Nucleic Acids Research, 2025, **53**, gkaf231
https://doi.org/10.1093/nar/gkaf231
Chemical Biology and Nucleic Acid Chemistry

Mapping small molecule–RNA binding sites via Chem-CLIP synergized with capillary electrophoresis and nanopore sequencing

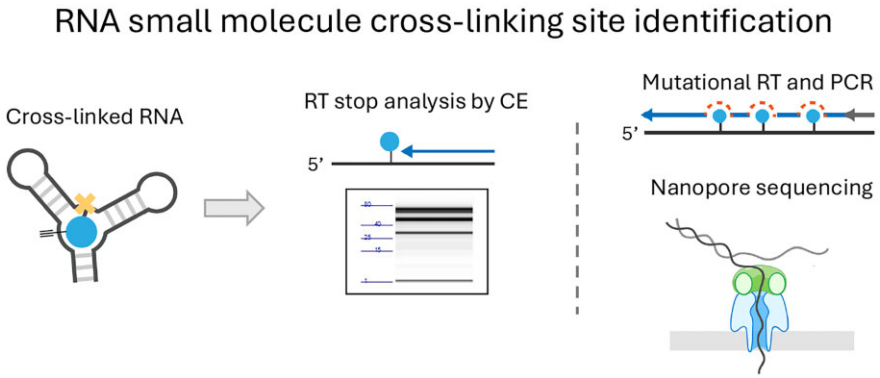
Xueyi Yang[1,2], Jielei Wang[1,2], Noah A. Springer[1,2], Patrick R.A. Zanon[1], Yilin Jia[1,2], Xiaoxuan Su[1,2], Matthew D. Disney[1,2,*]

[1]Department of Chemistry, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, 130 Scripps Way, Jupiter, FL 33458, United States
[2]Department of Chemistry, The Scripps Research Institute, 130 Scripps Way, Jupiter, FL 33458, United States
[*]To whom correspondence should be addressed. Email: mdisney@ufl.edu

Abstract

Target validation and identification of binding sites are keys to the development of bioactive small molecules that target RNA. Herein, we describe optimized protocols to profile small molecule–RNA interactions and to define binding sites of the small molecules in RNAs using covalent chemistry. Various reactive modules appended to an RNA-binding small molecule were studied for cross-linking to the RNA target. Electrophilic modules, whether *N*-chloroethyl aniline or diazirine, have reactive profiles consistent with induced proximity; however, probes with *N*-chloroethyl aniline were more reactive and more specific than those with a diazirine cross-linking moiety. Depending upon the identity of the cross-linking module, covalent adducts with different nucleotides that are proximal to a small molecule's binding site were formed. The nucleotides where cross-linking occurred were elucidated by using two different platforms: (i) automated capillary electrophoresis that identified a binding site by impeding reverse transcriptase, or "RT stops"; and (ii) nanopore sequencing where the cross-link produces mutations in the corresponding complementary DNA formed by reverse transcriptase-polymerase chain reaction amplification of the cross-linked RNA. These approaches are broadly applicable to aid in the advancement of chemical probes targeting RNA, including identifying binding sites and using covalent chemistry to screen for RNA-binding molecules in a high throughput format.

Graphical abstract



Introduction

RNA-targeted small molecules offer an alternative modality to oligonucleotide-based approaches to perturb RNA function [1–3]. Rather than targeting specific sequences, small molecules can bind to or degrade RNA structural elements, thereby affecting biological function [4–6]. Indeed, small molecules targeting RNA have been developed to displace proteins bound to RNA repeat expansions, thereby alleviating gain-of-function disorders [7], to direct pre-messenger RNA (mRNA) splicing outcomes [8–10], and to increase RNA turnover by affecting retention of introns or inclusion of a pre-mature termination codon [11]. With the advancement of small molecule approaches to modulate RNA function, methods for identifying engaged RNA targets are essential, particularly to enable efficient medicinal chemistry optimization, to determine mode of action, and to elucidate off-targets.

Received: December 3, 2024. **Revised:** March 5, 2025. **Editorial Decision:** March 6, 2025. **Accepted:** March 12, 2025

Cleavage- [12–16] and covalent chemistry-based [17–21], strategies have been developed to study the occupancy of RNA targets by small molecules *in vitro* and in cells. Cleavage-based approaches include Ribo-SNAP-Map [13], where an RNA binder is conjugated to a cleaving module, and ASO-Bind-Map [14], a competition experiment between an RNA-binding small molecule and an ASO (antisense oligonucleotide) that induces RNase H-mediated degradation. Target engagement is then studied by depletion of a target or inhibition of depletion, respectively.

The formation of a covalent bond between an RNA and a small molecule binder enables direct target engagement and identification of the binding site, as often employed in the protein-targeting field [22, 23]. Dubbed Chemical Cross-Linking and Isolation by Pull-down (Chem-CLIP), an RNA binder is appended with a cross-linking module, such as chlorambucil [16, 20, 21] or photoreactive diazirine [24–26], and a purification tag. The latter enables the pull-down of probe-modified transcripts with streptavidin beads or by direct copper-catalyzed azide-alkyne cycloaddition (CuAAC) [27, 28] to azide-functionalized beads. The sites of cross-linking can be identified by RNA-sequencing (RNA-seq) analysis, assuming that the cross-linked site impedes reverse transcription (RT) and leads to truncated complementary DNAs (cDNAs) [18]. This Chem-CLIP to Map Small Molecule-RNA Binding Sites (Chem-CLIP-Map-Seq) approach has been implemented transcriptome-wide to afford target engagement profiles (binding landscapes) for RNA-targeted small molecules [24, 26]. Subsequent to development of the Chem-CLIP platform [16, 18, 20, 21], Pearl-Seq was implemented to interrogate small molecule binding sites by assessing RT stops induced by cross-linking of photoaffinity probes [29].

In this report, various synergistic methods to study target engagement that are broadly applicable to RNA targets are described. Utilizing a validated aptamer system [29, 30], we designed and synthesized the aptamer-binding small molecule conjugated to several cross-linking modules and benchmarked them for reactivity using a fluorescence labeling approach. We then used capillary electrophoresis (CE; via a fragment analyzer) to identify rapidly positions of cross-linking ("RT stops") with relatively high accuracy (detectable RT stops have an error range around 1 nucleotide). Various reverse transcriptase (RT) enzymes were examined to determine their effectiveness in inducing mutations at the cross-linking site. By developing a nanopore sequencing approach (Chem-CLIP-MaP-Seq), cross-linking sites were identified at single-nucleotide-resolution by mutational profiling efficiently and economically. Collectively, these studies lay a foundation to profile RNA–small molecule interactions *in vitro* and in complex biological systems. Furthermore, this platform could be applied to screen for RNA-binding compounds in early discovery efforts as well as to define mode of action for more advanced compounds.

Materials and methods

In vitro transcription of RNA

The templates for *in vitro* transcription were purchased from Integrated DNA Technologies (IDT), Inc. (Supplementary Table S1). Polymerase chain reaction (PCR) amplification of the templates was carried out using Phusion High-Fidelity DNA Polymerase (New England Biolabs (NEB), catalog #M0530L) according to the manufacturer's procedure with corresponding primers (Supplementary Table S1). The reaction was thermally cycled using the following conditions: 98°C for 30 s; 35 cycles of 98°C for 7 s, 62°C for 30 s, and 72°C for 10 s; and a final extension at 72°C for 10 min. The PCR products were purified using RNAClean XP beads (Beckman Coulter, catalog #A63987) per the manufacturer's protocol with the following adjustments: 1.8 volumes of RNAClean XP beads and 3.5 volumes of isopropanol relative to the reaction mixture were added to the samples and mixed by pipetting. The mixture was incubated at room temperature (r.t.) for 15 min. A magnetic rack was used to separate the beads from solution, and 85% (v/v) ethanol was used to wash the beads three times. The PCR products were eluted in water at 0.5 volumes of the input.

In vitro transcription was carried out following the protocol from Nilsen *et al.* [31]. First, 10 × Transcription Buffer [1 M HEPES-KOH, pH 7.5, 0.1 M $MgCl_2$, 0.02 M spermidine–HCl, 0.4 M dithiothreitol (DTT), and 1 mg/ml bovine serum albumin] was prepared in nuclease-free water (ZYMO Research, catalog #W1001-30). The transcription reaction (1 ml total volume) was assembled at r.t. as follows: 429 µl of nuclease-free water, 100 µl of 10 × Transcription Buffer, 300 µl of 25 mM (each) rNTPs, 26 µl of 1 M $MgCl_2$, 20 µl of 0.5 M DTT, 50 µl of 0.1 mg/ml purified DNA template, 20 µl of RNase Inhibitor, Murine (NEB; catalog #M0314L) and 10 000 units T7 polymerase. The reaction mixture was incubated at 37°C for 16 h. At the end of the incubation, 10 units RNase-free DNase I (ZYMO Research, catalog #E1010) were added to the sample, which was incubated for an additional 1 h at 37°C. An equal volume of 2 × denaturing loading dye [8 M urea, 2 mM Tris, pH 7.5, 20 mM ethylenediaminetetraacetic acid (EDTA), 0.02% (w/v) orange G] was added to the sample, and the RNA was purified by gel electrophoresis using a denaturing 12% (w/v) polyacrylamide gel (15 cm × 20 cm). Separation was achieved by running the gel at 200 V for 3 h.

The band containing RNA, as determined by ultraviolet (UV) shadowing, was excised, and the RNA was extracted from the gel by tumbling the sample overnight at 4°C in 0.3 M NaCl. The volume of the eluted RNA was reduced by 2-butanol concentration, and the sample was precipitated with 2.5 volumes of ethanol at −80°C for 2 h. The RNA was pelleted by centrifugation at 16 000 × *g* for 30 min at 4°C. The pellet was washed twice with 70% (v/v) ice-cold ethanol, and residual ethanol was removed by vacuum concentration for 2 min. The RNA was dissolved in Nanopure water, and its concentration was measured by its absorbance at 260 nm at 90°C and the corresponding extinction coefficient using a Beckman DU800 UV/vis spectrophotometer.

Folding of RNA

RNA was folded following the protocol in reference [30]. Briefly, RNA was folded in 1.25 × Folding Buffer without $MgCl_2$ (1 × Folding Buffer: 50 mM HEPES, pH 7.4, 150 mM NaCl, and 5 mM $MgCl_2$) at 95°C for 5 min and snap-cooled on ice. Next, the sample was supplemented with 0.25 volume of 25 mM $MgCl_2$ (to afford a final concentration of 5 mM). The RNA was incubated at 37°C for 20 min followed by slowly cooling to r.t. by placing the sample on the bench top for 5 min.

Cross-linking of diazirine-based Chem-CLIP probes

The folded RNA in 1 × Folding Buffer was incubated with a diazirine probe (**3** or **5**) or dimethyl sulfoxide (DMSO, vehicle) at the indicated concentrations and incubated at r.t. for 30 min. The final concentration of RNA was 1 μM, and the concentration of dimethyl sulfoxide (DMSO) in all samples was 1% (v/v). The samples were then irradiated with UV light (365 nm) for 20 min using a UVP cross-linker (UV Stratalinker 2400; caps of the microcentrifuge tubes were open during irradiation). The cross-linked RNA was cleaned up using RNAClean XP beads as described in section "*In vitro transcription of RNA*". This step removes excess, unreacted compound that can interfere with enrichment/pull-down and analysis by TAMRA labeling or by fragment analyzer.

Cross-linking of chlorambucil-based Chem-CLIP probes

The folded RNA in 1 × Folding Buffer was incubated with a chlorambucil probe of interest (**2a**, **2b**, **2c**, **4a**, **4b**, or **4c**) at the indicated concentrations or DMSO (vehicle) at r.t. for 24 h to allow for cross-linking, while 12 and 16 h also produced reasonable signal. The final concentration of RNA was 1 μM and the concentration of DMSO in all samples was 1% (v/v). Cross-linked RNA was cleaned up using RNAClean XP beads as described in section "*In vitro transcription of RNA*". This step removes excess, unreacted compound that can interfere with enrichment/pull-down and analysis by TAMRA labeling or by fragment analyzer.

TAMRA labeling assay

RNA samples were cross-linked using methods described above, with a range of the probe concentrations (0.1, 1, 5, 20, 60, and 100 μM). For each sample, 10 μl of 1 μM cross-linked RNA was combined with 1.24 μl of "click-reaction mixture", comprising $CuSO_4$ (50 mM, 0.1 μl), THPTA (50 mM, 0.5 μl), sodium ascorbate (250 mM, 0.5 μl), TAMRA-azide (50 mM, 0.04 μl), and HEPES buffer (0.1 μl, 250 mM HEPES, pH 7.0). The mixture was incubated at 37°C for 2 h and cleaned up using RNAClean XP beads as described above. The RNA was eluted in 10 μl of nuclease free water. To the sample was added an equal volume of 2 × denaturing loading dye. The sample was mixed thoroughly by pipetting and then separated by electrophoresis using a denaturing 8% (w/v) polyacrylamide gel [1 × Tris–boric acid–EDTA (TBE)] buffer, 30 min at 170 V. The gel was imaged with an Azure Biosystems Sapphire Biomolecular to detect TAMRA-labeled RNA [excitation wavelength of 520 nm; emission wavelength of 565 nm; band pass (BP) of 24 nm]. The gel was then stained with SYBR Gold (1:10 000 dilution in 1 × TBE buffer; Invitrogen, catalog #S11494) to visualize the RNA for 10 min at r.t. with gentle shaking. SYBR signal was measured using an excitation wavelength of 488 nm, an emission wavelength of 518 nm, and BP of 22 nm.

