




A chemical probe based on the PreQ₁ metabolite enables transcriptome-wide mapping of binding sites

Sumirtha Balaratnam^{1,7}, Curran Rhodes^{1,7}, Desta Doro Bume¹, Colleen Connelly ¹, Christopher C. Lai¹, James A. Kelley¹, Kamyar Yazdani¹, Philip J. Homan^{2,3}, Danny Incarnato ⁴, Tomoyuki Numata^{5,6} & John S. Schneekloth Jr ¹✉

The role of metabolite-responsive riboswitches in regulating gene expression in bacteria is well known and makes them useful systems for the study of RNA-small molecule interactions. Here, we study the PreQ₁ riboswitch system, assessing sixteen diverse PreQ₁-derived probes for their ability to selectively modify the class-I PreQ₁ riboswitch aptamer covalently. For the most active probe (**11**), a diazirine-based photocrosslinking analog of PreQ₁, X-ray crystallography and gel-based competition assays demonstrated the mode of binding of the ligand to the aptamer, and functional assays demonstrated that the probe retains activity against the full riboswitch. Transcriptome-wide mapping using Chem-CLIP revealed a highly selective interaction between the bacterial aptamer and the probe. In addition, a small number of RNA targets in endogenous human transcripts were found to bind specifically to **11**, providing evidence for candidate PreQ₁ aptamers in human RNA. This work demonstrates a stark influence of linker chemistry and structure on the ability of molecules to crosslink RNA, reveals that the PreQ₁ aptamer/ligand pair are broadly useful for chemical biology applications, and provides insights into how PreQ₁, which is similar in structure to guanine, interacts with human RNAs.

¹Chemical Biology Laboratory, National Cancer Institute, Frederick, MD 21702, USA. ²Center for Cancer Research Collaborative Bioinformatics Resource, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. ³Advanced Biomedical Computational Science, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA. ⁴Department of Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Groningen, The Netherlands. ⁵Department of Bioscience and Biotechnology, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, 744 Motooka, Nishi-kuFukuoka-shi Fukuoka 812-8582, Japan. ⁶Biomedical Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba-shi, Ibaraki 305-8566, Japan. ⁷These authors contributed equally: Sumirtha Balaratnam, Curran Rhodes. ✉email: schneeklothjs@mail.nih.gov

Riboswitches are naturally occurring RNA sequences that influence bacterial gene expression by binding directly to small molecules^{1–3}. As they can bind specifically and selectively to small molecules, these important functional elements provide invaluable systems to study RNA-small molecule interactions^{4–6}. One such system is the PreQ₁ riboswitch, which binds to the small, modified nucleotide 7-aminomethyl-7-deazaguanine (known as PreQ₁). PreQ₁ riboswitches have been divided into three classes based on differences in function and size. Upon recognition of PreQ₁, the RNA changes conformation and alters gene expression at either transcriptional level or translational level^{7,8}. This system serves as a metabolic feedback sensor for PreQ₁ levels, allowing bacteria to control the expression of genes involved in one carbon metabolism in response to metabolite levels. While many bacteria contain diverse PreQ₁-responsive riboswitches in mRNAs, human mRNAs appear to be devoid of analogous elements⁹. To date, the study of PreQ₁ in human transcriptomes has mostly been limited to covalent modification of tRNAs by queuosine (of which PreQ₁ is a metabolic precursor in bacteria)¹⁰. PreQ₁ has also been utilized in a series of intriguing studies aimed at enzymatic tagging of complex RNAs^{11,12}. Owing to the high selectivity and affinity of the PreQ₁/aptamer interaction, as well as ease of chemical modification, it could find broad use in chemical biology and synthetic biology applications¹³. Additionally, PreQ₁ is structurally very similar to guanine and thus a PreQ₁-derived probe could provide insights into how this and other similar nucleobases recognize human RNA structures. Thus, a broad assessment of PreQ₁ binding in human transcriptomes would not only clarify the selectivity of the PreQ₁ aptamer interaction but also enable an unbiased examination of metabolite binding to human RNAs.

Reactive molecules that covalently modify RNAs have played a substantial role in understanding RNA biology by probing structure^{14,15}, controlling gene expression¹⁶, imaging RNAs^{17,18}, tagging RNAs with functional handles, and demonstrating target engagement for RNA-binding small molecules^{19–22}. However, examples of structurally characterized RNA/ligand complexes are relatively rare. Still, highly specific RNA/ligand pairs can find broad use in diverse applications. As interest in RNA as a target for small molecule drugs has increased, so has the need for information about the design of RNA-targeting probes^{23–25}. Understanding the chemical features that impact RNA crosslinking efficiency would influence probe design, as well as inform interpretation of results from experiments that use these probes. To date, there has been limited study on the factors that govern covalent crosslinking efficiency to RNA. Furthermore, the impact of tagging small ligands with reactive handles could be substantial. Aptamers can discriminate between highly similar ligands, even distinguishing fluoride from chloride²⁶, potentially representing a challenge in probe design. Drawing inspiration from the approaches above, we set out to design covalent probes targeting a PreQ₁ riboswitch, and to apply them for a broad assessment of binding in complex systems.

