Å. Their distances to the SHARC-exo determined 3′ ends range from 3 to 7 nucleotides, consistent with the global average for SHARC-exo (Fig. 3m). Together, these results demonstrate that SHARC-exo can capture static and alternative RNA-RNA interactions at near nucleotide resolution.

In addition to the ribosome, SHARC-exo also captured spatial distances in other non-coding RNAs, including the RPPH1 RNA in RNase P, the 7SL RNA in signal recognition particle (SRP), and U4/U6 snRNAs in the spliceosome (Supplementary Figs. 7–9). RNase P is a ribozyme that cleaves off the 5′ leader of tRNA precursors[40,41]. SHARC-exo captured five proximal nucleotide pairs in the range of 17–36 Å (compare to ~190 Å -- the overall length of RPPH1 structure. Supplementary Fig. 7). In the 7SL RNA, all SHARC-exo measured distances are in the range of 9–26 Å, except one at 77.5 Å, which is likely due to an alternative conformation previously predicted as a precursor in the SRP assembly[42] (Supplementary Fig. 8). U4 and U6 snRNAs form a stable complex in the spliceosome, and two DGs connecting U4 to U6 were detected (Supplementary Fig. 9a)[43]. Crosslinking sites were mapped to two regions in spatial proximity, including a 3-way junction (DG1, Supplementary Fig. 9b,c) and single-stranded regions near an intermolecular helix (DG2, Supplementary Fig. 9b, d). In both structures, exo trimming pinpointed the nucleotides in spatial proximity. Together these results demonstrated that SHARC-exo could measure spatial distances in a wide variety of RNAs in cells.

SHARC-exo distance measurements improve Rosetta-based RNA 3D modeling. Having demonstrated accurate distance measurements by SHARC-exo, next, we investigated whether these constraints can improve 3D structure prediction. For example, we focused on a specific region, h22-h24, in the 18 S rRNA (Fig. 4a, Supplementary Fig. 10a). SHARC-exo captured two major spatially proximal pairs of nucleotides at 7.8 and 21.0 Å (Fig. 4b, Supplementary Fig. 10b). Using these two distances as constraints and a linear pseudo-energy function, we modeled the 3D structure of this 18 S segment (see Methods in Supplementary Information). The addition of the constraints significantly reduced the RMSD distribution for all models and the top 200 models (Fig. 4c, d, Supplementary Fig. 10c). Clustering showed that the SHARC-exo constrained top model displays high topological similarity to the cryo-EM structure, while the de novo model deviates substantially (Fig. 4e, f, Supplementary Fig. 10d). In the native ribosome, the h22-h24 region is stabilized by interactions with other RNA and protein components. Despite using only two constraints in this unfavorable case, the resolution of the 3D model increased significantly. The
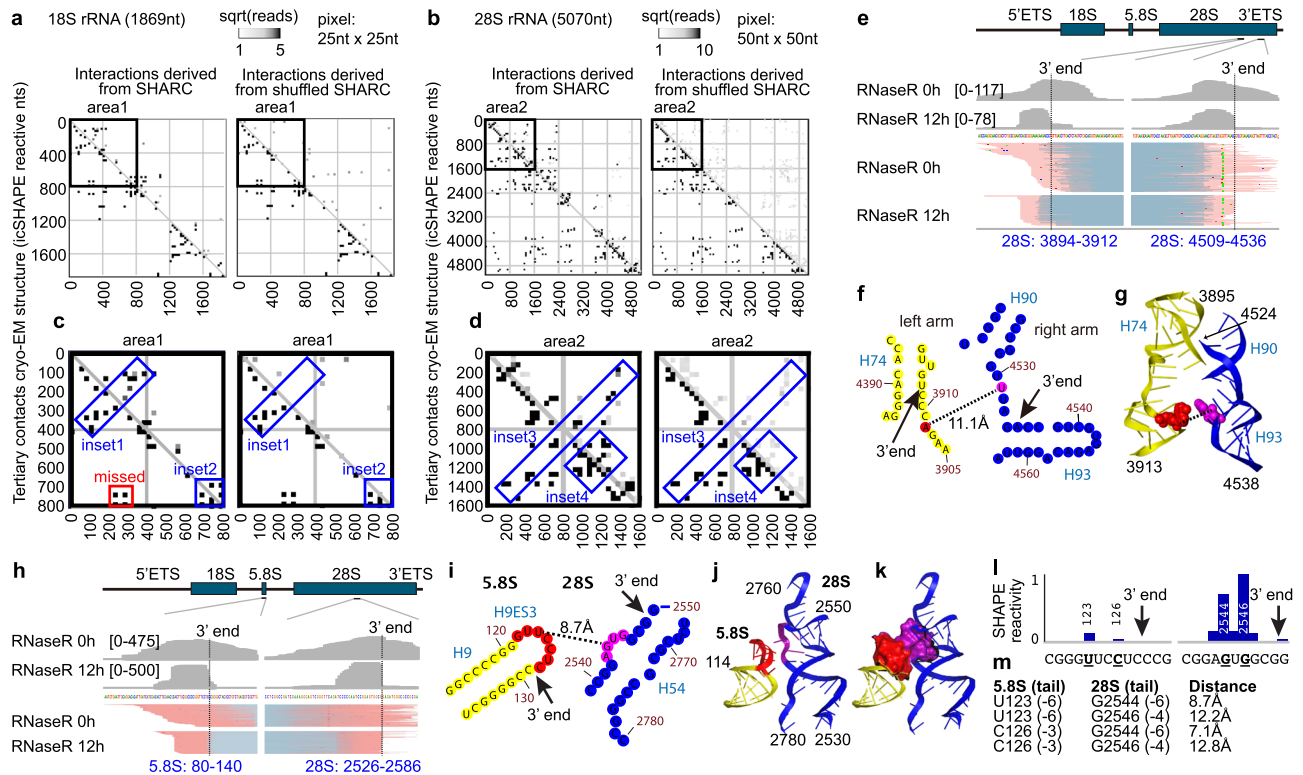
Fig. 3 SHARC-exo captures spatial distances within the human ribosome. a, **b** Comparing SHARC-exo captured spatial proximities to a cryo-EM structure model of the human ribosome in 25 nt × 25 nt (**a**) or 50nt × 50nt (**b**) windows (PDB:4V6X). icSHAPE-measured reactive nucleotides within 20 Å of each other in the 18S (**a**) and 28S (**b**) rRNAs are plotted on the lower-left corner of each square. SHARC-exo gapped reads with 2 arms within 20 Å of each other are plotted on the upper right corner, and the read numbers are square-root scaled. Positions of SHARC-exo reads are randomly shuffled as a control (right panels). **c**, **d** Zoom-in views of two areas in the 18S and 28S rRNAs. Blue rectangles highlight consistent distance measurements between SHARC-exo and cryo-EM. The red rectangle highlight regions missed by SHARC-exo. **e** An example tertiary proximity in the 28 S rRNA captured by SHARC-exo. **f** Secondary structure model of the two regions, showing the consensus 3′ ends of the gapped reads, expected crosslink sites, and distance. **g** 3D structure model of the crosslinked sites. **h**–**m** SHARC-exo captured an interaction between 5.8S and 28S rRNAs. **h** Gapped reads for the interactions between 5.8 S and 28 S rRNA. **i** The 3′ end, putative crosslinking sites and distance mapped onto the secondary structure model of rRNAs. **j**, **k** 3D model of the interaction, where the two loops involved in interactions are shown in red and purple. Interhelical stacking is shown in spheres (**k**). **l**, **m** icSHAPE measurement of nucleotide flexibility around the crosslinking sites (**l**) and all possible distances between the two interacting loops (**m**). **m** Distances between 2′OH groups at the nucleotides with high SHAPE reactivity. Tail length in the parentheses indicates the distance between the 3′ ends and the reactive nucleotides that are potentially crosslinked. Source data are provided as a Source Data file.
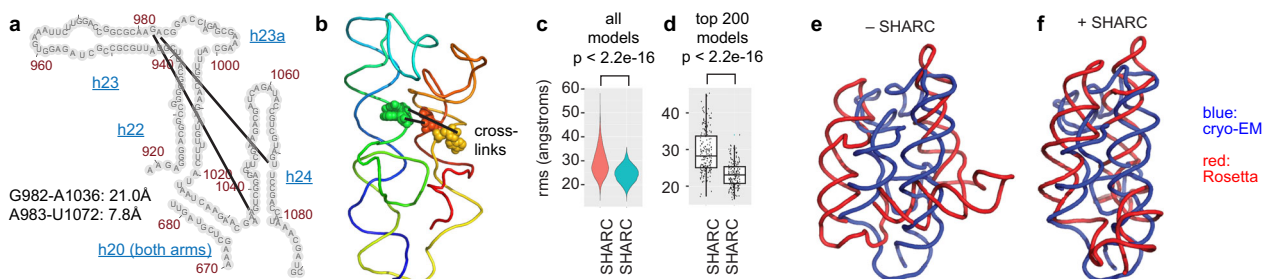
Fig. 4 Spatial distances captured by SHARC-exo improve 3D modeling. a Secondary structure model of h22-h24, showing the two crosslinks and their distances. Only the right side of h20 is included for modeling. **b** 3D model of h22-h24, showing the two crosslinks (PDB: 4V6X). **c**, Violin plots showing the distribution of RMSD values for all models with (n = 12,363) or without (n = 20,394) SHARC-exo constraints. p values shown above were calculated with a two-sided Wilcox rank-sum test. **d**, Top 200 models for each condition are shown as box plots. p Value shown above the plot was calculated with a two-sided Wilcox rank-sum test. For boxplots the median is marked by the solid line in the center of the box, the verticle length of the box represents the interquartile range (IQR) upper fence: 75th percentile + 1.5*IQR, lower fence: 25th percentile − 1.5*IQR, p values for Wilcox rank-sum tests are shown above violin and boxplots. **e** Comparing the top model from the de novo Rosetta run (red) with the cryo-EM model (blue). **f** Same as panel e, except that the Rosetta model was constrained with the two SHARC-exo distances measurements. Source data are provided as a Source Data file.