Quantification was carried out using the grayscale image component of ImageJ [32]. The 8-bit grayscale gel images were imported into ImageJ (Fiji) software for analysis. The background was subtracted using the "Rolling Ball" method with an appropriate radius to minimize uneven illumination. The entire row containing lanes of interest were selected using the "Rectangular" tool, and then defined using "Gels→Select First Lane". To account for additional background noise, an adjacent empty lane was also included in the selected area. The integrated density of each band was then able to be plotted using "Gels→Plot Lanes" and the area of individual peaks representing the intensities measured with the "Wand (Tracing) Tool" function. The intensity of the empty lane was subtracted from each band measurement. Data were normalized to the appropriate control, as specified for each figure, to allow for comparison between samples.

Enrichment of cross-linked RNAs for RT stop and mutational analysis

The RNA of interest (1 μM) was folded and cross-linked to a probe as described above in a total volume of 40 μl. To the cross-linked RNA sample was added 5 μl of "click-reaction mixture" composed of $CuSO_4$ (50 mM, 0.4 μl), THPTA (50 mM, 2 μl), sodium ascorbate (250 mM, 2 μl), disulfide-biotin-azide (50 mM, 0.2 μl), and HEPES buffer (0.4 μl, 250 mM HEPES, pH 7.0). The reaction mixture was incubated at 37°C for 2 h and then cleaned up using RNAClean XP beads as described above. For each sample, 100 μl of streptavidin-coated magnetic beads (Dynabeads™ MyOne™ Streptavidin C1, Invitrogen, 65 001) was used to capture the biotinylated RNA. The beads were pre-equilibrated in 1 × Dulbecco's Phosphate Buffered Saline (DPBS) by washing twice with 400 μl of 1 × DPBS on a magnetic rack. The cross-linked RNA was then added to the beads, and the slurry was incubated at r.t. for 1 h with gentle shaking. The beads were then washed three times with 400 μl of 1 × High-Salt Washing Buffer [10 mM Tris–HCl, pH 7.0, 1 mM EDTA, 4 M NaCl, and 0.2% (v/v) Tween-20] followed by three times with 400 μl of 1 × DPBS on a magnetic separation rack. For each wash, the microcentrifuge tubes on the rack were turned 180° to allow the magnetic beads to migrate to the opposite side of the tube. After the final wash, the supernatant was removed, and the beads were resuspended with 10 μl freshly prepared and pre-mixed TCEP (200 mM, 5 μl) and K_2CO_3 (600 mM, 5 μl), and incubated at 37°C for 30 min to allow for cleavage of the disulfide bonds. The free thiols were then capped by the addition of iodoacetamide (400 mM, 10 μl) and incubation at r.t. for 30 min with shaking. The RNA released from the beads (now present in the solution) was cleaned up using RNAClean XP beads as described above.

Quantification of enrichment using RiboGreen Dye

To compare the efficiency of pull-down and enrichment, Quant-iT RiboGreen Dye (Invitrogen, catalog #R11491), a fluorescent RNA-binding dye, was utilized. Aliquots of "Input" samples were taken immediately before addition of the streptavidin beads (4 μl taken from total of 40 μl solution), whereas aliquots of "Enriched" samples were taken following the cleanup of eluted RNA from a total of 10 μl solution. To measure the RNA concentration, 1 μl of "Non-enriched" or "Enriched" sample was diluted with 49 μl of 1 × TBE buffer. The Quant-iT RiboGreen Dye was diluted 1:2000 with 1 × TBE buffer, and 50 μl of this solution was added to each diluted RNA solution in a 96-well black plate (Corning, catalog #3686). Fluorescence intensity (relative fluorescence units, RFU) was measured using a Tecan Infinite M1000 Pro with excitation/emission wavelengths of 500 nm/525 nm. Fluo-

rescence intensity in each well was adjusted by subtraction of background signal ("Background fluorescence signal"), from the average measured fluorescence from three wells containing no RNA (buffer and RiboGreen Dye only). Percent pull-down was calculated by taking the ratio of the adjusted "Non-enriched" and "Enriched" values for each sample individually, shown using Equation 1.

$$\%\text{Pull}-\text{down} = \frac{\frac{\text{Fluorescence signal from Enriched RNA}-\text{Background fluorescence signal}}{\text{Fluorescence signal from Non}-\text{enriched RNA}-\text{Background fluorescence signal}}}{3.6} \cdot 100 \quad (1)$$

The correction factor of 3.6 was included to account for the difference in volumes of the "Non-enriched" samples (1 µl of the reaction was measured directly as the "Non-enriched" concentration) and "Enriched" samples [36 µl of the 40 µl reaction was carried forward to the pull-down steps and eluted at the end in 10 µl of water; 1 µl of the eluted RNA was used to measure the concentration. This results in a theoretical 3.6-fold (36 µl/10 µl) higher concentration of RNA, assuming 100% recovery].

Reverse transcription of RNA

Typically, 40 ng of RNA (either enriched or non-enriched) was used as template for each RT reaction. However, for negative control samples that have low RNA concentration (<1 ng/µl), for example, the enriched samples treated with DMSO or control probe, one-third of the total volume after enrichment was used as template for the RT reaction. As little as 1 ng of RNA was detectable in our hands.

SuperScript III

To induce RT stops, cross-linked RNA was reverse transcribed using SuperScript III (SSIII; Invitrogen, catalog #18 080 093) following the manufacturer's protocol. Briefly, 1–40 ng RNA was mixed with 1 µl of 2 µM Aptamer RT Primer, 1 µl of 10 mM dNTPs, and nuclease-free water to reach a final volume of 13 µl. The primer was annealed by heating the samples at 75°C for 5 min then slowly cooling to 35°C at a rate of 0.1°C/s in a thermocycler. Alternatively, samples were snap cooled on ice for 3 min. No significant differences were observed upon analysis of the samples using the two different annealing procedures.

To each sample, 4 µl of 5 × First Strand Buffer (provided by the manufacturer), 1 µl of 0.1 M DTT, 1 µl of RNaseOUT, and 1 µl (200 units) of SSIII were added. Reverse transcription was carried out at 55°C for 1 h, followed by heat inactivation of the SSIII by incubating the samples at 70°C for 15 min. Subsequently, 1 µl (5 units) of RNase H (NEB, catalog #M0297L) and 1 µl (5 units) of RNase T1 (Thermo Scientific, catalog #EN0541; diluted to 5 units/µl in 1 × Dilution Buffer: 50 mM Tris–HCl, pH 7.5, and 2 mM EDTA) were added to each sample to digest the RNA. Following incubation at 37°C for 30 min, the cDNA was purified using 1.8 × volumes of Ampure XP beads (Beckman Coulter, catalog #A63881) and 3.5 volumes of isopropanol. After washing with 85% (v/v) ethanol twice, the cDNA was eluted from the beads with 10 µl of Nanopure water. The amount of cDNA was quantified using an Nanodrop Microvolume Spectrophotometer (Thermo Scientific).

SuperScript II and SuperScript IV

RNA mutational RT was performed following a previously published protocol [33]. The RNA eluted from RNAClean XP beads was mixed with 0.2 µl of 10 µM Aptamer RT Primer, 1 µl of 10 mM dNTPs and nuclease-free water to reach a final volume of 10 µl. The primer was annealed by heating at 70°C for 5 min and immediately cooled on ice for 2 min. Next, 2.22 × freshly prepared RT buffer (111 mM Tris–HCl, pH 8.0, 167 mM KCl, 13.3 mM $MnCl_2$, 22 mM DTT and 2.22 M betaine) was added to the RNA-primer mixture and incubated at r.t. for 2 min, after which 200 U Superscript II (Invitrogen, catalog #18 064 014) was added. The samples were incubated in a thermocycler as follows: 10 min at 25°C, 90 min at 42°C; 10 cycles of 2 min at 42°C and 2 min at 50°C; followed by 10 min at 70°C to inactivate the enzyme. An alternative protocol was also tested for Superscript II and Superscript IV (Invitrogen, catalog #18 090 010), where the RT buffer did not contain betaine. The RNA samples were incubated in a thermocycler at 25°C for 10 min, 42°C for 3 h, and 70°C for 15 min. The reverse transcribed samples were treated with RNase H and RNase T1, cleaned up and quantified as described above for SuperScript III.

Induro RT

RNA was reverse transcribed by Induro Reverse Transcriptase (Induro RT, NEB, catalog #M0681L) following the manufacturer's protocol. In brief, 0.2 µl of 10 µM Aptamer RT Primer and 1 µl of 10 mM dNTPs were added to the RNA sample, and the total volume was brought to 10 µl total volume using nuclease-free water. The RNA/primer was denatured at 65°C for 5 min and immediately cooled on ice for 2 min. The mixture was then supplemented with 4 µl 5 × RT Buffer (supplied by the manufacturer), 0.2 µl (8 units) of RNase inhibitor (NEB, catalog #M0314), 200 units of Induro RT, and nuclease-free water to a final volume of 20 µl. Reverse transcription was carried out at 55°C for 30 min, followed by heat inactivation of the enzyme by incubating the sample at 95°C for 1 min. The samples were then treated with RNase H and RNase T1, cleaned up, and quantified as described above for SuperScript III.

TGIRT III

RNA was reverse transcribed by TGIRT III (Ingex) following a previously described protocol [34]. The RNA was mixed with 0.2 µl of 10 µM Aptamer RT Primer and 1 µl of 10 mM dNTPs, and the total volume was brought to 15 µl using nuclease-free water. The primer was annealed by heating the sample at 65°C for 5 min and immediately cooling on ice for 2 min. Subsequently, 4 µl freshly prepared 5 × FS Buffer (200 mM Tris–HCl, pH 8.0, 300 mM KCl, 24 mM $MnCl_2$, and 40 mM DTT) was added to the reaction followed by the addition of 200 units TGIRT III. Reverse transcription was carried out at 25°C for 10 min, 57°C for 1.5 h and 70°C for 15 min. The samples were then treated with RNase H and RNase T1, cleaned up, and quantified as described above for SuperScript III.

MarathonRT

RNA was reverse transcribed by MarathonRT (Kerafast) following the previously described protocol [35]. Briefly, RNA was mixed with 0.2 µl of 10 µM Aptamer RT Primer and 1 µl of 10 mM dNTPs. The RNA and primer were annealed by heating at 65°C for 5 min followed by immediately cooling on

ice for 2 min. Next, 2.22 × freshly prepared RT buffer (111 mM Tris–HCl, pH 8.0, 167 mM KCl, 13.3 mM $MnCl_2$, and 22 mM DTT) was added to the RNA-primer mixture and incubated at r.t. for 2 min. MarathonRT (40U) was added to the reaction mixture, and the samples were incubated in a thermocycler at 25°C for 10 min, 42°C for 3 h, and 70°C for 15 min. The samples were then treated with RNase H and RNase T1, cleaned up, and quantified as described above for SuperScript III.

Analyzing RT products using a fragment analyzer

The RT products were assessed using Agilent 5300 Fragment Analyzer equipped with capillaries that are 55 cm in length (catalog #A2300-1250-5580). Samples were prepared following the manufacturer's protocol for the Small RNA Kits (Agilent, catalog #DNF-470–0275). Briefly, 2 µl of eluted cDNA was heated at 70°C for 10 min followed by addition of 18 µl of Small RNA Diluent (provided with the kit). To account for differences in mobility of cDNA relative to RNA in general and Aptamer 21 in particular, a custom ladder was created by mixing DNA oligonucleotides with lengths of 16 (100 nM), 28 (25 nM), 40 (15 nM), 52 (12 nM), 64 (10 nM), and 79 (8 nM) nucleotides (Supplementary Table S1), with sequences identical to that of Aptamer 21 cDNA; different concentrations of each oligonucleotide were used as longer oligonucleotides bind to more dye molecules than smaller ones and hence give stronger signals. The method for separation included a 30 s pre-run at 11.5 kV, followed by a 50 s sample injection at 8.0 kV, with 45 min of 11.5 kV for separation. Gel images were created using the ProSize Data Analysis software. The size of each truncated cDNA bands was estimated by the software by comparison to the customed ladder using the "point to point" curve fitting option.

For analysis involving raw electropherograms, the raw data were exported from the ProSize Data Analysis software using the "time" option rather than the "nt" option. Because well-to-well variability in the time it takes for the full-length aptamer cDNA to elute from the capillary is observed (https://www.agilent.com/cs/library/applications/application-poly-tails-rna-fragment-analyzer-5994-5325en-agilent.pdf) [36], a time correction value (TCV) was calculated and applied for each well by normalizing to the time of elution (TE) for the 79 nt peak in the custom Aptamer ladder according to the equations below.

$$\mathrm{TCV} = \frac{\mathrm{TE_{well}} - \mathrm{TE_{marker}}}{\mathrm{TE_{Ladder}} - \mathrm{TE_{marker}}}$$

After calculating the TCV for each well, a normalized time scale was calculated for each well via the equation below:

$$t_{\mathrm{corr}} = \frac{(t_{\mathrm{n}} - \mathrm{TE_{marker}})}{\mathrm{TCV}} + \mathrm{TE_{marker}}$$

where t_{corr} is the corrected time at each time point (t_{n}). To normalize for differences in loading between the wells, normalized intensity ($\mathrm{RFU_{norm}}$) at each time point was calculated for each well via the equation below:

$$\mathrm{RFU_{norm}} = \frac{\mathrm{RFU}_{t_n}}{\frac{MaxF_{\mathrm{well}}}{MaxF_{\mathrm{Ladder}}}}$$

where MaxF is the maximum fluorescence that corresponds to the full-length (79 nt long) aptamer peak. The normalized intensity at each corrected time was plotted and overlayed in GraphPad Prism.