availability of deeper sequencing coverage and denser constraints will likely further improve the resolution of 3D modeling.

To further benchmark the SHARC-exo method, we performed in vitro crosslinking of a well-characterized model RNA, the P4-P6 domain of the tetrahymena ribozyme[6]. SHARC-exo revealed 17 DGs with highly variable abundance, where the low abundance ones are likely artifacts from RNA misfolding or crosslinking (Supplementary Fig. 10f). We focused on 5 abundant DGs where each contains >3% of total reads. The RNase R trimming shortened the 3′ ends and improved the measurement of spatial distances by 11–23 Å, based on the analysis of the crystal structure[6] (Supplementary Fig. 10g–h). The RNase R refined distances, in the range of 19–41 Å, were still longer than the minimal cross-linkable distances between the two arms (e.g., Supplementary Fig. 10i-j). These results demonstrate that the crosslinking and RNase R trimming successfully captured spatial distances, but there is still space for further improvement. The addition of the SHARC and SHARC-exo derived distances significantly reduced overall RMSD distribution for the complete set of all models and the top 100 (Supplementary Fig. 10k). Models constrained by SHARC-exo distances are much more compact than non-constrained or SHARC-constrained ones (Supplementary Fig. 10l–o). This result further confirmed the usefulness of distance measurements in 3D modeling.

SHARC-exo captures alternative RNA conformations. Many flexible regions in the ribosome, especially the ES, play essential roles in translation[44,45]. However, they are often at low resolution or not resolvable with crystallography or cryo-EM due to their dynamic nature[36]. In SHARC-exo data, reads with two arms that span >40 Å are predominantly located in the ES (96.52%, vs. 3.41% for core tertiary, and 0.08% for dsRNA, Fig. 5a, Supplementary Fig. 11a–c). Consistent with this, the ESs, even if visible by cryo-EM, have a higher B-factor, indicating higher flexibility, while the core segments are considerably better resolved (average resolution 5.4 Å, ranges between 2.2 and 21 Å, Supplementary Fig. 11d, e)[36]. Two ESs in the 28 S rRNA, 78ES30 and 79ES31, have the highest read coverage with between-arm distances >40 Å (hub1 and hub2, Fig. 5b, c). Hub1 makes extensive contacts on many regions on the ribosome, most of which are other flexible ESs (Fig. 5d, Supplementary Fig. 11f). For example, the top six DGs connecting hub1 are all located on the surface of the ribosome, among which the top-ranked is hub2 (Fig. 5d, Supplementary Fig. 11g). The flexibility of both hub1 (78ES30) and its partners make it possible for them to reach each other. Using Rosetta and a single distance constraint between hub1 and hub2, we found that the two regions can be modeled in spatial proximity (from 128 to 16 Å, Fig. 5e, f) without any clashes with other parts of the ribosome surface (Supplementary Fig. 11h, i).

Next, we examined more complex RNA-RNA interactions between the 5.8S and 28S and between 18S and 28S rRNAs (Supplementary Figs. 12 and 13). We discovered both spatial proximal nucleotide pairs and distant ones that likely represent intermediates during ribosome assembly. Two significant regions in the 5.8 S interact extensively with the 28 S (Supplementary Fig. 12a). Among the top 6 DGs connecting 5.8S and 28S, DGs 1, 4, and 5 capture direct contacts, DGs 3 and 6 are likely due to the alternative conformations of ESs on the 28 S that allow the formation of intermolecular contacts, which were not captured by cryo-EM, underlining the power of the SHARC method (Supplementary Fig. 12b)[36]. The remaining DG2 connects two regions that cannot reach each other in the mature ribosome but are supported by extremely high sequencing coverage (Supplementary Fig. 12a, close to DG1). This is likely explained due to

spatial proximity during the assembly of the ribosome. The interactions that we captured between 18S and 28S ESs suggest a highly dynamic nature of the translation machine (Supplementary Fig. 13). Together with the alternative conformations in 7SL (Supplementary Fig. 8), these results suggest that SHARC-exo captures static and dynamic structures in cells.

SHARC-exo reveals compact folding of the 7SK RNA. The noncoding RNA 7SK plays an essential role in transcriptional regulation[46,47]. Still, the structural basis of its function is largely unknown, except for a few small regions that were solved by crystallography and NMR[48–53]. For the full-length 7SK, 331 nt in humans, both secondary and tertiary structures remain uncertain. Wassarman and Steitz proposed the first secondary structure model with four major helices, a "linear model" based on chemical probing (Supplementary Fig. 14a, b)[54]. Deep phylogenetic analysis together with manual adjustments revealed a consistent global secondary structure model across metazoans (Marz model, or "circular model"), featuring eight helical regions, among which a terminal helix (M1) circularizes 7SK[55]. More recent work using the evolutionary coupling method that detects spatial interactions failed to identify the M1 terminal helix (Supplementary Fig. 14c–e)[56]. In vivo icSHAPE[21], a measurement of 1D nucleotide flexibility only provided consistent but not conclusive evidence for the overall validity of helical regions in the Marz model (Supplementary Fig. 14f–g). Here, using SHARC-exo in combination with low-resolution methods PARIS and CLIP, we conclusively demonstrate the existence of the circular model and extensive tertiary contacts within this RNA that suggest compact 3D folding.