Library preparation and nanopore sequencing

The cDNA from reverse transcribed RNA using Super Script II reverse transcriptase (mutational read-through) was amplified using Phusion High-Fidelity DNA Polymerase (NEB, catalog #M0530L) according to the manufacturer's protocol. Briefly, 4 µl of 5 × Phusion HF Buffer, 0.4 µl of 10 mM dNTPs, 0.1 µl of 100 µM PCR primers (RT primer also serves as the reverse primer for PCR amplification; Supplementary Table S1), 0.2 µl/0.4 units of Phusion™ High–Fidelity DNA Polymerase, and 0.6 µl of DMSO were mixed with 0.1–10 ng of cDNA, and nuclease-free water to afford a final volume of 20 µl. The PCR amplification was carried out using the following conditions: 98°C for 30 s; 25 cycles of 98°C for 7 s, 62°C for 30 s and 72°C for 10 s; and a final extension at 72°C for 10 min. The quality of the PCR products was assessed by running a 2 µL aliquot of the PCR reaction mixture on a 10% (w/v) denaturing polyacrylamide gel. The PCR products were imaged by staining with SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen, catalog #S11494). The PCR products were cleaned up using Ampure XP beads as described above in section "*Reverse transcription of RNA*". The amount of PCR amplicons was quantified Nanodrop Microvolume Spectrophotometer (Thermo Scientific).

The RNA sequencing library was prepared using a Native Barcoding Kit 24 V14 (Oxford Nanopore Technology, SQK-NBD114.24) following the manufacture's protocol. In brief, 10–50 ng of DNA amplicon was used as input for each sample. Since the DNA amplicon was <100 bp, for the first clean-up step involving SPRI beads, 1.8 volumes of beads and 3.5 volumes of isopropanol were added to ensure proper binding of short amplicons. For the second clean-up step involving SPRI beads, 1.8 volumes of beads were used (no addition of isopropanol). The library was sequenced using a MinION R10.4.1 flow cell (Oxford Nanopore Technology, catalog #FLO-MIN114) using MinKNOW software per the manufacturer's protocol.

RNA sequencing data analysis

Sequencing data were base-called using Dorado basecaller with the following command: "dorado basecaller dna_r10.4.1_e8.2_400bps_sup@v4.3.0./pod5/ -r –kit-name SQK-NBD114-96 > all_calls.bam". The base-called reads were demultiplexed by command "dorado demux –output-dir./barcodes/ –no-classify all_calls.bam". The base-called reads were then aligned to target sequence (in .fasta format) using Dorado aligner. The alignment was executed similarly for each barcode, for example, "dorado aligner Aptamer21.fasta./barcodes/SQK-NBD114-96_barcode01.bam > ./aligned/sample1_aligned.bam". The resulting bam files were sorted and indexed using samtools as follows: "samtools sort./aligned/sample1_aligned.bam -o./sorted/sample1_sorted.bam"; "samtools index -@ 8 -M./*sorted.bam". The per base mutation information was summarized using samtool mpileup: "samtools mpileup -aa -d 0 -B -Q 10 -f Aptamer21.fasta./sorted/sample1_sorted.bam 〉 ./pileup/sample1_sorted.pileup" for each sample. A customed code was used to extract per base coverage and mutation information (including substitution, deletion, and insertion) from each mpileup file. The mutation rates were

calculated as the number of mutations divided by the coverage at each nucleotide position. The mutation rates were either plotted directly for each probe, or normalized by subtracting the control probe (**4a** or **5**) from the respective HAP derived probe (**2a** or **3**).

For the analysis of direct RNA sequencing results, "rna004_130bps_hac@v3.0.1" base-calling mode was used with the "–emit-moves" option on, enabling downstream visualization of raw data. The data were then processed in the same manner as the DNA sequencing data described above.

Nucleoside digestion of covalent adducts

Aptamer 21 RNA was reacted with **2a**, as described above in section "*Cross-linking of chlorambucil-based Chem-CLIP probes*". Following clean-up of the RNA via RNAClean XP beads, 100 ng of RNA as measured by Nanodrop Microvolume Spectrophotometer (Thermo Scientific) was digested to nucleosides using a commercial enzyme mix (NEB, catalog #M0649S) in a total volume of 20 µl, per the manufacturer's protocol. The 20 µl reaction was diluted to 100 µl using Liquid Chromatography-Mass Spectrometry (LC-MS) grade water. Afterwards, 1 µl of the digestion reaction was injected with a Vanquish HPLC system (Thermo Scientific) into an Orbitrap Exploris 120 (Thermo Scientific) mass spectrometer. The digested nucleosides were separated using the following water:acetonitrile (0.1% formic acid) gradient: 2 min at 2% acetonitrile, a 5 min 2%–98% acetonitrile gradient, then 2 min at 98% acetonitrile. Data were analyzed in the Thermo Scientific Freestyle software by searching for the expected m/z values (+1, +2, and + 3 charges) for adducts of **2a** to each of the four nucleosides.

Results

Compound design and synthesis to study target engagement in an aptamer model system

The heteroaryldihydropyrimidine (HAP) ligand, **1**, was previously selected to bind the RNA "Aptamer 21" with high affinity and selectivity, making it a suitable model system for both *in vitro* and cellular applications (Fig. 1A and B) [29, 30]. HAP was designed based on a compound that affects the assembly of the hepatitis B virus [37] and contains a triazole linkage substituted with a 2-(2-ethoxymethoxy)ethanol, which can readily be replaced with alternative functional groups, for example various electrophiles, via CuAAC [27, 28]. The high affinity of HAP for the aptamer as well as its straightforward modification makes this model system desirable for both *in vitro* and in cellular applications [38]. Various reactive modules were installed onto the HAP compound, including *N*-chloroethyl aniline [16, 20, 21] and diazirine [24–26] (Fig. 1B).

N-chloroethyl anilines such as chlorambucil react with nucleic acids through *in situ* formed aziridinium ions and have been appended to RNA/DNA ligands to induce structure- or sequence-specific modifications [16, 20, 39, 40]. Traditional nitrogen mustards used as chemotherapeutics are "bifunctional" ligands, capable of generating cytotoxic interstrand cross-links in DNA by using both chloroethyl substituents [41, 42]. Monofunctional nitrogen mustards (i.e. only a single chloroethyl group) are expected to mitigate this DNA-mediated toxicity [43], while still capable of effective modification of a target RNA [21]. To expedite downstream studies, we synthesized two related *N*-chloroethyl aniline molecules bearing an alkyne handle (**2a**, **2b**; Fig. 1B) in addition to the HAP-chlorambucil conjugate **2c**. Moreover, a HAP-diazirine conjugate, **3**, was also synthesized (Fig. 1B), which upon UV irradiation will form reactive diazo or carbene intermediates [44]. Respective control probes lacking the HAP binding module (**4a–c**, **5**) were also synthesized to assess selective reactivity of the HAP-conjugates with the aptamer (Fig. 1B).

Together, these Chem-CLIP probes share the same general structure, harboring an RNA-binding moiety (HAP), a reactive module (*N*-chloroethyl aniline or diazirine) and an enrichment handle (alkyne). Each of these components is programmable, that is they can be modified for various purposes to enable target validation of other RNA binders (by replacement of HAP) and identification of novel chemical moieties that cross-link to RNA (by replacement of the *N*-chloroethyl aniline or diazirine). These probes were used to develop a platform to study target engagement and to map small molecule binding sites (Fig. 1C).

A fluorescence-based labeling assay to assess target engagement

As a first comparison of the HAP Chem-CLIP probes, we developed a simple reactivity-based assay (Fig. 2A) [45]. In brief, the probes were incubated with Aptamer 21 to allow formation of cross-links; for **3**, the samples were also irradiated with UV light. Cross-linking was detected by CuAAC between the probe's alkyne handle and an azide-functionalized tetramethylrhodamine (TAMRA) fluorescent dye, and subsequent visualization by polyacrylamide gel electrophoresis (PAGE). Our intent for analysis by PAGE was to enable simultaneous quantification of the extent of reaction with the RNA aptamer (TAMRA signal) and RNA loading via SYBR staining. However, as the extent of TAMRA labeling increased, interference with SYBR staining was observed (Fig. 2B–D), likely due to steric impendence SYBR binding to cross-linked RNA. Thus, the extent of TAMRA labeling afforded by cross-linking of a HAP-conjugate was compared to its respective control probe.

All small molecule probes harboring the HAP binding moiety labeled Aptamer 21, albeit to varying extents (Fig. 2B–D and Supplementary Fig. S1). To evaluate the effect of nonspecific cross-linking stemming from the reactive modules themselves, the TAMRA signal intensities of HAP-conjugates were compared to the respective control probes, e.g. **2a** versus **4a** (Fig. 2B–D). For all three pairs of probes (*N*-chloroethyl aniline probes **2a** and **4a**, **2b** and **4b**, along with diazirine probes **3** and **5**), the TAMRA fluorescence intensities were concentration-dependent, concomitant with an increase in HAP-to-control probe signal ratio. Specifically, TAMRA signal for **2a** peaked at 20 µM, with the maximum HAP-to-control probe ratio reaching 14 ± 2 at 60 µM. Compound **2b** exhibited a more rapid increase in TAMRA signal, achieving saturation at 5 µM and attaining a peak HAP-to-control probe signal ratio of 23 ± 9, before nonspecific cross-linking of **4b** was observed at higher concentrations. The HAP-diazirine **3** displayed a steadier trend in both TAMRA fluorescence and signal ratio across the test range of concentrations, although the overall magnitudes for both were markedly lower compared to the chlorambucil-based probes. Notably, the peak ratio observed at 100 µM was only around 4 ± 0.9, which was ~3.5 ± 0.9-fold and 5.8 ± 2.6-fold less than observed for the **2a/4a** and **2b/4b** pairs, respectively. This differ-

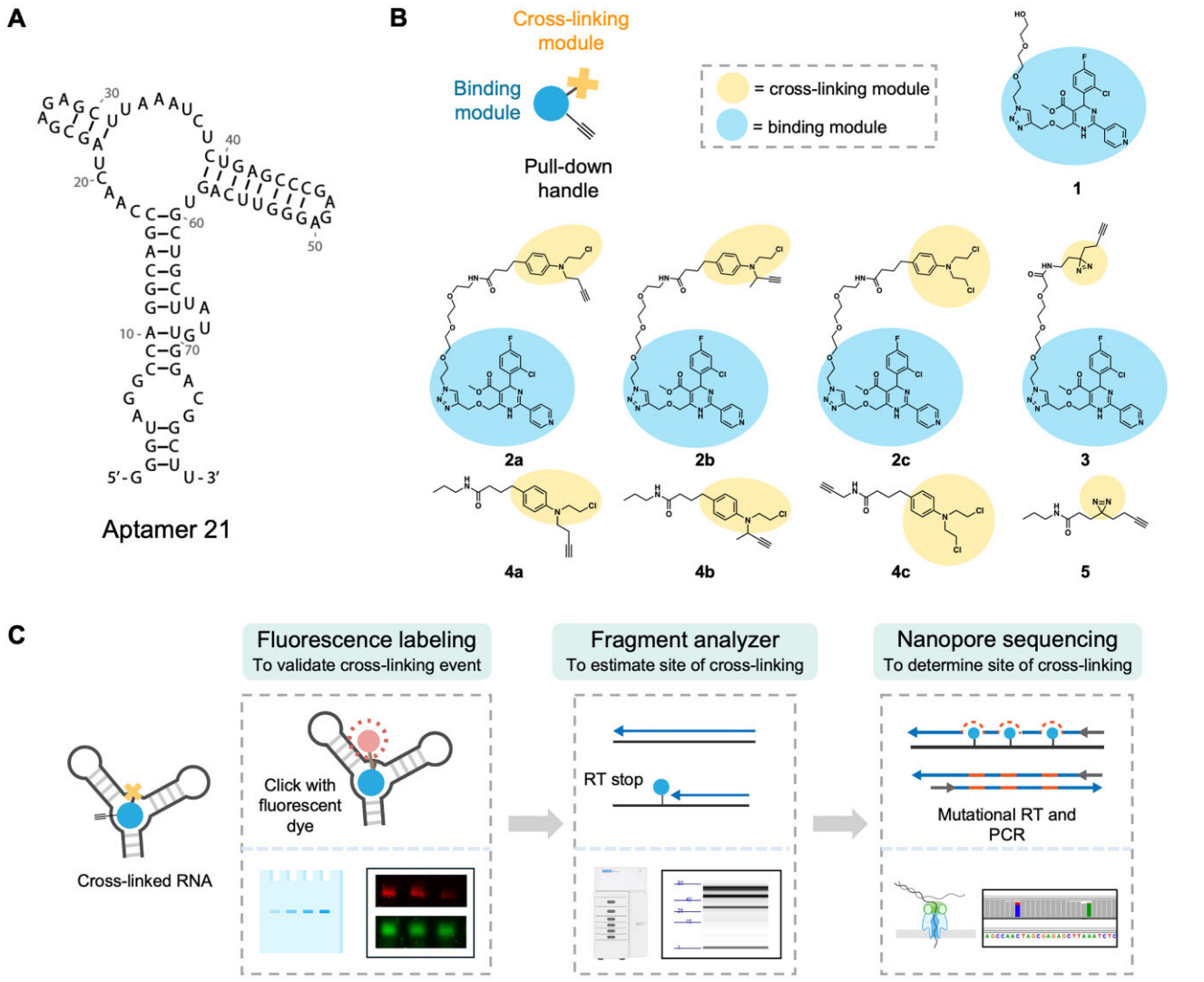


Figure 1. Development of methods to define small molecule binding sites within a target RNA by using covalent chemistry. (**A**) Secondary structure of Aptamer 21. (**B**) Design of small molecule probes to define binding sites and specific probe structures. The probes are modular in nature, each containing an RNA-binding module, a cross-linking module and an alkyne handle. (**C**) Schematics of three different workflows to assess binding and map small molecule binding sites. TAMRA labeling was used to confirm cross-linking events, followed by RT stop analysis using a fragment analyzer. Finally, cross-linked sites were defined by nanopore sequencing via either RT stops or mutational mapping.

ence for the chlorambucil- versus diazirine-based Chem-CLIP probes suggests that the former (**2a, 2b**) have a higher cross-linking efficiency. The lower extent of cross-linking observed for the diazirine probe could be due to a variety of factors, including inefficient activation by UV light, quenching of the reactive carbene intermediate with water [44, 46], or less favorable positioning of the reactive group observed for other biomolecules [25, 47].

As a further comparison of the reactivity of the probes, competition experiments were completed. In each experiment, Aptamer 21 was incubated with: (i) a Chem-CLIP probe containing a reactive module and an alkyne handle and (ii) a probe with a chlorambucil reactive module but lacking the alkyne handle (**2c**, Fig. 1C). For example, the reactivity of **2a**, which contains HAP, a *N*-chloroethyl aniline, and an alkyne, was studied in the presence of **1** (HAP) and **2c**. Both **1** and **2c** reduced the extent of TAMRA labeling due to reaction of **2a** dose-dependently, where the hydroxyl compound (**1**) competed with **2a** to a greater extent than **2c** did (Supplementary Fig. S2A). This trend was also observed in competition experiments between **2b** and **2c** or **1** (Supplementary Fig. S2B) as well as between **3** and **2c** or **1** (Supplementary Fig. S2C). These somewhat surprising results suggest that addition of the electrophiles is modestly affecting molecular recognition of the aptamer by the HAP-binding element. We therefore studied whether the probes reacted with previously published control RNA construct (21-R, Supplementary Fig. S3) that does not bind HAP [30]. Each of the HAP Chem-CLIP probes, **2a**, **2b**, and **3**, reacted to the greatest extent with Aptamer 21, as compared to 21-R by ~11.5 ± 1-fold (raw signals of 7003 ± 609 versus 608 ± 29; 60 μM treatment), 8 ± 3-fold (raw signals of 11 164 ± 3007 versus 1377 ± 372; 60 μM treatment), and 5 ± 0.8-fold (raw signals of 3419 ± 496 versus 678 ± 52; 60 μM treatment), respectively (Fig. 2 and Supplementary Fig. S3). In contrast, the control Chem-CLIP probes, **4a**, **4b**, and **5** reacted less with Aptamer 21 than with 21-R, with corresponding ratios of 0.28 ± 0.03 (raw signals of 515 ± 50 and 1845 ± 94, re-

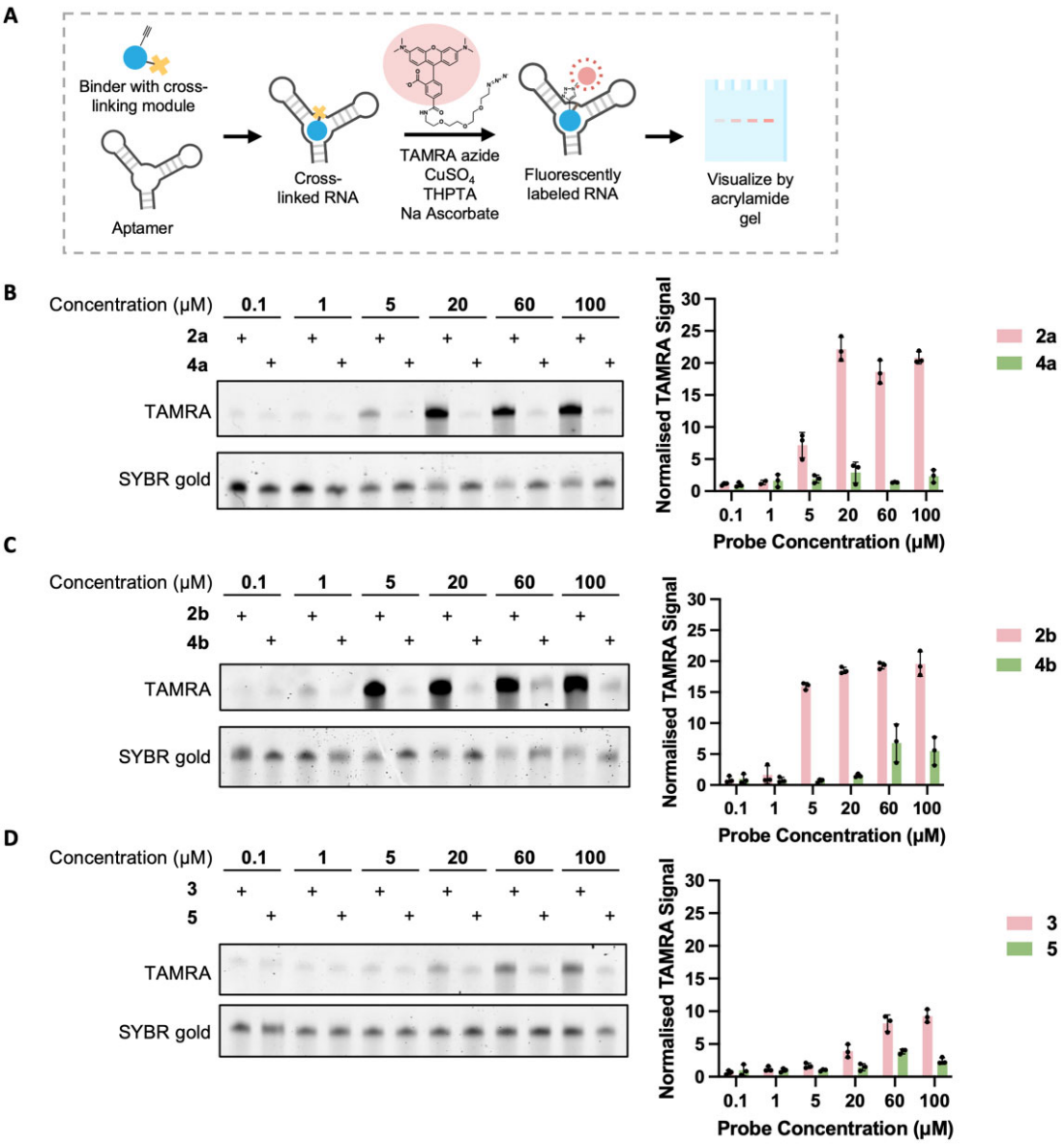


Figure 2. TAMRA labeling assay to determine cross-linking events. (**A**) Scheme of the cross-linking and TAMRA labeling assay. (**B**–**D**) Dose responses of the reactions of probes 2a and 4a (chlorambucil-based) (**B**) 2b and 4b (chlorambucil-based) (**C**); and 3 and 5 (diazirine-based) (**D**) with Aptamer 21 (1 μM), and quantification thereof. These studies also informed the optimal concentration for distinguishing HAP-driven cross-linking from nonspecific reaction of the cross-linking module, i.e. control probes. TAMRA signals were normalized to the group exhibiting the weakest TAMRA fluorescence, the signals of the control compounds (4a, 4b, 5, respectively) at 0.1 μM.

spectively; 60 μM treatment), 0.53 ± 0.17 (raw signals of 3733 ± 1084 and 7074 ± 1045, respectively; 60 μM treatment), 0.97 ± 0.30-fold (raw signals of 1559 ± 232 and 1 607 ± 428, respectively; 60 μM treatment). Comparison of the Aptamer 21-to-21-R ratios for **2a/4a**, **2b/4b**, and **3/5** probe pairs further demonstrated the specificity of the HAP-functionalized molecules towards Aptamer 21, with respective fold differences of 41 ± 6, 15 ± 7 and 5 ± 2.

Validation of the fluorescence-based labeling assay by mass spectral analysis

To verify that TAMRA labeling indeed detects formation of RNA-Chem-CLIP probe adducts, the **2a**-labeled Aptamer 21 was subjected to mass spectrometry analysis after nucleoside digestion. While *N*-chloroethyl anilines preferably react with N7 of guanosine [42], only modifications of adenosine and cytidine by **2a** were detected (Supplementary Fig. S4). The lack of guanine residues near the compound binding site is likely the primary reason for this observation. Although to lesser extent, chlorambucil derivatives react with N3 of cytosine and N3 of adenine [48], which are likely the cross-linking sites of **2a**, although this could not be unambiguously assigned from the mass spectra.

Mapping compound binding sites as indicated by RT stops by capillary electrophoresis

The covalent cross-links from reaction with *N*-chloroethyl aniline (UV-independent) or diazirine (UV-triggered) probes can induce RT stops at the site of cross-linking, enabling the binding site to be identified by electrophore-

sis (Fig. 3A) [18, 49, 50]. Previously, *in vitro* binding site mapping was completed by primer extension using a radioactively or fluorescently labeled primer or by subcloning and sequencing the resultant cDNA [18, 29]. To improve throughput, we developed an approach that identifies binding sites using a fragment analyzer capillary electrophoresis (CE) system that enables high throughput (96-well format) and quantitative determination of RT stop sites with small amounts of input cDNA (as little as 50 pg) (https://www.agilent.com/cs/library/usermanuals/public/quick-guide-dnf-470-small-rna-kit-SD-AT000130.pdf) [51], without the need of PCR amplification. Following covalent adduct formation, Aptamer 21 was reverse transcribed with SuperScript III, an enzyme commonly used to generate RT stops at the sites of cross-linking [5]. The RT primer is complementary to the 3′ end of the RNA, and thus the positions of the RT stops can be calculated from the lengths of the truncated cDNA products which were determined via CE and fitted to a custom ladder. The ladder was generated from six oligonucleotides of varying lengths that contain the same sequence as the reverse transcribed cDNA. This customized ladder was necessary for accurate size determination, as sequence and structural elements of the Aptamer 21 RT products resulted in differential electrophoretic mobility relative to the single-stranded RNA ladder provided by the vendor of undefined sequence (Supplementary Fig. S5A).

Following 24 h incubation of Aptamer 21 with **2a** (20 μM) and subsequent RT with SuperScript III, the cleaned-up cDNA products were directly analyzed via CE. (Note: detectable cross-linking was also observed after 12 and 16 h.) For DMSO-treated samples, a minor peak was observed at length of 65 nucleotides, corresponding to nucleotide A14 in Aptamer 21; the percentage of truncated cDNAs was 6 ± 0.2% (Fig. 3B and Supplementary Fig. S5C). Likely, this RT stop is induced by a stable structure formed in this region (Fig. 1A). In **2a**-treated samples, two primary RT stop products with lengths of 45 and 60 nucleotides (assuming RT stops terminate at the position prior to adduct formation), corresponding to adducts formed with nucleotides A34 and A19 of Aptamer 21 (Fig. 3B and D and Supplementary Fig. S5C). The RT stops observed at A34 and A19 (note the neighboring nucleotide C20) are consistent with the adenosine adducts observed in mass spectral analysis (Supplementary Fig. S4). The percentage of truncated cDNAs at A34 and A19 were 19 ± 0.1% and 32 ± 0.3%, respectively. In contrast, only 9 ± 0.3% of RT stops were observed for control probe **4a**, like that of DMSO-treated samples and consistent with the low extent of TAMRA labeling observed at the 20 μM concentration (Fig. 3B, Supplementary Figs S1A and S5C). For compound **3**, two UV-dependent RT stops were identified with lengths of 41 and 59 nucleotides, corresponding to U38 and C20 in Aptamer 21, with the percentage of truncated cDNA at these positions equaling 2 ± 0.3% and 5 ± 0.4%, respectively (Fig. 3C and Supplementary Fig. S5D). No UV-dependent RT stops were observed for control probe **5**. The total percentage of truncated cDNAs for **3** was significantly less than that observed for **2a** (52 ± 0.3% versus 6 ± 0.7%), in agreement with the TAMRA labeling assays where **2a** had a four-fold higher TAMRA/control probe ratio than **3.** The RT stop sites and percentage of truncated cDNA for **2b** were like **2a** (Supplementary Fig. S5B). Thus, only **2a** was carried forward for the downstream studies.

To enrich covalently modified RNAs and hence increase the signal from RT stops, particularly for diazirine probe **3**, the alkyne functional handle was clicked to disulfide-biotin-azide, which enabled pull-down via streptavidin beads and elution of enriched RNA under reducing conditions. For compound **2a**, 30 ± 6% of input RNA was pulled down, whereas only 5 ± 0.2% of input RNA was pulled down by control probe **4a** (Supplementary Fig. S5E). In contrast, only 0.6 ± 0.1% of the Aptamer 21 RNA was pulled down by **3**. Despite the lower overall pull-down, **3** significantly enriched Aptamer 21 to a greater extent than control probe **5** (0.11 ± 0.03%) or DMSO (0.07 ± 0.05%). Furthermore, no pull-down by **3** was observed in the absence of UV irradiation (0.02 ± 0.04% pulldown). Both observations demonstrate the dependence of the pull-down on both the HAP binding component and UV-activated diazirine cross-linking (Supplementary Fig. S5E). These findings agree with the RT stop and TAMRA labeling assays.