Using SHARC-exo, we discovered extensive secondary and tertiary contacts among the helices and single-stranded regions (Fig. 6a, Supplementary Fig. 15). These contacts suggest tight folding of the 7SK RNA in cells. In particular, the two most extended helices, M3 and M7, are packed together (a subset of the contacts shown in Fig. 6b). To validate the compact folding of 7SK, we reanalyzed our recent PARIS and previously published eCLIP data[21,57] (Fig. 6c, d, Supplementary Figs. 16-17). PARIS validated the local structures in the Marz model in both human and mouse cells (M3, M4/M5, and M7), especially the terminal helix M1 (DG1 in Supplementary Fig. 16a–c). In addition, PARIS revealed proximity between distant regions (Supplementary Fig. 16c–e, DGs 2-3). These long-range contacts suggest direct contacts between M3 and M7 since psoralen crosslinking requires stable structures, where at least two base pairs are needed to sandwich a psoralen molecule[58]. In addition to 2-segment (1-gap) reads that represent RNA duplexes, PARIS also captures more complex structures in the form of multi-segment reads, where two structures that form together in one molecule are crosslinked ligated and sequenced (Fig. 6e, Supplementary Fig. 16f)[34,38]. Therefore, multi-segment reads provide direct evidence that three or more segments are close to each other in space in the same RNA molecule. The multi-segment reads connect the 5′ end M3 to the 3′ end M7 and their surrounding sequences. Together, these PARIS data suggest compact folding of the 7SK RNA.

CLIP experiments occasionally crosslink a protein molecule to more than one RNA fragment in spatial proximity[59]. Proximity ligation can join these fragments in one sequencing read (Fig. 6d, Supplementary Fig. 17). We reanalyzed the extensive collection of eCLIP datasets and found that LARP7, an integral component of the 7SK complex, is strongly crosslinked to multiple locations, including the M1, M3, and M7-M8 (Supplementary Fig. 17a,b). The LARP7 eCLIP gapped reads confirmed the local 7SK secondary structures (M3, M6, and M7), the terminal helix M1. They revealed long-range structures that bring M3 and M7 to
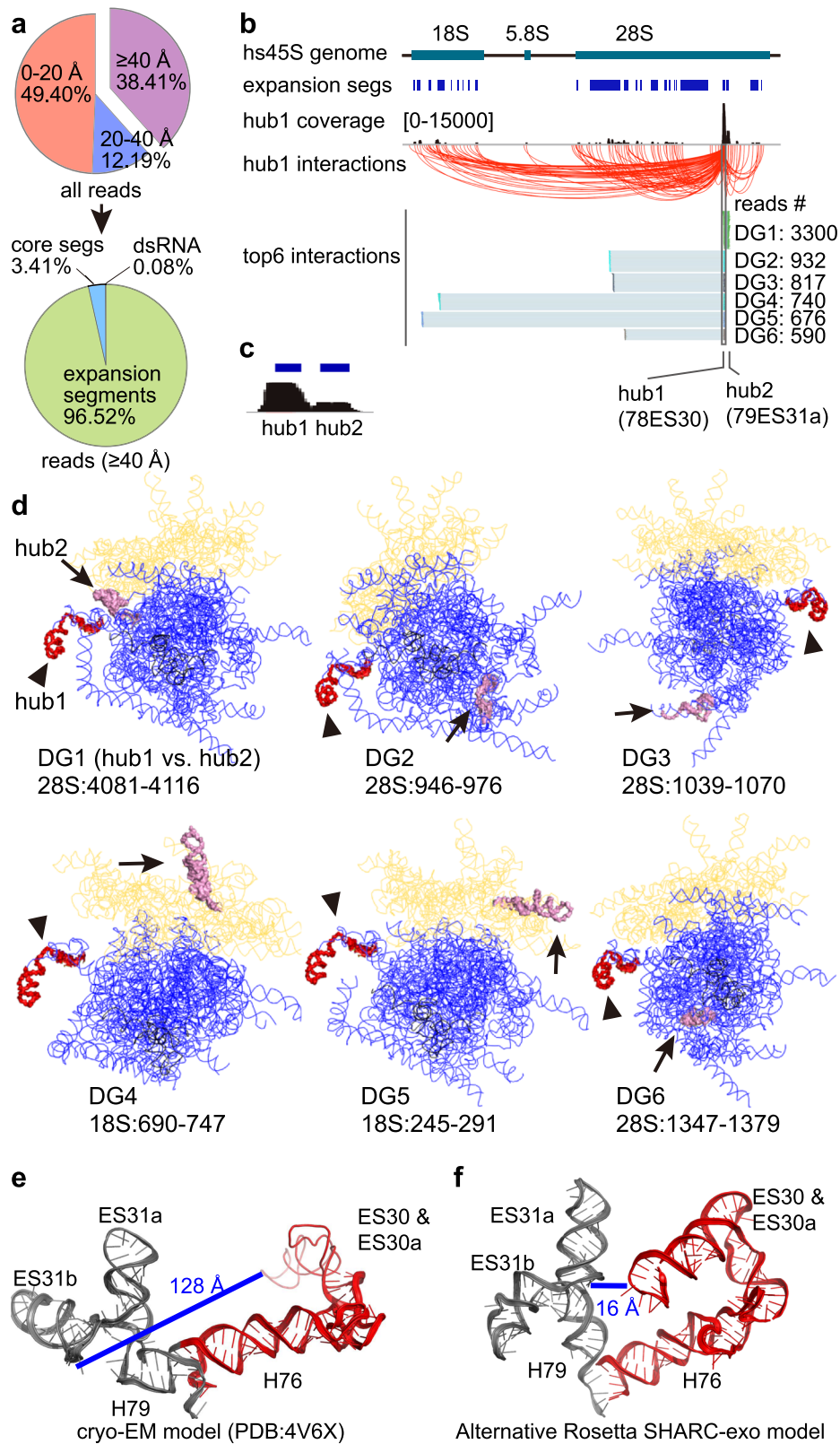
spatial proximity, similar to PARIS (Fig. 6d, Supplementary Fig. 17c,d, DGs 1–3). Together, our integration of 3 orthogonal approaches, SHARC-exo, PARIS, and eCLIP, provided strong support for the circular model (Marz model) of the 7SK secondary structure and suggest a compact folding of the 7SK helices, in particular the direct contacts between M3 and M7 (Fig. 6f–g).

Discussion

This study reports a series of reversible crosslinkers, SHARC, that can capture spatial proximity in RNA with high efficiency. We develop an exo trimming strategy that improved resolution in both SHARC and PARIS to near single nucleotides (~3–7 nucleotides from the 3′ end), therefore generally applicable to various types of crosslinkers. The high throughput SHARC-exo