The enriched (pulled down) RNA was then subjected to RT and RT stop analysis via CE. For **2a**, nearly all RT products following the pull-down procedure correspond to the two RT stop sites identified without pull-down and nearly no full-length, unmodified RNA remained (Fig. 3B). No noticeable RT products were observed for untreated (DMSO) or **4a**-treated Aptamer 21 (Fig. 3B), suggesting these sites of modification are induced by the aptamer-binding HAP component. Similarly, for **3**, minimal full-length RNA was observed, and the two primary RT stop sites observed were the same as those identified in the absence of pull-down (Fig. 3C). The enriched samples exhibited different mobility on CE compared to the non-enriched samples, which is likely due differences in salt concentration. Nonetheless, the similarity in RT stop patterns between the enriched and non-enriched samples supports that the RT stop sites are identical (Fig. 3C and Supplementary Fig. S5D). Importantly, these two RT stops were UV-dependent both in the input and enriched fractions, indicating that the RT stops are specifically induced from the UV activation of the diazirine moiety (Fig. 3C).

The RT stop sites identified for **2a** and **3** mapped to the multibranched loop on aptamer 21 (Fig. 3D), close to the proposed binding site of HAP binder [29]. Two cross-linked sites were detected for each small molecule probe, which could be attributed to the flexibility of the reactive module and the polyethylene glycol (PEG) linker. Although less likely, the binder might have slightly different binding modes, leading to cross-linking at various sites.

Identification of a reverse transcriptase(s) suitable for mutational profiling of small molecule–RNA adducts

After demonstrating that CE can map small molecule binding sites as indicated by the length of the truncated cDNA products, we sought to obtain single-nucleotide-resolution of these sites via mutational profiling and RNA sequencing (MaP-Seq). To date, mutational mapping is predominantly utilized in the context of RNA structural probing, where unstructured RNA bases are labeled with chemical reagents such as dimethyl sulfate (DMS) [34, 52] and *N*-acyl-imidazoles or isatoic anhydrides for selective 2′-hydroxyl acylation (SHAPE) [53]. Compared to the covalent RNA binders investigated here, adducts formed by DMS or SHAPE probing are typically smaller, and

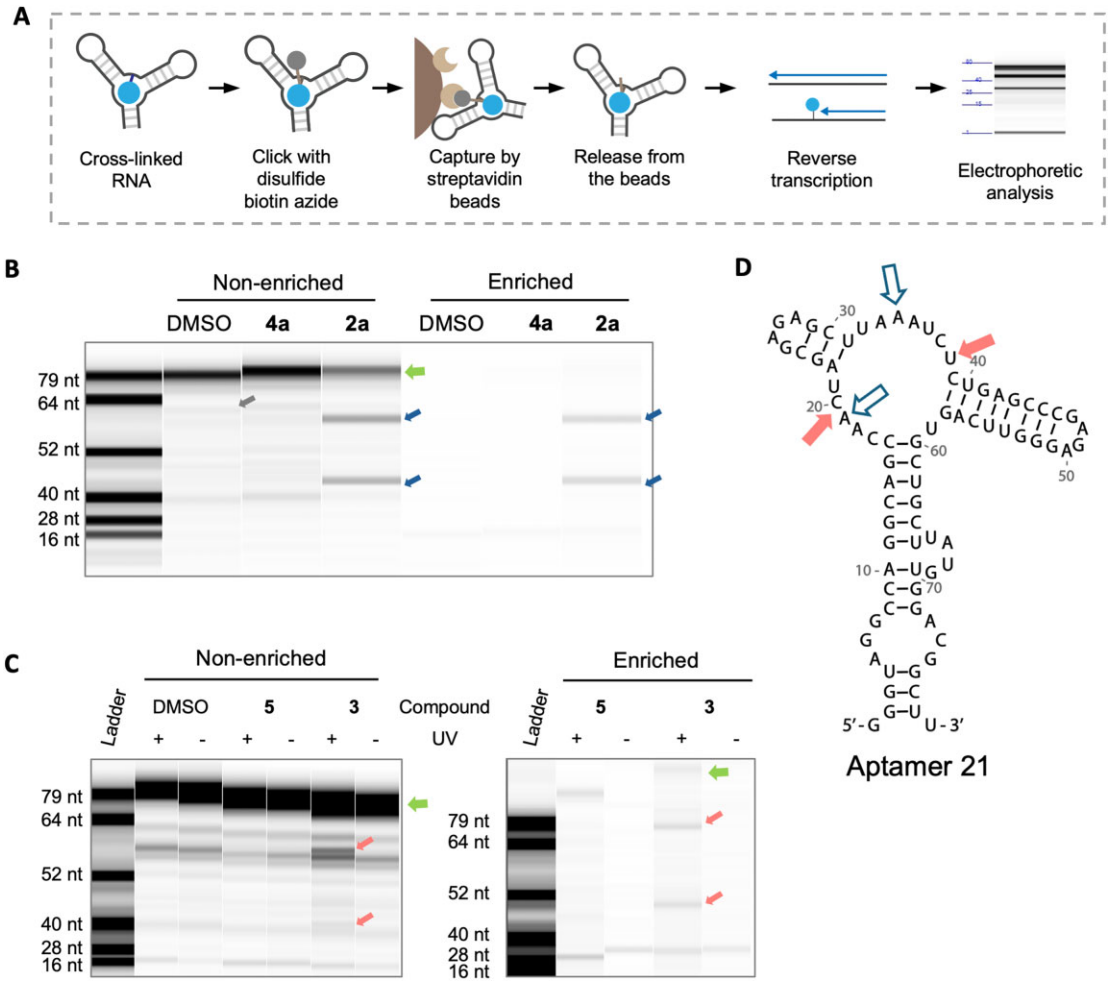


Figure 3. Identifying cross-linked (binding) sites using capillary electrophoresis. (**A**) Scheme of Chem-CLIP and RT stop analysis using capillary electrophoresis. (**B**) Representative fragment analyzer trace to identify cross-linking sites of 2a (chlorambucil-based Chem-CLIP probe) in Aptamer 21 by RT stops (indicated with arrows; RT stops in DMSO and 2a-treated samples, angled arrows; full-length RNA, horizontal arrows), with or without affinity enrichment. Only cross-linked RNAs remain after the enrichment step. (**C**) Representative fragment analyzer traces identified unique, UV-dependent RT stop sites for 3 (indicated with arrows; RT stops in 3-treated samples, angled arrows; full-length RNA, horizontal arrows), thus identifying HAP binding sites. (**D**) Secondary structure of Aptamer 21 with cross-linked nucleotides indicated with arrows (2a, hollow arrows; 3, solid arrows).

thus the protocols established for these methods require validation for mapping small molecule binding sites identified by covalent chemistry. Indeed, we explored how adduct size affects read through of cross-linked nucleotides, where the hypothesis is that small molecule adducts induce mutations.

To find the optimal conditions that enable reverse transcriptase to read through large compound adducts, a panel of reverse transcriptases that have been reported in mutational profiling studies was evaluated, including Induro RT [54], TGIRT III [34], MarathonRT [35], SuperScript II [55], and SuperScript IV [56]. Compound **2a** was selected to assess RT read-through as it labeled Aptamer 21 with TAMRA to the greatest extent and afforded the largest percentage of truncated cDNAs (~50%) without needing to enrich the cross-linked RNA (Figs 2 and 3). The adduct generated by reaction of **2a** with Aptamer 21 is dubbed **2-Ak** (Fig. 4A and Supplementary Fig. S6A). A CE experiment was used to quantify the disappearance of these truncated cDNA products and the concomitant restoration of the full-length cDNA, indicating read-through of the covalent adducts. Upon cross-linking of **2a**, RNA samples were directly subjected to RT using literature-reported mutational conditions for each RT (Supplementary Table S2). The cDNA products generated by each RT were quantified to measure the percent read-through (or full-length product). Significant reduction of truncated cDNAs was observed for Induro RT, MarathonRT, and SuperScript II relative to SuperScript III, indicating the enzymes read through the cross-linked site, but not for TGIRT III and SuperScript IV (Fig. 4B and Supplementary Fig S6B). Quantification of the electrophoresis traces from Induro RT, MarathonRT and SuperScript II confirmed that these enzymes produced 10%–30% more full-length cDNA relative to SuperScript III, with a maximum percentage of full-length product around 75%.

For small molecule probes with lower reactivities or cross-linking efficiencies such as diazirine-based Chem-CLIP probe **3**, an enrichment step may be required to enhance the cross-linking signal. The pull-down procedure results in larger adducts due to the CuAAC reaction, which may further reduce read-through efficiency. We therefore studied how adduct size and structure affect the processivity of Induro RT, MarathonRT, and SuperScript II using the highly reactive probe **2a** to ensure sufficient signal in CE. Three different adducts of increasing size were compared: (i) Aptamer 21–**2a** adduct (no CuAAC reaction; "**2a-Ak**" adduct); (ii) Aptamer 21–**2a** adduct after CuAAC reaction with disulfide biotin

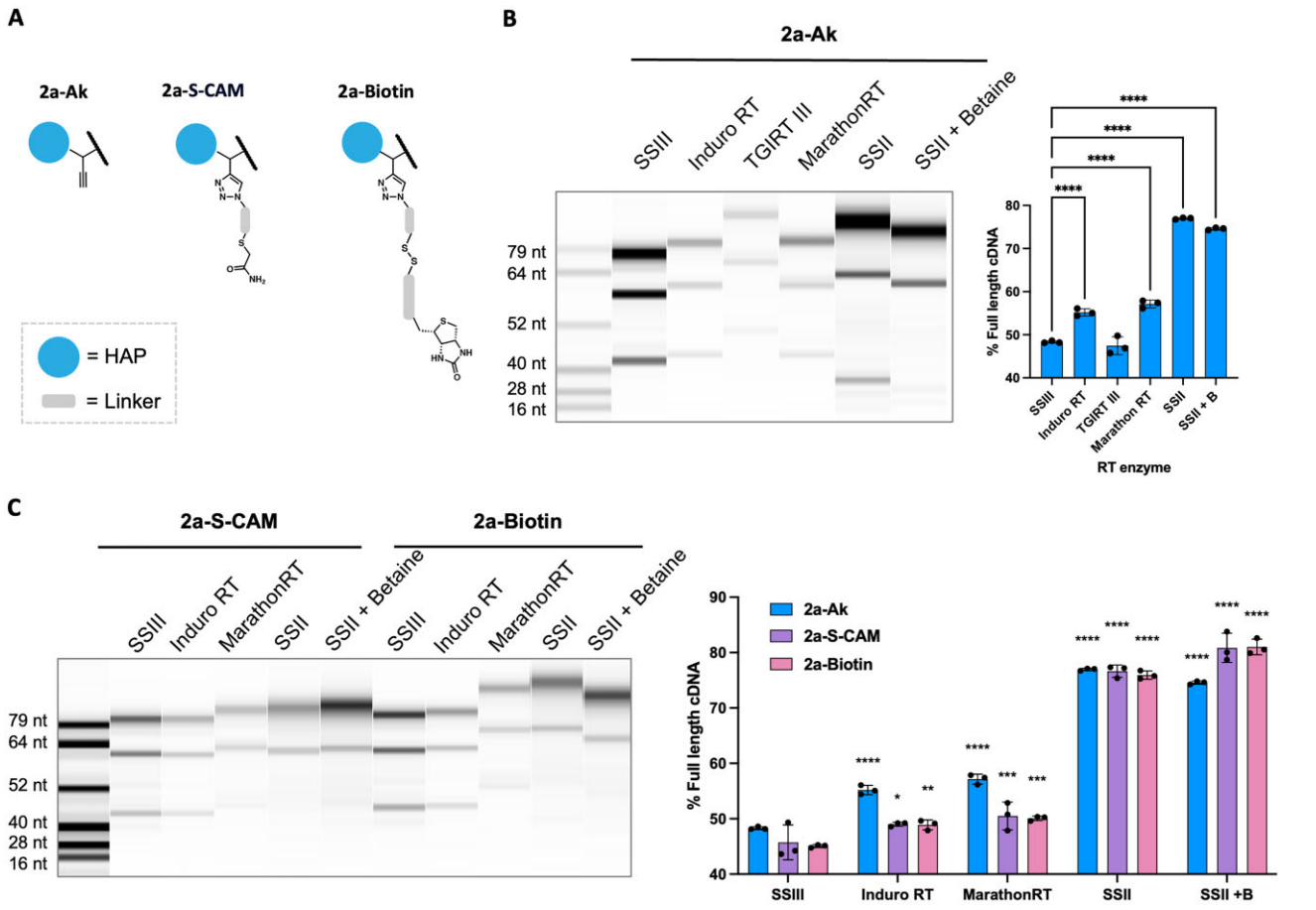


Figure 4. Studying the read-through of various reverse transcriptases to enable binding site mapping by mutational profiling. (**A**) Schematic structures of different modifications of Aptamer 21 after reaction with various Chem-CLIP probes. 2a-Ak: HAP-binding module with an alkyne handle; 2a-S-CAM (HAP-S-carbamidomethyl): HAP alkyne clicked with disulfide biotin azide, cleaved by TCEP and capped with iodoacetamide; 2a-Biotin, HAP alkyne clicked with disulfide biotin azide without cleavage. (**B**) Identification of RT enzymes capable of reading through a compound adduct using capillary electrophoresis. "+B" indicates that Betaine was added to the RT reaction. (**C**) Representative fragment analyzer traces and quantification thereof to assess RT enzyme processivity through adducts of different sizes. "+B" indicates that betaine was added to the RT reaction. $*P < .05$; $**P < .01$; $***P < .001$; and $****P < .0001$, as determined by a one-way Analysis of Variance (ANOVA) with multiple comparisons.

azide, reduction by TCEP and capping with iodoacetamide ("**2a-S-CAM**" adduct); and (iii) Aptamer 21–**2a** adduct after "clicking" on disulfide biotin azide without reduction by TCEP ("**2a-Biotin**" adduct) (Fig. 4A and Supplementary Fig. S6A). [Note: We observed differences in cDNA mobility, likely due to varying salt components derived from the RT buffers (Supplementary Table S2). However, this mobility variation does not interfere with the primary goal of rapidly screening suitable RT enzymes.] The processivity of Induro RT and MarathonRT decreased modestly as the adduct size increased, from 55% ± 1% to 49 ± 2% and from 57 ± 1% to 54 ± 0.4%, respectively (Fig. 4B and C and Supplementary Fig. S6B and C). Interestingly, the percentage full-length cDNA produced by SuperScript II was not affected by the structures of the adducts, ranging from ∼75% to ∼80% for all conditions, rendering it the best RT candidate for the mutational mapping workflow.