Fig. 5 SHARC-exo reveals dynamic conformations of expansion segments in the ribosome. a Distribution of reads based on their distances between the two arms. For reads with minimal distance >40 Å, the vast majority are mapped to the expansion segments (lower panel). **b** All hub1 interactions are shown by red arcs. Among top-ranked DGs connecting hub1, the highest abundance expansion segment (78ES30), 4 of them are within the 28S (DGs 1–3 and 6), and 2 of them with the 18S (DGs 4 and 5). **c** Zoom-in view of hub1 and hub2, the two most highly connected dynamic regions in the rRNAs. **d** Locations of the hub1 (red-colored, indicated by arrowheads) and its targets (gray, indicated by arrows). Blueline: 28S rRNA. Yellow line: 18S rRNA. Blackline: 5.8S rRNA. **e, f** Cryo-EM (**e**) and a representative Rosetta (**f**) model of the hub1-hub2 region (28S:3936–4175) in the ribosome. Distances are between the 2′OH groups at nucleotides 4000 and 4123. **f** The Rosetta model was constructed with a single constraint between nucleotides 4000 and 4123. Source data are provided as a Source Data file.
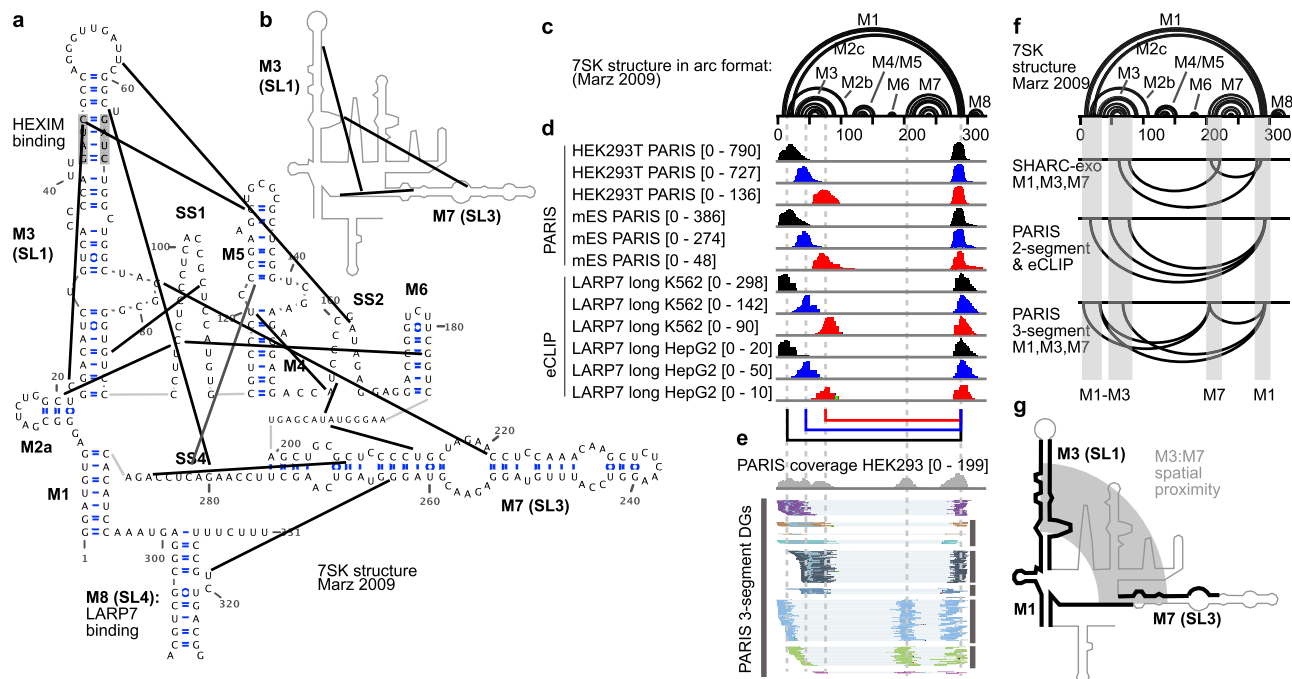
Fig. 6 SHARC-exo, PARIS, and eCLIP reveal compact folding of the 7SK RNA. a Secondary structure model of the human 7SK RNA (Marz 2009, helices M1-M8 and single-stranded regions SS1–4) and 15 SHARC-exo derived spatial constraints (thick black lines). HEXIM and LARP7 binding sites are labeled. **b** SHARC-exo derived tertiary proximities between M3 and M7. **c** Marz 2009 secondary structure model of 7SK in arc format. **d** eCLIP and PARIS captures long-range contacts among M1, M3, and M7. Each track shows the coverage of one DG connecting two regions. **e** PARIS two-gap (3-segment) reads that support long-range contacts in the compact 7SK RNA. The 5 vertical dash lines align the major peaks interacting with each other. **f** Comparison of long-range contacts derived from SHARC-exo, PARIS 2-segment, eCLIP 2-segment, and PARIS 3-segment reads. **g** Model of spatial proximity between M3 and M7 as determined by all 3 methods.

method measures spatial distances between nucleotides either within an RNA or between different RNA molecules in living cells with high efficiency. We show that SHARC-exo distance information can be used to constrain Rosetta-based 3D RNA modeling, therefore opening the possibility of understanding the 3D structures of the entire transcriptome in vivo. Using the ribosome as an example, we demonstrate that SHARC-exo also reveals highly heterogeneous conformations of ESs in cells, challenging to characterize using conventional physical methods. Finally, we integrated SHARC-exo with two other methods, PARIS and CLIP, to conclusively determine a secondary structure model for the 7SK RNA and reveal a compact folding of the multiple helices. These results highlight significant advancements compared to previous methods for RNA 3D structure analysis.

Future improvements and extensions of the SHARC-exo principle will further enhance its versatility and reliability and broaden its applications. First, despite our careful in vitro analysis of the SHARC reagents, the kinetics of the two-step acylation and hydrolysis reactions are difficult to characterize experimentally and theoretically because they are not necessarily decoupled or orthogonal. These reactions are likely different in vitro and in vivo, making it challenging to develop a simplified model to

study them in detail. Nevertheless, a better understanding of these reactions is important for further improvement of crosslinking chemistry. The current RNase R trimming is not 100% efficient, likely due to the presence of SHARC mono adducts on RNA. Even though our computational identification of trimming stop sites uses the 3′ end medians, which is robust against outliers, further improvement of this experimental step will increase the resolution and accuracy. Most cellular RNAs are associated with proteins. Incorporating RNA–protein interactions and protein structure information will enable 3D modeling of RNP complexes in cells. Current acylation-based crosslinkers apply to all four nucleotides yet are limited to flexible ones. In some highly structured RNAs, the number of flexible and, therefore, cross-linkable nucleotides might be moderate (Supplementary Fig. 18). Critical spatially proximal nucleotides may be non-reactive, making it potentially challenging to capture such constraints. In the future, the development of chemical crosslinkers that react with other functional groups in RNA with reduced bias will further improve the efficiency, resolution, and dynamic range of the distance measurements. Current modeling methods that can use experimental constraints, such as Rosetta, are extremely computationally expensive. With the ability to measure spatial

distances in high throughput, new computational tools are urgently needed further to exploit the rich structural information in the SHARC-exo data and enable more rapid 3D modeling for larger RNAs and deconvolution of structural ensembles on a transcriptome-wide scale. Targeted enrichment coupled with SHARC-exo can be applied to many low-abundance RNAs to study their structures[24]. We anticipate that direct high through-put analysis of RNA 3D structures in vivo will reveal new principles of RNA structure formation and function. Given the critical roles of RNA in human genetic and infectious diseases, in vivo 3D structural information is invaluable for developing RNA-based and RNA-targeted therapeutics.

Methods

Synthesis of activated dicarboxylic acids. 1,1′-Oxalyldiimidazole was purchased from Tokyo Chemical Industries. All other activated dicarboxylic acids were synthesized (Supplementary Fig. 1). The dicarboxylic acid (0.20 mmol) was dissolved in 0.1 mL of DMSO. To this was added a solution of CDI (0.40 mmol) in DMSO (0.1 mL) and the resulting mixture was kept under nitrogen at room temperature for 1 h. Heavy bubbling was observed in all cases, which stopped after ~10 min. The resulting 1.0 M solution of activated dicarboxylic acids was used immediately in crosslinking experiments, without further purification. Successful activation was confirmed for all compounds and full analysis (^{1}H NMR, ^{13}C NMR, and MS) was obtained. Note that imidazole is formed as a byproduct in the reaction and is present in all spectra, which can be found in the Supplementary Information.

In vitro crosslinking of model RNA. The model RNA **1** was purchased from Integrated DNA Technologies. Nine μl of 10 μM RNA **1** in 0.06 M MOPS, pH 7.5; 0.1 M KCl; 2.5 mM MgCl₂, was heated to 95 °C for 2 min and then slowly cooled to room temperature. One μL of 1 M activated dicarboxylic acid stock solution in DMSO was added and the mixture was incubated for 4 h at room temperature. Reactions were quenched by the addition of 9 volumes of precipitation solution (0.33 M NaOAc, pH 5.2, glycogen 0.2 mg/mL) and 30 volumes of absolute ethanol. RNA was precipitated for 1 h at −20 °C and then centrifuged (21,000 × g) for 40 min at 4 °C. The pellet was washed with 70% ethanol, air-dried, and resuspended in 10 μL RNase-free water. Precipitated RNA was analyzed using 20% PAGE and imaged using Sybr Gold and a Bio-Rad Gel Documentation System (Image Lab software, v6.0.1) and safeVIEW-MINI2 Imaging System. The distribution between unreacted RNA **1** and crosslinked RNA was determined by quantifying the band intensity with ImageJ (V1.52t). All experiments were performed in triplicate.

Reversal of in vitro crosslinked RNA. Five microlitres of 10 μM crosslinked RNA in water were diluted with 45 μL 100 mM borate buffer pH 10.0 and incubated for 2 h at 37 °C. Reactions were quenched by addition of 50 μL of precipitation solution (0.33 M NaOAc, pH 5.2, glycogen 0.2 mg/mL) and 300 μL of absolute ethanol. RNA was precipitated for 1 h at −20 °C and then centrifuged (21,000 × g) for 40 min at 4 °C. The pellet was washed with 70% ethanol, air-dried, and resuspended in 10 μL RNase-free water. RNA was analyzed using 20% PAGE and imaged using Sybr Gold and a Bio-Rad Gel Documentation System and safeVIEW-MINI2 Imaging System. The distribution between unreacted RNA 1and cross-linked RNA was determined by quantifying the band intensity with ImageJ (V1.52t). All experiments were performed in triplicate.

Cell culture. HeLa (CCL-2) and HEK293T (CRL-3216) cells were purchased from ATCC and maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco) + 10% fetal bovine serum (FBS, Gibco) + Pen/Strep antibiotic, in 37 °C incubators with 5% CO₂. All cell cultures were handled according to protocols approved by the University of Southern California.

SHARC crosslinker preparation for crosslinking. SHARC reagents were made by dissolving 1-part SHARC reactant in 200 μl anhydrous DMSO (Sigma, 276855) and 2 parts CDI Sigma, 115,533) in 250 μl DMSO. Dissolved SHARC reactant was pipetted into the tube containing CDI. After briefly vortex and spinning down, a needle was inserted into the top of the 1.5 mL centrifuge tube to allow the CO₂ product to escape. Mixed solutions were left at room temperature to react for 30–60 min before crosslinking.

In vivo crosslinking. Hela and HEK293T cells with 80% confluency in a 10 cm dish were washed twice with 1× phosphate-buffered saline (PBS). Then cells were collected, resuspended in 1× PBS, and transferred into a 1.5 ml tube with a final volume of 900 μl. For each tube of cells, added 100 μl of SHARC crosslinker to make the final concentration of 0, 5, 12.5, and 25 mM. Cells were incubated in a rotator at room temperature for 30 min. After crosslinking, crosslinking solution was removed and cells were washed twice with 1× PBS.

Extraction of crosslinked RNA (TNA method, adapted from[1]). For each 10 cm dish cell, added 100 μl of 6 M GuSCN (Sigma, 368975) and lysed cells with vigorous manual shaking for 1 min. Then, cell lysate was added 12 μl of 500 mM EDTA (Invitrogen™, 15575020), 60 μl of 10× PBS (Invitrogen™, AM9625), and water to a final volume of 600 μl. Each sample was passed through a 25 or 26 G needle about 20 times to further break the insoluble material. Proteinase K (PK) (Thermo Scientific™, EO0492) was added to a final concentration of 1 mg/ml, and PK treatment was performed at 37 °C for 1 h on a shaker at 1000–1200 × g. After PK digestion, 60 μl of 3 M sodium acetate (pH 5.3) (Invitrogen™, AM9740), 600 μl of water-saturated phenol (pH 6.6) (Invitrogen™, AM9712), and 1 volume pure isopropanol were added to precipitate total nucleic acids by spinning at 17,000 × g for 20 min at 4 °C. After twice washing using 70% ethanol, total nucleic acids were resuspended in 300 μl of nuclease-free water. For 100 μg of TNA samples, 50 units of TURBO™ DNase (Invitrogen™, AM2239) were added to remove DNA at 37 °C for 20 min. Then added 20 μl of 3 M sodium acetate, an equal volume of water-saturated phenol, two-volume of pure isopropanol to precipitate RNA sample by spinning 20 min at 12,000 × g at 4 °C.

RNA fragmentation. A 10 μg of cross-linked RNA was fragmented using 10 μl of RNase III (NEB, M0245) with 5mM MnCl₂ and 1× supplied shortcut buffer at 37 °C for 5 mins. After incubation, an equal volume of phenol was immediately added to stop the reaction. Then the one-tenth volume of 3 M sodium acetate (pH 5.3), 3 μl of GlycoBlue (Invitrogen™, AM9516), three-volume of pure ethanol were added to precipitate RNA. Fragmented RNA was resuspended in RNase-free water.

DD2D purification of cross-linked RNA. *First dimension gel*. Prepare 8% 1.5 mm thick denatured first dimension gel using the UreaGel system (National Diagnostics, EC-833) with MOPS buffer (Fisher, BP2900500). Briefly, 3.2 ml UreaGel concentrate, 5.8 ml UreaGel diluent, 1 ml 10× MOPS buffer, 80 μl 10% of APS, and 4 μl TEMED (Thermo Scientific™, 17919) were mixed to make 8% first dimension gel. Loading dsRNA ladder (NEB, N0363S) as molecular weight marker. Run the first dimension gel at 30 W for 7–8 min in 1× MOPS buffer. After electrophoresis was finished, staining the gel with SYBR Gold (Invitrogen™, S11494) in 1× MOPS buffer and excising each lane between 50 nt to topside from the first dimension gel. The second dimension gel can usually accommodate three gel splices.

Second dimension gel. Prepare the 16% 1.5 mm thick urea denatured second dimension gel using the UreaGel system with MOPS buffer. Briefly, 6.4 ml UreaGel concentrate, 2.6 ml UreaGel diluent, 1 ml 10× MOPS buffer, 80 μl 10% of APS, and 4 μl TEMED were mixed to make 16% first dimension gel. Using prewarmed 1× MOPS buffer to fill the electrophoresis chamber to facilitate denaturation of the cross-linked RNA. Run the second dimension at 30 W for 50 min to maintain high temperature and promote denaturation. Gels were imaged using the iBright FL1500 Image System (iBright Analysis Software, v3.1.2). A gel containing the cross-linked RNA above the diagonal from the 2D gel was excised and crushed for RNA extraction.

RNase R treatment. RNase R is a 3′ → 5′ exonuclease that is capable of unwinding and digesting double-stranded RNA with a 3′ overhang. Purified crosslinked RNAs from DD2D gel were treated with 20 units of RNase R (Biovision, M1228) in 1× RNase R digestion buffer with 5 mM ATP at 45 °C for 2, 12, and 24 h, respectively. Control RNA was without RNase R treatment. After RNase R treatment, the one-tenth volume of 3 M sodium acetate (pH 5.3), 3 μl of GlycoBlue, three-volume of pure ethanol were added to precipitate RNA.

Proximity ligation. Purified RNA fragments were proximity ligated by T4 RNA Ligase1 (NEB, M0437M). Briefly, 2 μl of 10× ligation buffer, 5 μl of T4 RNA Ligase, 1 μl of SuperaseIn (Invitrogen™, AM2696) and 1 μl of 0.1 mM ATP were added to 10 μl of purified dsRNA fragments[2]. Ligation mixture was incubated at room temperature overnight. After ligation, the samples were boiled for 2 min to stop the reaction. After heat denaturation, samples were centrifuged to remove the precipitate and then precipitated by ethanol.

Reverse crosslinking. To proximity ligated RNA fragments, 5× decrosslinking buffer (500 mM Boric acid, pH 11) was added, and nuclease-free water was added to bring decrosslinking buffer to 1×. Samples were incubated for 2 h at 45 °C to guarantee reversal (this is higher than the temperatures used in the in vitro experiments). After reverse crosslinking, RNA was purified with three-volume of ethanol and 1 μl of GlycoBlue.