In vitro binding site mapping via nanopore sequencing

RNA-sequencing via nanopore sequencing was then used to determine where mutations due to cross-linking occurred within Aptamer 21. Although not applicable to the ∼80 nt long aptamer studied herein, the long-read sequencing capability of nanopore technology enables sequencing of DNA amplicons at any length quickly and cost-effectively, without the need for fragmentation. Thus, the workflow developed herein should be easily adaptable to longer RNAs. Of interest for our study was whether the binding sites determined by assessing RT stops and mutational profiling agree.

After mutational RT by SuperScript II, cDNA amplification via PCR, sequencing of DNA amplicons, and data analysis (Fig. 5A), the mutation rate at each nucleotide position was calculated and compared. We first examined the background mutation rate of unmodified Aptamer 21. Approximately 90% of the bases have mutation rate <2%, while nucleotides 32, 36, 46, 53 have mutation rates around 4% (Supplementary Fig. S7). These mutations could be due to reduced enzyme processivity through structured regions (nt 20–60), intrinsic background mutations of SuperScript II, or errors during nanopore sequencing. Thus, it is essential to first examine the background mutation rate across the reference sample prior to assessing the compound-treated groups in these type of mutational profiling studies. To effec-

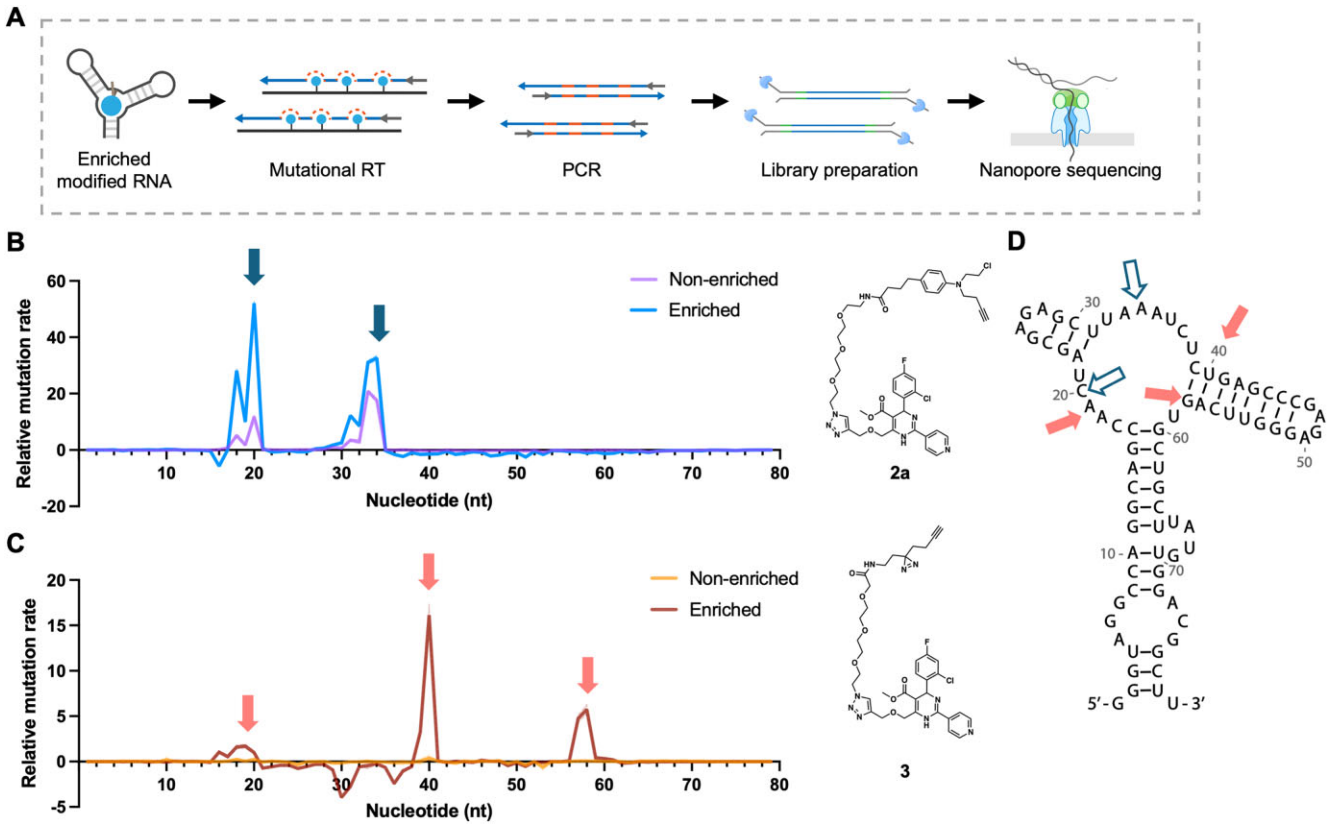


Figure 5. Mutational mapping followed by nanopore sequencing to identify cross-linked sites within an RNA target. (**A**) Scheme of Chem-CLIP-MaP-Seq workflow. (**B**) Representative mutational profile of Aptamer 21 cross-linked to 2a obtained from nanopore sequencing. Relative mutation rate of 2a-treated samples at each nucleotide position was calculated by subtracting the mutational rate observed in 4a-treated samples. (**C**) Representative mutational profile of Aptamer 21 cross-linked to diazirine probe 3 obtained from nanopore sequencing. Relative mutation rate of 3-treated samples at each nucleotide position was calculated by subtracting the mutational rate observed in 5-treated samples. (**D**) Secondary structure of Aptamer 21 overlaid with cross-linked sites for 2a (hollow arrows) and 3 (solid arrows) from mutational profiling studies. Note that affinity enrichment is key for probes in which the extent of cross-linking is < 5%.

tively demonstrate compound-induced mutations, the background mutation rate at each nucleotide was subtracted from compound-treated samples. The unnormalized mutational rate was also analyzed to verify that the subtraction did not bias the analysis (Supplementary Fig. S7).

We subsequently analyzed both probe pairs **2a** and **4a** as well as **3** and **5** without affinity enrichment, denoted as "Non-enriched" samples. Comparison of **2a**- and **4a**-treated samples afforded two regions (C16-C20, and G29-A34) with mutational rates 3σ higher for **2a**-treated samples (determined via a Z-factor test [57]; Fig. 5B and Supplementary Fig. S7A). To account for mutations observed in DMSO- (background) and control probe **4a**-treated samples (nonspecific reaction of the cross-linking module), mutational rates were normalized by subtracting the mutational rates observed at each nucleotide in **4a**-treated samples from the rates observed in **2a**-treated samples. Nucleotides C20 and A34 were considered primary sites of cross-linking as they had the highest mutation rates, 9 ± 0.7%, and 5 ± 0.2%, respectively, compared to other nucleotides within the same region (Fig. 5B and Supplementary Fig. S7A). Moreover, the nucleotides where mutations were observed matched closely with estimates from CE analysis of RT stops, at nucleotides A19 and A34 in Aptamer 21 (Fig. 3B). These data support that the mutations were produced by SuperScript II when encountering small molecule cross-linking. Furthermore, LC-MS analysis of the adducts after digestion to nucleotides revealed only modifications to adenosine and cytidine were formed by **2a** (Supplementary Fig. S4). These results suggest that the mutations observed by sequencing occurred precisely at the site of covalent adduct formation (i.e. not before or after the cross-linked nucleotide). Interestingly, when the same workflow was completed for **3** (non-enriched samples), no mutations were detected (Fig. 5C and Supplementary Fig. S7B).

As **3** had ~5 times lower reactivity in the TAMRA labeling assay and less prominent RT stops than **2a**, we hypothesized that an enrichment step might improve the signal to noise for probes that are less reactive or have lower cross-linking efficiencies. Mutational profiling was therefore completed for both **2a** and **3** and their corresponding control probes (**4a** and **5**, respectively) after adding enrichment steps to the workflow; that is, after cross-linking, the adducts were clicked to disulfide biotin azide, pulled down with streptavidin beads, and eluted for mutational sequencing. To examine whether the enrichment itself affected RNA base integrity and, in turn, the mutational profiling result, the mutation rates of non-enriched and enriched **2a**-treated samples were compared. Significant differences (>3σ) in mutation rates upon **2a**-treatment were only observed nearby the cross-linking site (C17–C20 and G29–A34 near C20 and A34, respectively), with no changes detected in other regions (Fig. 5B and Supplementary Fig. S7A). Therefore, the enrichment steps themselves did not induce mu-

tations, rather the mutations can be traced to cross-linking by the Chem-CLIP probes.

The addition of the enrichment steps indeed improved the detection of mutations induced by the cross-linking of **2a** and **3**. For **2a**, the mutational rate at C20 and A34 increased by approximately three-fold, from 9 ± 0.7% to 45 ± 1% at C20 and from 5 ± 0.2% to 16 ± 0.6% at A34 (Fig. 5B and Supplementary Fig. S7A). For compound **3**, for which mutations could not be detected without enrichment, three regions with mutations were observed compared to **5**-treated samples, and the primary sites were at A18 (4 ± 0.2% mutation rate), U40 (19 ± 1%), and G58 (7 ± 0.5%) (Fig. 5C and Supplementary Fig. S7B), demonstrating the effectiveness and importance of enrichment for weaker RNA binders and/or less reactive probes.

In mutational profiling, an intriguing question is which type of mutations (substitution, deletion, or insertion) are triggered by the small molecule probe. The profiles of each mutation type for **2a** and **3** were examined and compared. Notably, **2a** and **3** induced different types of mutations. For **2a**, substitutions were predominantly observed at the primary sites, C20 and A34, while deletions were observed 5′ to the primary sites, at A18, U31, and U32 (Supplementary Fig. S7C). For **3**, deletions were mainly observed at the mutation sites, at U40, G58, and U59 (Supplementary Fig. S7D). These observations suggested that different cross-linking chemistries and sites uniquely impact enzyme read through, resulting in unique mutational profiles for each small molecule probe.

Comparison of binding site mapping by RT stop analysis and mutational profiling

The RT stop analysis by CE and mutational profiling for **2a** are in overall agreement where both methods identified two positions where cross-linking occurred. Both methods identified A34 while RT stop analysis identified A19 and mutational profiling identified C20 as the second position of cross-linking (Fig. 5D). For **3**, differences were observed when using the two methods. In CE analysis, RT stops were observed at nt C20 and U38 while mutations were observed at nt A19, U40, and G58 in nanopore sequencing analysis; that is two common binding sites were identified at nt 19/20 and 38/40, although with different extents of RT stop and mutation rates. The lack of identifiable RT stop at nt G58 in the CE method may be explained by the poor resolution of oligonucleotides <30 nt long in the electropherograms, given that the expected length of the RT product (21 nt) may not separate well from the primer. Moreover, the mutational profile for **3** is largely consistent with the cross-linked sites identified by Mukherjee *et al*. using an RT stop method with a closely related diazirine probe [29]. The general agreement between CE and Nanopore sequencing results reinforces the validity and effectiveness of both methods in identifying cross-linking sites.

Although bearing the same RNA binding module, **2a** and **3** had distinct cross-linking sites identified from both RT stop and mutational profiling. We reasoned that perhaps the slight differences in the PEG linker and cross-linker structures could lead to a shift in probe positioning, affecting the set of reacted bases. Nonetheless, the HAP binding site was mapped to the multibranch loop of Aptamer 21 for both probes (C20 and A24 for **2a** and A19, U40 and G58 for **3**) and by both methodologies (Figs 3B and C and 5B and C). The mutational profiling data could be integrated with computational approaches, including molecular docking and molecular dynamics simulations, to elucidate the precise binding interactions of each probe.

In general, a lower mutation rate was observed for probe **2a** and **3** than RT stop rate obtained from CE analysis. For example, the percentage mutation rate at C20 from nanopore sequencing was 12 ± 5%, whereas the percentage truncated cDNAs for **2a** estimated at A19 on Aptamer 21 was 36 ± 0.3% (Figs 3B and 5B). One contributing factor is the incomplete read-through of cross-linked sites by Superscript II where read-through percentage was ∼50%–65%. Additionally, there is a possibility that the RT enzyme could read through a modified base without introducing a mutation, possibly influenced by the location of cross-link within the nucleotide (Watson–Crick versus Hoogsteen face, for example).