Adapter ligation. Reverse crosslinked RNAs were heated at 80 °C for 90 s, then snapped cooling on ice. To each sample, 3 μl of 10 μM ddc adapter /5rApp/ AGATCGGAAGAGCGGTTCAG/3ddC/, 1 μl of T4 RNA ligase 1, 2 μl of DMSO, 5 μl of PEG8000, 1 μl of 0.1 M DTT, 1 μl of SuperaseIn and 2 μl of 10x T4 RNA ligase buffer were added to perform adapter libation at room temperature for 3 h. After adapter ligation, the following reagents were added to remove free adapters: 3 μl of 10x RecJf buffer (NEBuffer 2, B7002S), 2 μl of RecJf (NEB, M0264S), 1 μl of 5′Deadenylase (NEB, M0331S), 1 μl of SuperaseIn, Reaction was incubated at 37 °C for 1 h. Then 20 μl of water was added to each sample to make a total volume of

50 µl and Zymo RNA clean and Concentrator-5 (Zymo Research, R1013) was used to purify RNA.

Reverse transcription. SuperScript IV (SSIV) (Invitrogen™, 18090010) was used to perform reverse transcription. The reaction buffer was optimized Mn^{2+} buffer (1×): 50 mM Tris-HCl (PH 8.3), 75 mM CH_3COOK, and 1.5 mM $MnCl_2$. Briefly, 1 pmol of barcoded RT primer and 1 µl of 10 mM dNTP were added to RNA samples and heated at 65 °C for 5 min in a PCR block, chilling the samples on ice rapidly. Then 4 µl of 5× Mn^{2+} buffer, 2 µl of 0.1 M DTT, 1 µl of SuperaseIn and 1 µl of SSIV were added to each sample. The mixed sample was incubated at 25 °C for 15 min, 42 °C for 10 h, 80 °C for 10 min; hold at 10 °C. After reverse transcription, 1 µl RNase H and RNase A/T1 mix were added and incubated at 37 °C for 30 min in a thermomixer to remove RNA. Synthesized cDNA was purified using Zymo DNA clean and Concentrator-5.

cDNA circularization and library generation. 1 µl of CircLigase™ II ssDNA Ligase (Lucigen, CL9021K), 1 µl of 50 mM $MnCl_2$ and 10× CircLigaseII™ buffer were added to cDNA sample and performed circularization at 60 °C for 100 min. An 80 °C treatment for 10 min was followed to stop the reaction. The circularized cDNA products were directly used to library PCR. Library PCR preparation was performed[3]. PCR products were run on 6% native TBE gel. A gel containing DNA products from 175 bp and topside (corresponding to >40 bp insert) was excised and crushed for DNA extraction.

In vitro SHARC-exo analysis of the P4-P6 RNA. The P4–P6 (PDB: 1HR2) DNA with T7 promoter (TAATACGACTCACTATAG) was purchased from twist bioscience. After PCR amplification, the DNA was cleaned up using the Qiagen PCR Purification Kit and purified using an 8% native polyacrylamide gel. The P4-P6 (1HR2) RNA was transcribed using the MEGAscript T7 Transcription Kit from Thermo Fisher (AM1334) from 136 ng of DNA template and purified on denatured polyacrylamide gels. 10 µg of P4–P6 RNA, 10 µL of refolding buffer, and water was added to a final volume of 44 µL per sample. The RNA was then denatured by incubating at 90 °C for 5 min followed by snap cooling on ice. 1 µL of 500 mM $MgCl_2$ was then added to each sample while cold and then mixed. Samples were then allowed to come to room temperature over several minutes to refold. After refolding, either 5 µL of DMSO for controls or 5 µL 50 mM DPI was added to each sample. Samples were then incubated at room temperature for 30 min. After incubation, samples were purified using ethanol precipitation. The crosslinked RNA was then converted into cDNA libraries as described above. In particular, we divided the crosslinked RNA fragments from the DD2D gels into 2 fractions, where one was treated with RNase R at 37 °C for 2 h, while the other was not treated. The cDNA library was sequenced on a MiSeq machine.

SHARC-Seq analysis

Mapping. BCL files were converted to fastq files using bcl2fastq2 Conversion Software (v2.20.0). The 3′end adapters of sequencing data were removed using Trimmomatic (v0.36). PCR duplicates were removed using readCollapse script from the icSHAPE pipeline. After removing 5′ header, reads were mapped to manually curated hg38 genome using STAR (v2.7.0 f) program[4]. The parameters used are as follows: STAR --runThreadN 8 --runMode alignReads --genomeDir OuputPath --readFilesIn SampleFastq --outFileNamePrefix Outprefix --genomeLoad NoSharedMemory outReadsUnmapped Fastx --outFilterMultimapNmax 10 --outFilterScoreMinOverLread 0 --outSAMattributes All --outSAMtype BAM Unsorted SortedByCoordinate --alignIntronMin 1 --scoreGap 0 --scoreGapNoncan 0 --scoreGapGCAG 0 --scoreGapATAC 0 --scoreGenomicLengthLog2scale -1 --chimOutType WithinBAM HardClip --chimSegmentMin 5 --chimJunctionOverhangMin 5 --chimScoreJunctionNonGTAG 0 --chimScoreDropMax 80 --chimNonchimScoreDropMin 20.

Classify alignments. The primary mapping alignments were extracted from SampleAligned.sortedByCoord.out.bam using SAMtools (v1.8), and classified into six different types using gaptypes.py (https://github.com/zhipenglu/CRSSANT)[5]. cont.sam, continuous alignments; gap1.sam, non-continuous alignments with one gap; gapm.sam, non-continuous alignments with more than one gaps; trans.sam, non-continuous alignments with the two arms on different strands or chromosomes; homo.sam, non-continuous alignments with the two arms overlapping each other; bad.sam, non-continuous alignments with complex combinations of indels and gaps. Gap1. and gapm alignments containing splicing junctions and short 1–2 nt gaps were filtered out using gapfilter.py (https://github.com/zhipenglu/CRSSANT). Then filtered gap1.sam, filtered gapm.sam and trans.sam were used to analyze RNA structures and interactions.

Cluster alignments to groups. Filtering alignments were assembled to DGs and NGs using the crssant.py script (https://github.com/zhipenglu/CRSSANT). After DG clustering, crssant.py verifies that the DGs do not contain any non-overlapping reads, i.e., any reads where the start position of its left arm is greater than or equal to the stop position of the right arm of any other read in the DG. If the DGs do not contain any non-overlapping reads, then the following output files ending in the following are written: Sample.sam: SAM file containing alignments that were

successfully assigned to DGs, plus DG and NG annotations; dg.bedpe: bedpe file listing all duplex groups.

Visualization of SHARC-seq data in Integrative Genomics Viewer. Assembled alignments with DGs tag were displayed using integrative Genomic Viewer (IGV)[6] visualization tool (V.2.8.13). The bed output file (from crssant.py script) can be visualized in IGV, where the two arms of each DG can be visualized as two "exons", or as an arc that connects far ends of the DG.

Structure analysis of rRNAs. To analyze the RNase R trimming efficiency (e.g., Fig. 2f), we examined gapped alignments against the ribosome cryo-EM structure. For each alignment, we calculated the minimum physical distance between the two arms in the ribosome. Then the nucleotides involved in the minimal distances were recorded (counting from the 3′ end of each arm). In a hypothetical example, we found that the minimal distance between the two arms in one read was between the 10th nt from the right arm and the 15th nt from the left arm (both counting from the 3′ ends). Then this tuple (10, 15) is considered one point on the heatmap (Fig. 2f and Supplementary Fig. 5b). After all the minimal distance nucleotides are calculated, their frequencies are plotted in the heatmap in square root scale.