Overall, we observed consistency between the mutational profile of probes **2a** and **3** with the corresponding RT stop analysis. Notably, the RT stops identified from CE were within ±1 nucleotide from the binding sites identified by mutational mapping. The high sensitivity of the fragment analyzer and the use of a custom cDNA ladder contributed to the observed accuracy. Collectively, these data demonstrated the feasibility and effectiveness of using Chem-CLIP, mutational mapping, and nanopore sequencing approaches (Chem-CLIP-MaP-Seq) to identify compound binding site for small molecule probes with different levels of reactivity.

Direct RNA sequencing of Aptamer 21–2a using nanopore technology

One advancement of nanopore technology is its ability to sequence native RNA molecules through RNA-specific flow cells, termed direct RNA sequencing. We thus thought to investigate how cross-linked RNAs behave in the direct RNA sequencing experiment. Libraries containing Aptamer 21 modified with **2a** ("**2a-Ak**" adduct) or DMSO as a control were prepared and sequenced without the enrichment steps. To evaluate the effects of cross-linking, changes in read coverage, mutation profiles, and raw electronic signals were assessed (Supplementary Fig. S8A). A drop of read coverage >10% at the 5′ end of the RNA was observed when comparing **2a**-treated samples with DMSO, indicating premature read termination or read stops, while no significant change in mutational profile or raw signals was observed between two groups. To quantify read stops, the beginning position of each aligned read (indicates a stop) were cumulated, affording two major sites of read stop at nucleotides A22, and around U31 and A33 (Supplementary Fig. S8B), which are in close proximity to the cross-linked site of **2a** within Aptamer 21. We reasoned that since the adduct **2a-Ak** is relatively large that the RNA might be stalled at the cross-linked site when passing through the nanopore. It is possible that smaller adducts would not cause this read stop and in turn impact the raw signal or the base-calling accuracy of the RNA, as demonstrated for DNA modifications [58].

Discussion

In this study, we described the development and implementation of a platform to identify small molecule binding sites within an RNA target by using covalent chemistry. These methods include a fluorescent labeling assay to assess formation of a covalent bond between a reactive probe and an

RNA, indicative of binding, an RT stop assay using capillary electrophoresis to estimate the location of the binding site, and mutational profiling via nanopore sequencing to identify cross-linked sites at single-nucleotide resolution (Fig. 6). These methods can be applied to the target validation of any type of Chem-CLIP probe, i.e. a compound containing an RNA binder, a reactive module, and a purification tag, and can be used as a screening platform for identifying selective binders of structured RNAs or novel chemical moieties that react with RNA.

We observed consistency in the extent of reactivity of the probes in the assays that were developed. For example, the *N*-chloroethyl anilines were ∼3–6-fold more reactive than the diazirine probe, as determined from both TAMRA labeling and RT stop analysis. For mutational mapping, the mutation rate of *N*-chloroethyl aniline **2a** at the major site of cross-linking was around 12% without affinity enrichment, while no mutations were observed for the diazirine probe under the same conditions. We estimate that probes that label >5% of their RNA target (as determined by RT stop analysis) might be suitable for mutational profiling without the need for affinity enrichment.

The low reactivity of the diazirine probe is perhaps unsurprising as quenching of the diazo or carbene species by water is unavoidable in aqueous solution [44]. In the case of Aptamer 21 and **3**, the flexible PEG linker between the diazirine and the HAP binder might compound the effect of quenching by orienting the diazirine away from the RNA target and into solvent. Rational design of the linker using molecular docking assisted drug design to position the diazirine near a favorable substrate could increase the reactivity of this probe. Despite their moderate reactivity, diazirines have high specificity when conjugated to an RNA binding module, showing minimal cross-linking to a low affinity RNA mutant in the TAMRA labeling assay. Thus, diazirines can be used for cellular target identification and selectivity studies transcriptome- and proteome-wide [22, 24].

In contrast, the chlorambucil derivatives of HAP exhibited high reactivity with minimal nonspecific cross-linking in the absence of the RNA-binding component. Although one concern with chlorambucil-based probes is nonspecific DNA or RNA alkylation, these studies show that appending a potent and specific RNA-binding small molecule to the *N*-chloroethyl aniline can result in significant target selectivity. Conversely, the reactivity and selectivity of small molecule probes to low affinity RNA targets significantly decline, as demonstrated by **2a** and **4a** and control RNA 21-R. Such a phenomenon has also been observed with other warheads, particularly bleomycin where conjugation of the natural product with a selective RNA binder attenuates its activity against DNA [16]. Moreover, the sites where cross-linking occurred could be determined by both RT stop analysis as determined from CE and from mutational profiling by nanopore sequencing without the need for enrichment, in contrast to diazirine-based probes. Collectively, these data suggest that *N*-chloroethyl anilines are superior to diazirines, traditionally used in other methods [16, 18, 20, 21, 29], to map the binding sites of small molecules.

The high reactivity of the chlorambucil-based probes enabled us to examine a panel of reverse transcriptases for mutational profiling. Although the field of RNA structural probing is moving away from the use of SuperScript II due to its relatively high background mutation and drop-off rates for RNA > 500 nucleotides [34, 56], this work demonstrated that large adducts to nucleobases may require Super Script II for effective read-through. It is reasonable to infer that the behavior of each enzyme towards adducts at different positions on the RNA (e.g. 2′OH versus nucleobase), as well as adducts varying in length, molecular weight, and rigidity could be distinct; thus, it is important to evaluate different RT enzymes when working with various adducts. Alternatively, installing a cleavable linker between the RNA binder and the cross-linking module in the Chem-CLIP probe could reduce the complexity of the adduct structure, enabling read through by more enzymes and expanding the available options for mutational profiling. To implement the current workflow on cellular RNAs or those exceeding 500 nt in length, whose analysis might be affected by the reduced processivity of Superscript II, mild mRNA fractionation, polyadenylation, and oligo-dT priming followed by either next-generation sequencing (NGS) or nanopore sequencing could help overcome this limitation. Another option would be to combine the RT stop site mapping strategies developed in Chem-CLIP-Map-Seq with nanopore sequencing to achieve full-length RNA read coverage (see below).

This study utilized nanopore sequencing to detect sites within an RNA target that were cross-linked to a small molecule. Complementary to NGS studies, which features short-read sequencing, nanopore sequencing reads native RNA or DNA amplicons in its original length, ranging from 20 bp to >10 kb [59]. This property is essential for effective sequencing of DNA amplicons of varying size, which is commonly produced in *in vitro* and in cell gene-specific mutational profiling assays. Moreover, it could be beneficial for the downstream application of Chem-CLIP-MaP-Seq, allowing for isoform-specific small molecule binding site identification in cells, as implicated in the Nano-DMS-MaP-Seq methods, as well as direct identification of multiple binding sites on a single RNA molecule, bypassing the need for deconvolution from shorter reads [52]. In addition, a MinION nanopore sequencer priced at ∼$2000 and operable from a standard laptop democratizes advanced sequencing methods when access to more expensive platforms is limited or cost-prohibitive. While nanopore sequencing is known to have a higher error rate than other technologies, the methods used in this study demonstrate sufficient robustness to address these limitations effectively with the proper experimental controls. Although not fully explored in this report, ongoing work aims to leverage these capabilities to further minimize artifacts introduced during sample preparation. Direct RNA sequencing is another attractive feature enabled by the nanopore technology. Although our attempt to sequence the native cross-linked RNA resulted in read stops, it is possible that distinctive raw electronic signal at cross-linked sites can be identified for less complicated adducts, which could provide a new methodology for binding site mapping of RNA-targeted small molecules. Collectively, these characteristics position the platform described herein as a promising tool for advancing transcriptome-wide studies of RNA-ligand interactions in cellular systems, offering broader potential applications in RNA-targeting research.

While both CE and nanopore sequencing developed in this study are effective for evaluating small molecule binding sites, they offer different advantages. CE provides a simple and sensitive approach to rapidly assess the site of cross-linking with relatively high accuracy, generally within ±1 nt of the binding site, as confirmed by mutational profiling. As the resolution between 1 and 30 nt on CE is low, binding sites located at the

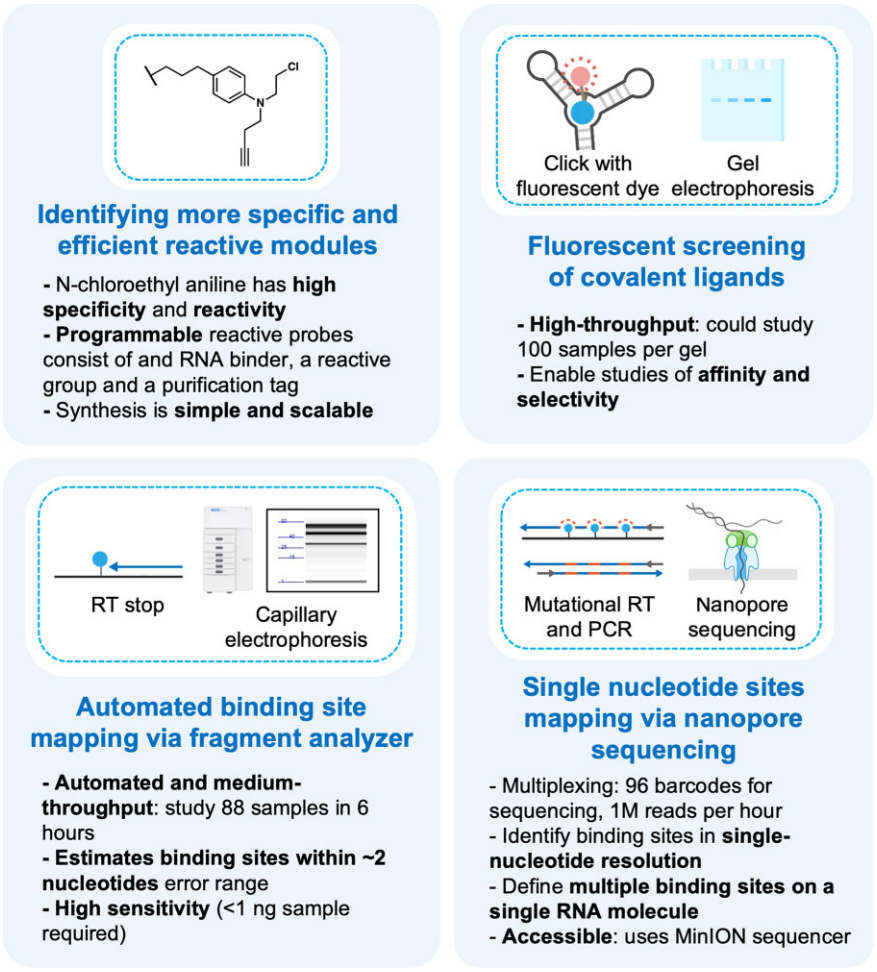


Figure 6. A general platform for RNA binder identification and target validation. *N*-chloroethyl aniline was identified to exhibit high reactivity and selectivity toward Aptamer 21. Fluorescence-based reactivity screening allows rapid identification of covalent probes reactive with RNA. Capillary electrophoresis analysis via fragment analyzer enabled automated binding site identification with high accuracy and sensitivity. Mutational mapping using nanopore sequencing identifies cross-linked sites at single-nucleotide resolution cost effectively.

3′ end of the RNA may not be captured effectively. This issue may be circumvented by increasing the length of cDNA products, using RT primers with additional adaptor sequences on the 5′ end.

This study also explored an alternative approach to mapping small molecule–RNA binding sites, mutational profiling, which offers several advantages compared to traditional RT stop-based methods and other established techniques [16, 18, 20, 21, 26, 29]. RT stops can arise from artifacts such as RNA hydrolysis or regions of high structural stability that block RT, leading to false positives for cross-linked sites. In contrast, mutational mapping mitigates these issues, offering a more accurate and reliable representation of small molecule–RNA interactions. Indeed, Compared to CE, nanopore sequencing requires a lengthier preparation process, but provides significantly more information, including the mutation rate at each base, the types of mutations (substitution, deletion, and insertion), a single-molecule mutational profile (i.e. detect several cross-linking events on one molecule of RNA), and can be applied to both *in vitro* and cellular studies of RNAs of varying length.

Overall, the methods described herein for assessing RNA–small molecule binding sites provide a programmable platform for future discovery and validation of RNA–small molecule interactions. Appending reactive modules (with functional handles) to potential RNA-binding compounds enables a myriad of techniques to validate target engagement and define binding sites within target RNAs. Fluorescence labeling via CuAAC reaction allows for rapid and high-throughput identification of novel RNA binders or covalent probes. Capillary electrophoresis of RT stops enables a medium throughput and quantitative assessment of RNA–small molecule binding sites with limited RNA input. Nanopore sequencing provides a cost-effective method for assessing binding sites of small molecule probes with single-nucleotide resolution for transcripts of (nearly) any length. Although this work focused on *in vitro* validation of a known aptamer-small molecule pair, future work may apply these methods to assess target engagement and binding in cells or for discovery of novel small molecule–RNA interactions.

Acknowledgements

Author contributions. all authors contributed to writing review & editing, M.D.D. conceptualization, funding acquisition, project administration writing - original draft. X.Y. methodology, data curation, formal analysis, resources, writing - original draft. J.W. data curation, formal analysis, re-

sources,. N.A.S. data curation, formal analysis, resources. P.R.A.Z resources. Y.J. and X.S. resources. We thank J. Childs-Disney for advice and critical review of the manuscript.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

M.D.D. is a co-founder of Expansion Therapeutics and Ribonaut Therapeutics.

Funding

This work was supported by the National Institutes of Health (R35 NS116846 and R01 CA249180 to M.D.D) and by the German Research Foundation (DFG) through a Walter Benjamin fellowship to P.R.A.Z. Funding to pay the Open Access publication charges for this article was provided by National Institutes of Health (R35 NS116846 and R01 CA249180).