The SHARC-seq reads aligned to 45S pre-rRNA (NR_046235.3) were collected and used to construct the interaction matrix. To build the physical interaction map of 28S rRNA and 18S rRNA, the cryo-EM model of the 28S rRNA and 18S rRNA was downloaded from RCSB Protein Data Bank (PDB) (ID: 4V6X). Watson-Crick and non-Watson-Crick base pairs were analyzed using the DSSR software (v1.7.7)[7]. The 3D structures of the ribosome were visualized by the PyMOL system (Educational version, https://pymol.org/2/). Spatial distances in the cryo-EM model were extracted directly for use. The resolution of the human ribosome cryo-EM model is highly variable across the entire complex (PDB: 4V6X, Supplementary Fig. 2c from[8]). Although the average resolution is 5.4 Å, the lowest goes to 21 Å. The ribosome structure analysis and conclusions are based on longer distance intervals, e.g., 0–20 and 20–40 Å. The modeling runs also used 0–20 and >20 Å as the intervals for penalty calculations. In addition, the low-resolution regions are confined to the expansion segments and do not affect the analysis of the stable core regions (as we showed in the B-factor plot, Supplemental Fig. 11d, e). In our analysis of the expansion segments, the distances that we focused on are much longer than 20 A (Fig. 5e, f, Supplementary Fig. 11h, i). Therefore, the limited resolution of the ribosome model does not affect the analysis.

Structure analysis of representative RNAs. In order to accurately and easily analyze SHARC-seq data, pseudogenes and multicopy genes from gencode, refGene, and Dfam were masked from hg38 genome. And then a single copy of them was added back as a separated "chromosome". For example, multicopy of snRNAs were masked from the hg38 assembly genome, and 9 snRNAs (U1, U2, U4, U5, U6, U11, U12, U4atac, and U6atac) were concatenated into one reference, separated by 100 nt "N"s, was added back. The curated hg38 genome contained 25 reference sequences, or "chromosomes", masked the multicopy genes, and added back single copies. This reference is best suited for the PARIS analysis. SHARC-seq reads were mapped to representative RNAs were collected and used for IGV visualization.

Cross-linking distance analysis. The ribose 2′OH in every flexible nucleotide (single-stranded or icSHAPE activated) was used to calculate the cross-linking distance. The minimum distance between two arms' flexible nucleotide was used to analyze the minimum distance distribution. The distance between No.3 to No.7 flexible nucleotides from the 3′end of each arm was used for 3–7 distance distribution analysis.

rRNA dynamic structure analysis. The core and expansion segment boundaries of rRNA were derived from Chandramouli et al. (2008)[9] and Wakeman and Maden (1989)[10]. The SHARC-seq reads with ≥40 Å between two arms were collected and separated to core and expansion alignments. The dynamic reads were selected based on the rules that one arm mapped to the same region of rRNA, other arm mapped to different regions. The selected dynamic alignments were loaded to IGV for visualization.

Computational Modeling of h22–24 region in the 18S. Rosetta software (version 2020.08.61146) was used to model RNAs for this study[11,12]. Helices of secondary structure regions were pre-built with the example command below to save computational expense: rna_helix.py -seq cag cug -resum 5–7 27–29 -o example_helix_1.pdb. The 920–1080 nucleotide region of the human 18S RNA was modeled with and without SHARC determined constraints. For the model set without SHARC constraints, no cst file or flag was used. For the model set with SHARC constraints, the following linear energy function example command was used to assign constraints for 2′OH atoms that participated in the crosslinking reaction so that between 0A–20A there is no energy penalty and to apply a linearly scaling energy penalty if the atoms are >20 A apart.

AtomPair O2′ 63 O2′ 117 LINEAR_PENALTY 10.0 0 10 1.0. Here 10.0 is the ideal distance between the atoms in Å, 0 is the energy penalty assigned to the range, 10 is the tolerance for the energy trough and 1.0 is the slope constraint.

Models were built with the command shown below using fasta, secondary structure and constraint files (for modeling set containing the SHARC data, otherwise no constraint file). For the native file to get a rms for these files, the 920–1080 region of the 18S RNA was cut out using pymol and renumbered using renumber_pdb_in_place.py.

rna_denovo.static.linuxgccrelease -nstruct 1000 -fasta../18s_920_1080.fasta -s../18s_920_1080_helix_1.pdb../18s_920_1080_helix_2.pdb-secstruct_file../18s_920_1080.secstruct -cst_file../18s_920_1080.cst -native../18s_920_1080_renumbered.pdb -minimize_rna true -out:file:silent 18s_920_1080_tert.out

After modeling runs were finished, models were extracted using easy_cat.py. Example: easy_cat.py directory. To extract the top 1% scoring models from the bulk of the models for each run condition the following command was used (in this example 200 models are extracted): silent_file_sort_and_select.py [example_file.out] -select 1–200 -o [example_file_cluster.out]. The lowest 1% energy models were then clustered from each run condition to inspect the different pose topologies that existed within the lowest energy scoring models. Clustering was done with the command shown below (in this example 10 clusters are made using a cluster radius of 5 Å: rna_cluster.static.linuxgccrelease -in:file:silent [example_file.out] -out:nstruct 10 -cluster:radius 5 -out:file:silent [example_file_cluster.out]. Clusters were extracted with the following command. The -no_replace_names flag here is used to prevent clusters from being renamed: python extract_lowscore_decoys.py [example_cluster_file.out] -no_replace_names.

Computational 3D modeling of the P4-P6 RNA. The P4-P6 secondary structure was determined from PDB:1HR2 and is as follows:….(((((.(….(((((((….(((.((((((((..(((((((((((….)))))))))…………….)))….).)))).)))…))))))…).)))))((…(((((.(((((((((…..))))))..)))))…))..Models were generated using the following sample command line: rna_denovo.static.linuxgccrelease -nstruct 1000 -fasta../1hr2.fasta -secstruct_file../1hr2.secstruct -s../1hr2_helix* -cst_file../1hr2_trim_tert.cst -native../1hr2_chain_a_native.pdb -out:file:silent 1hr2_20A.out -minimize_rna true. Models generated without constraints had the –cst_file flag and cst file omitted from the command. The top 5 DGs by number of reads in SHARC-exo data, each with >3% of total reads, were used to constrain modeling with the equivalent DG being used for SHARC-constrained models. Linear atom pair constraint was set so that distances within 20 Å carried no penalty and distances greater then 20 Å were penalized with a slope of 1. RMSD values for models against the 1HR2 crystal structure was calculated. The top 100 scoring models from each group were clustered into 5 groups with a cluster radius of 5 Å. Wilcox Ranked Sum Tests were performed between each two groups of top 100 models with the following R command wilcox.test (rmsd ~ group, data = XXX, exact = FALSE, alternative = 'greater').

Analysis of hub1-hub2 alternative conformations. First, the minimal 28 S segment that contained both regions are limited to residues 3935 – 4175. Secondary structure was determined by running x3dna on the extracted segment[7]: x3dna-dssr -i=input.pdb -o=dssr.out. As the structure did not contain bases for all models so we went by hand to assign addition bases pairs base on geometry. Secondary structure is as follows: ((..(((((((((.(((((.((((.………..(((((…((((.………(.((…………..……….)).)..…………..))))…………).)))))).)))).)).).)))))))…((((((…..((((..((.(((((((….)))))))………………((((((((((…..)))))))))…)))).)))))…..)).

We generated 15317 models using FARFAR by command: rna_denovo.default.macosclangrelease -s stem_1.pdb stem_2.pdb helix_1.pdb helix_2.pdb -nstruct 1000 -fasta test.fasta -secstruct_file test.secstruct -minimize_rna false -cst_file test.cst, where the pdbs contained the original static structures of helices and stems not included in the contact. Linear atom pair constraint was set such that distances within 20 Å carry no penalty, while distance above 20 Å is penalized at a slope of 1. For each of the models, we checked for steric clashes of the rest of the RNA and local proteins by comparing the distance between each phosphate of each modeled RNA nucleotide to the phosphate of each remaining RNA and the c-alpha of an atom of each amino acid. A clash was defined as a distance of fewer than 5 Å.

Analysis of 7SK structures using PARIS. PARIS data from human and mouse cells were used to generate DGs for 7SK[13]. To analyze the secondary structures of 7SK, we clustered HEK293T PARIS non-continuous alignments on 7SK using CRSSANT[3].