Data availability

The data underlying this article are available in the article and in its online supplementary material. Sequencing files were deposited in Mendeley, available at: Reserved https://doi.org/10.17632/rszx769fn9.2.

References

1. Childs-Disney JL, Yang X, Gibaut QMR *et al.* Targeting RNA structures with small molecules. *Nat Rev Drug Discov* 2022;**21**:736–62. https://doi.org/10.1038/s41573-022-00521-4
2. Garner AL. Contemporary progress and opportunities in RNA-targeted drug discovery. *ACS Med Chem Lett* 2023;**14**:251–9. https://doi.org/10.1021/acsmedchemlett.3c00020
3. Dhuri K, Bechtold C, Quijano E *et al.* Antisense oligonucleotides: an emerging area in drug discovery and development. *J Clin Med* 2020;**9**:2004. https://doi.org/10.3390/jcm9062004
4. Ding Y, Tang Y, Kwok CK *et al.* *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 2014;**505**:696–700. https://doi.org/10.1038/nature12756
5. Rouskin S, Zubradt M, Washietl S *et al.* Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* 2014;**505**:701–5. https://doi.org/10.1038/nature12894
6. Kaur J, Sharma A, Mundlia P *et al.* RNA–small-molecule interaction: challenging the "undruggable" tag. *J Med Chem* 2024;**67**:4259–97. https://doi.org/10.1021/acs.jmedchem.3c01354
7. Bush JA, Aikawa H, Fuerst R *et al.* Ribonuclease recruitment using a small molecule reduced c9ALS/FTD r(G_4C_2) repeat expansion *in vitro* and *in vivo* ALS models. *Sci Transl Med* 2021;**13**:eabd5991. https://doi.org/10.1126/scitranslmed.abd5991
8. Chen JL, Zhang P, Abe M *et al.* Design, optimization, and study of small molecules that target Tau pre-mRNA and affect splicing. *J Am Chem Soc* 2020;**142**:8706–27. https://doi.org/10.1021/jacs.0c00768
9. Wang J, Schultz PG, Johnson KA. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc Natl Acad Sci USA* 2018;**115**:4604–12. https://doi.org/10.1073/pnas.1800260115
10. Naryshkin NA, Weetall M, Dakka A *et al.* SMN2 splicing modifiers improve motor function and longevity in mice with spinal muscular atrophy. *Science* 2014;**345**:688–93. https://doi.org/10.1126/science.1250127
11. Keller CG, Shin Y, Monteys AM *et al.* An orally available, brain penetrant, small molecule lowers huntingtin levels by enhancing pseudoexon inclusion. *Nat Commun* 2022;**13**:1150. https://doi.org/10.1038/s41467-022-28653-6
12. Mikutis S, Rebelo M, Yankova E *et al.* Proximity-induced nucleic acid degrader (PINAD) approach to targeted RNA degradation using small molecules. *ACS Cent Sci* 2023;**9**:892–904. https://doi.org/10.1021/acscentsci.3c00015
13. Li Y, Disney MD. Precise small molecule degradation of a noncoding RNA identifies cellular binding sites and modulates an oncogenic phenotype. *ACS Chem Biol* 2018;**13**:3065–71. https://doi.org/10.1021/acschembio.8b00827
14. Childs-Disney JL, Tran T, Vummidi BR *et al.* A massively parallel selection of small molecule–RNA motif binding partners informs design of an antiviral from sequence. *Chem* 2018;**4**:2384–404. https://doi.org/10.1016/j.chempr.2018.08.003
15. Angelbello AJ, Rzuczek SG, McKee KK *et al.* Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proc Natl Acad Sci USA* 2019;**116**:7799–804. https://doi.org/10.1073/pnas.1901484116
16. Rzuczek SG, Colgan LA, Nakai Y *et al.* Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat Chem Biol* 2017;**13**:188–93. https://doi.org/10.1038/nchembio.2251
17. Hostetter AA, Osborn MF, DeRose VJ. RNA-Pt adducts following cisplatin treatment of *Saccharomyces cerevisiae*. *ACS Chem Biol* 2012;**7**:218–25. https://doi.org/10.1021/cb200279p
18. Velagapudi SP, Li Y, Disney MD. A cross-linking approach to map small molecule–RNA binding sites in cells. *Bioorg Med Chem Lett* 2019;**29**:1532–6. https://doi.org/10.1016/j.bmcl.2019.04.001
19. Moon MH, Vock IW, Streit AD *et al.* Disulfide tethering to map small molecule binding sites transcriptome-wide. *ACS Chem Biol* 2024;**19**:2081–6. https://doi.org/10.1021/acschembio.4c00538
20. Guan L, Disney MD. Covalent small-molecule–RNA complex formation enables cellular profiling of small-molecule–RNA interactions. *Angew Chem Int Ed* 2013;**52**:10010–3. https://doi.org/10.1002/anie.201301639
21. Yang WY, Wilson HD, Velagapudi SP *et al.* Inhibition of non-ATG translational events in cells via covalent small molecules targeting RNA. *J Am Chem Soc* 2015;**137**:5336–45. https://doi.org/10.1021/ja507448y
22. Cisar JS, Cravatt BF. Fully functionalized small-molecule probes for integrated phenotypic screening and target identification. *J Am Chem Soc* 2012;**134**:10385–8. https://doi.org/10.1021/ja304213w
23. Parker CG, Galmozzi A, Wang Y *et al.* Ligand and target discovery by fragment-based screening in human cells. *Cell* 2017;**168**:527–41. https://doi.org/10.1016/j.cell.2016.12.029
24. Tong Y, Gibaut QMR, Rouse W *et al.* Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J Am Chem Soc* 2022;**144**:11620–5. https://doi.org/10.1021/jacs.2c01929
25. Balaratnam S, Rhodes C, Bume DD *et al.* A chemical probe based on the PreQ1 metabolite enables transcriptome-wide mapping of binding sites. *Nat Commun* 2021;**12**:5856. https://doi.org/10.1038/s41467-021-25973-x
26. Tong Y, Su X, Rouse W *et al.* Transcriptome-wide, unbiased profiling of ribonuclease targeting chimeras. *J Am Chem Soc* 2024;**146**:21525–34. https://doi.org/10.1021/jacs.4c04717
27. Tornøe CW, Christensen C, Meldal M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem* 2002;**67**:3057–64. https://doi.org/10.1021/jo011148j
28. Rostovtsev VV, Green LG, Fokin VV *et al.* A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* 2002;**41**:2596–9. https://doi.org/10.1002/1521-3773(20020715)41:14⟨2596::Aid-anie2596⟩3.0.Co;2-4
29. Mukherjee H, Blain JC, Vandivier LE *et al.* PEARL-seq: a photoaffinity platform for the analysis of small molecule–RNA

interactions. *ACS Chem Biol* 2020;**15**:2374–81. https://doi.org/10.1021/acschembio.0c00357
30. Lau JL, Baksh MM, Fiedler JD *et al.* Evolution and protein packaging of small-molecule RNA aptamers. *ACS Nano* 2011;**5**:7722–9. https://doi.org/10.1021/nn2006927
31. Nilsen TW, Rio DC, Ares M Jr. High-yield synthesis of RNA using T7 RNA polymerase and plasmid DNA or oligonucleotide templates. *Cold Spring Harb Protoc* 2013;**2013**:pdb.prot078535. https://doi.org/10.1101/pdb.prot078535
32. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;**9**:671–5. https://doi.org/10.1038/nmeth.2089
33. Weidmann CA, Mustoe AM, Jariwala PB *et al.* Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA. *Nat Biotechnol* 2021;**39**:347–56. https://doi.org/10.1038/s41587-020-0709-7
34. Zubradt M, Gupta P, Persad S *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing *in vivo*. *Nat Methods* 2017;**14**:75–82. https://doi.org/10.1038/nmeth.4057
35. Guo L-T, Adams RL, Wan H *et al.* Sequencing and structure probing of long RNAs using MarathonRT: a next-generation reverse transcriptase. *J Mol Biol* 2020;**432**:3338–52. https://doi.org/10.1016/j.jmb.2020.03.022
36. Houston L, Pike W. *Analyzing Poly(A) Tails of In Vitro Transcribed RNA with the Agilent Fragment Analyzer System*. Santa Clara, California. United States of America: Agilent Technologies, Inc., 2022.
37. Stray SJ, Bourne CR, Punna S *et al.* A heteroaryldihydropyrimidine activates and can misdirect hepatitis B virus capsid assembly. *Proc Natl Acad Sci USA* 2005;**102**:8138–43. https://doi.org/10.1073/pnas.0409732102
38. Fukunaga K, Dhamodharan V, Miyahira N *et al.* Small-molecule aptamer for regulating RNA functions in mammalian cells and animals. *J Am Chem Soc* 2023;**145**:7820–8. https://doi.org/10.1021/jacs.2c12332
39. Di Antonio M, McLuckie KIE, Balasubramanian S. Reprogramming the mechanism of action of chlorambucil by coupling to a G-quadruplex ligand. *J Am Chem Soc* 2014;**136**:5860–3. https://doi.org/10.1021/ja5014344
40. Kundu GC, Schullek JR, Wilson IB. The alkylating properties of chlorambucil. *Pharmacol Biochem Behav* 1994;**49**:621–4. https://doi.org/10.1016/0091-3057(94)90078-7
41. Mattes WB, Hartley JA, Kohn KW. DNA sequence selectivity of guanine–N7 alkylation by nitrogen mustards. *Nucl Acids Res* 1986;**14**:2971–87. https://doi.org/10.1093/nar/14.7.2971
42. Hartley JA, Bingham JP, Souhami RL. DNA sequence selectivity of guanine-N7 alkylation by nitrogen mustards is preserved in intact cells. *Nucl Acids Res* 1992;**20**:3175–8. https://doi.org/10.1093/nar/20.12.3175
43. Kohn KW, Spears CL, Doty P. Inter-strand crosslinking of DNA by nitrogen mustard. *J Mol Biol* 1966;**19**:266–88. https://doi.org/10.1016/S0022-2836(66)80004-9
44. Jiang Y, Zhang X, Nie H *et al.* Dissecting diazirine photo-reaction mechanism for protein residue-specific cross-linking and distance mapping. *Nat Commun* 2024;**15**:6060. https://doi.org/10.1038/s41467-024-50315-y
45. Suresh BM, Li W, Zhang P *et al.* A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc Natl Acad Sci USA* 2020;**117**:33197–203. https://doi.org/10.1073/pnas.2012217117
46. Brunner J, Senn H, Richards FM. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J Biol Chem* 1980;**255**:3313–8. https://doi.org/10.1016/S0021-9258(19)85701-0
47. Farley SE, Hashimoto R, Evangelista J *et al.* Trifunctional fatty acid derivatives: the impact of diazirine placement. *Chem Commun* 2024;**60**:6651–4. https://doi.org/10.1039/d4cc00974f
48. Florea-Wang D, Pawlowicz AJ, Sinkkonen J *et al.* Reactions of 4-[*bis*(2-chloroethyl)amino]benzenebutanoic acid (chlorambucil) with DNA. *Chem Biodivers* 2009;**6**:1002–13. https://doi.org/10.1002/cbdv.200800327
49. Inoue T, Cech TR. Secondary structure of the circular form of the tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA* 1985;**82**:648–52. https://doi.org/10.1073/pnas.82.3.648
50. Stern S, Wilson RC, Noller HF. Localization of the binding site for protein S4 on 16 S ribosomal RNA by chemical and enzymatic probing and primer extension. *J Mol Biol* 1986;**192**:101–10. https://doi.org/10.1016/0022-2836(86)90467-5
51. Technologies A. *Agilent DNF-470 Small RNA Kit Quick Guide*. Santa Clara, California. United States of America: Agilent Technologies Inc., 2024.
52. Bohn P, Gribling-Burrer A-S, Ambi UB *et al.* Nano-DMS-MaP allows isoform-specific RNA structure determination. *Nat Methods* 2023;**20**:849–59. https://doi.org/10.1038/s41592-023-01862-7
53. Siegfried NA, Busan S, Rice GM *et al.* RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 2014;**11**:959–65. https://doi.org/10.1038/nmeth.3029
54. Romero-Agosto G, Cox E, Rouskin S. DMS-MapSeq analysis of antisense oligonucleotide binding to lncRNA PANDA. *Curr Protoc* 2024;**4**:70038. https://doi.org/10.1002/cpz1.70038
55. Homan PJ, Favorov OV, Lavender CA *et al.* Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci USA* 2014;**111**:13858–63. https://doi.org/10.1073/pnas.1407306111
56. Grzywacz K, Chełkowska-Pauszek A, Plucinska-Jankowska M *et al.* The Evaluation of SHAPE-MaP RNA structure probing protocols reveals a novel role of Mn2+ in the detection of 2′-OH Adducts. *Int J Mol Sci* 2023;**24**:7890. https://doi.org/10.3390/ijms24097890
57. Zhang J-H, Chung TDY, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 1999;**4**:67–73. https://doi.org/10.1177/108705719900400206
58. Georgieva D, Liu Q, Wang K *et al.* Detection of base analogs incorporated during DNA replication by nanopore sequencing. *Nucleic Acids Res* 2020;**48**:e88. https://doi.org/10.1093/nar/gkaa517
59. Wang Y, Zhao Y, Bollas A *et al.* Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021;**39**:1348–65. https://doi.org/10.1038/s41587-021-01108-x

Received: December 3, 2024. **Revised:** March 5, 2025. **Editorial Decision:** March 6, 2025. **Accepted:** March 12, 2025