Analysis of 7SK structures using LARP7 eCLIP. LARP7 eCLIP data in HepG2 and K562 cells were downloaded from ENCODE[14] and analyzed as follows. First reads mapped to 7SK were extracted from the mapped bam files (chr6:52995620–52995951 in hg38 coordinates). Reads with CIGAR gap flags D and N are extracted. All reads with D flags are converted to N for consistency. Then all reads with "N" were divided into three groups based on read start using the script readspan7SK.py and short-span reads were used to construct local structures.

References
1. Cech, T. R & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
2. Lu, Z. & Chang, H. Y. Decoding the RNA structurome. *Curr. Opin. Struct. Biol.* **36**, 142–148 (2016).
3. Byeon, G. W. et al. Functional and structural basis of extreme conservation in vertebrate 5′ untranslated regions. *Nat. Genet.* **53**, 729–741 (2021).
4. Wu, X. & Bartel, D. P. Widespread influence of 3′-end structures on mammalian mRNA processing and stability. *Cell* **169**, 905–917 e911 (2017).
5. Shabalina, S. A., Ogurtsov, A. Y. & Spiridonov, N. A. A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34**, 2428–2437 (2006).
6. Cate, J. H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
7. Lu, Z. et al. Structural modularity of the XIST ribonucleoprotein complex. *Nat. Commun.* **11**, 6163 (2020).
8. Miao, Z. & Westhof, E. RNA structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **46**, 483–503 (2017).
9. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
10. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505**, 701–705 (2014).
11. Ding, Y. et al. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505**, 696–700 (2014).
12. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
13. Velema, W. A. & Kool, E. T. The chemistry and applications of RNA 2′-OH acylation. *Nat. Rev. Chem.* **4**, 22–37 (2020).
14. Tian, S. & Das, R. RNA structure through multidimensional chemical mapping. *Q. Rev. Biophys.* **49**, e7 (2016).
15. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* **3**, 954–962 (2011).
16. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).
17. Homan, P. J. et al. Single-molecule correlated chemical probing of RNA. *Proc. Natl Acad. Sci. USA* **111**, 13858–13863 (2014).
18. Cheng, C. Y. et al. Consistent global structures of complex RNA states through multidimensional chemical mapping. *Elife* **4**, e07600 (2015).
19. Lu, Z. & Chang, H. Y. The RNA base-pairing problem and base-pairing solutions. *Cold Spring Harb. Perspect. Biol.* https://doi.org/10.1101/cshperspect.a034926 (2018).
20. Aw, J. G. et al. In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* **62**, 603–617 (2016).
21. Lu, Z. et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267–1279 (2016).
22. Ziv, O. et al. COMRADES determines in vivo RNA structures and interactions. *Nat. Methods* **15**, 785–788 (2018).
23. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global mapping of human RNA-RNA interactions. *Mol. Cell* **62**, 618–626 (2016).
24. Zhang, M. et al. Optimized photochemistry enables efficient analysis of dynamic RNA structuromes and interactomes in genetic and infectious diseases. *Nat. Commun.* **12**, 2344 (2021).

25. Velema, W. A., Park, H. S., Kadina, A., Orbai, L. & Kool, E. T. Trapping transient RNA complexes by chemically reversible acylation. *Angew. Chem. Int. Ed. Engl.* https://doi.org/10.1002/anie.202010861 (2020).

26. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).

27. Cruz, J. A. & Westhof, E. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat. Methods* **8**, 513–521 (2011).

28. Theis, C. et al. RNA 3D modules in genome-wide predictions of RNA 2D structure. *PLoS ONE* **10**, e0139900 (2015).

29. Miao, Z. et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* **26**, 982–995 (2020).

30. Park, H. S., Kietrys, A. M. & Kool, E. T. Simple alkanoyl acylating agents for reversible RNA functionalization and control. *Chem. Commun.* **55**, 5135–5138 (2019).

31. Spitale, R. C. et al. RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9**, 18–20 (2013).

32. Lu, Z., Gong, J. & Zhang, Q. C. PARIS: psoralen analysis of RNA interactions and structures with high throughput and resolution. *Methods Mol. Biol.* **1649**, 59–84 (2018).

33. Vincent, H. A. & Deutscher, M. P. Substrate recognition and catalysis by the exoribonuclease RNase R. *J. Biol. Chem.* **281**, 29769–29775 (2006).

34. Fischer-Hwang, I., Lu, Z., Zou, J. & Weissman, T. Cross-linked RNA secondary structure analysis using network techniques. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/668491v1 (2019).

35. Petrotchenko, E. V. et al. Use of proteinase K nonspecific digestion for selective and comprehensive identification of interpeptide cross-links: application to prion proteins. *Mol. Cell Proteom.* **11**, M111.013524 (2012).

36. Anger, A. M. et al. Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80–85 (2013).

37. Petrov, A. S. et al. Secondary structures of rRNAs from all three domains of life. *PLoS ONE* **9**, e88222 (2014).

38. Zhang, M. et al. Classification and clustering of RNA crosslink-ligation data reveal complex structures and homodimers. Preprint at *bioRxiv* https://doi.org/10.1101/2021.08.01.454689 (2021).

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

40. Wu, J. et al. Cryo-EM structure of the human ribonuclease P holoenzyme. *Cell* **175**, 1393–1404 e1311 (2018).

41. Pitulle, C., Garcia-Paris, M., Zamudio, K. R. & Pace, N. R. Comparative structure analysis of vertebrate ribonuclease P RNA. *Nucleic Acids Res.* **26**, 3333–3339 (1998).

42. Kuglstatter, A., Oubridge, C. & Nagai, K. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nat. Struct. Biol.* **9**, 740–744 (2002).

43. Charenton, C., Wilkinson, M. E. & Nagai, K. Mechanism of 5′ splice site transfer for human spliceosome activation. *Science* **364**, 362–367 (2019).

44. Leppek, K. et al. Gene- and species-specific Hox mRNA translation by ribosome expansion segments. *Mol. Cell* **80**, 980–995.e913 (2020).

45. Wild, K. et al. MetAP-like Ebp1 occupies the human ribosomal tunnel exit and recruits flexible rRNA expansion segments. *Nat. Commun.* **11**, 776 (2020).

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

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

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

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

50. Bourbigot, S. et al. Solution structure of the 5′-terminal hairpin of the 7SK small nuclear. *RNA* **22**, 1844–1858 (2016).

51. Durney, M. A. & D'Souza, V. M. Preformed protein-binding motifs in 7SK snRNA: structural and thermodynamic comparisons with retroviral TAR. *J. Mol. Biol.* **404**, 555–567 (2010).

52. Martinez-Zapien, D. et al. The crystal structure of the 5′ functional domain of the transcription riboregulator 7SK. *Nucleic Acids Res.* **45**, 3568–3579 (2017).

53. Pham, V. V. et al. HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat. Commun.* **9**, 4266 (2018).

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

55. Marz, M. et al. Evolution of 7SK RNA and its protein partners in metazoa. *Mol. Biol. Evol.* **26**, 2821–2830 (2009).

56. Weinreb, C. et al. 3D RNA and functional interactions from evolutionary couplings. *Cell* **165**, 963–975 (2016).

57. Van Nostrand, E. L. et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* **13**, 508–514 (2016).

58. Cimino, G. D., Gamper, H. B., Isaacs, S. T. & Hearst, J. E. Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annu. Rev. Biochem.* **54**, 1151–1193 (1985).

59. Sugimoto, Y. et al. hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1. *Nature* **519**, 491–494 (2015).

Acknowledgements

Author contributions

Z.L. and W.A.V. conceived the project. W.A.V. performed the SHARC synthesis and crosslink characterizations. R.V.D. and Z.L. developed the SHARC-exo strategy. K.L. and J.B. performed the cellular RNA structure experiments. Z.L., M.Z., K.L., J.B., and W.H.L performed the sequencing data analysis. R.V.D. and J.D.Y. performed Rosetta modeling. Z.L., W.A.V., and J.D.Y. wrote the paper with input from all coauthors.

Competing interests

The authors declare no competing interests.

Additional information