The class-I PreQ₁ riboswitch aptamers (PreQ₁-I) represent an ideal system for this study due to their relatively small size and its conserved ligand binding sites. For example, the PreQ₁ riboswitch from *Bacillus subtilis* (Bs-preQ₁-RS) consists of just 34 nucleotides. Upon binding to PreQ₁, it folds into an H-type pseudoknot; this conformational change results in the transcriptional down-regulation of *queCDEF* operon^{7,27,28}. Similarly, *Staphylococcus saprophyticus* (Ss-preQ₁-RS) carries a related class-I PreQ₁ riboswitch that also functions by transcriptional termination²⁹. For the PreQ₁ riboswitch from *Thermoanaerobacter tengcongensis* (Tt-preQ₁-RS) binding to PreQ₁ also results in the formation of an H-type pseudoknot structure. However, in this case the riboswitch acts at the translational level by regulating the

accessibility of the Shine-Dalgarno sequence and ribosome binding^{8,30,31}. While the mechanism and sequences of these PreQ₁-I class riboswitches differ they all have tight and selective interactions with the PreQ₁ ligand. For example, the Tt-preQ₁-RS has an equilibrium dissociation constant of 2 nM for PreQ₁³⁰ and the Bs-preQ₁-RS constant has been reported as 20 nM from in-ligand probing⁸. In addition to this tight binding, these aptamers have been studied extensively using both X-ray crystallography and NMR, providing detailed information about the mode of ligand binding^{29–34}. Other studies on this system include how it responds to other natural metabolites, engineering to remodel ligand specificity^{8,13,34}, and investigations for inhibitor design²⁹. PreQ₁ riboswitches from other organisms have been studied in detail as well, revealing multiple distinct classes of independently evolved riboswitches that function by different mechanisms⁷. In fact, PreQ₁ riboswitches are among the most frequently occurring riboswitches in Nature³⁵. This example of convergent evolution demonstrates a remarkably high diversity of RNA sequences that recognize PreQ₁ with high affinity and selectivity. Given the extensive work reported on the PreQ₁ riboswitches, its potential for use in chemical biology and synthetic biology applications, the role of PreQ₁ in influencing gene expression, the frequency of natural aptamers for PreQ₁, and the similarity of PreQ₁ to guanine, efforts to probe the binding of this metabolite to human RNAs are warranted.

Here, we report the design and evaluation of a series of covalent probes for the PreQ₁-I class riboswitch. In total, a series of 16 probes are reported, including both electrophilic and photochemically activated crosslinkers. Since the PreQ₁-I class riboswitches have a small aptamer domain and have conservation of PreQ₁ binding, we utilize multiple PreQ₁-I class aptamers (Bs, Ss, Tt) from different species in this study to ensure that the best probes could be broadly recognized by these riboswitches. Gel-based and mass spectrometry experiments are conducted to quantify crosslinking efficiency and characterize the covalent adduct. We report a remarkable effect of linker length and crosslinker structure on crosslinking efficiency. None of the electrophilic probes showed evidence of covalent modification, however photocrosslinking probes had varying efficiency. One of these probes was selected for further in-depth characterization. The probe was found to photocrosslink only to the bases of the aptamer, not sugar or phosphate moieties. Additionally, the photocrosslinking probe retained activity in a transcriptional termination assay. An X-ray co-crystal structure of the probe in complex with the aptamer demonstrated that the mode of binding was unperturbed by the introduction of the crosslinker sidechain. We also demonstrate the potential utility of this approach for labeling, target engagement and enrichment studies. Transcriptome-wide mapping studies (Chem-CLIP) in the presence of the bacterial aptamer reveal an unexpectedly high specificity for the bacterial aptamer over all other human RNAs. In total human RNA (without aptamer added), the probe enriched several transcripts significantly. Validation of these genes with competitive assays revealed several specific interactions between PreQ₁ and human RNAs.

Results

Synthesis and evaluation of electrophilic probes. To design covalent probes targeting the PreQ₁-I class riboswitch aptamers, a structure-informed strategy was taken. From multiple available structures of ligand-bound riboswitches it was apparent that the pendant amine from PreQ₁ was involved with binding but was also mostly solvent exposed^{29,30,32–34}. Furthermore, proximal to the binding site were multiple different potential nucleophilic groups. For example, the 2'-OH of G11 or base side chains of G4

and G5 from the *Thermoanaerobacter tengcongensis* (*Tt*-preQ₁-RS) (PDB: 3Q50) aptamer are all near the binding site^{29–32}. We thus designed and synthesized a series of probes that contain commonly used electrophilic warheads for covalent modifications that could potentially react with these groups. As described in Table 1, these compounds included Michael acceptors (1–3), acylating reagents (4–9), and the chlorambucil derivative that has been used in Chem-CLIP experiments (10) and has been reported to successfully alkylate RNA in a non-specific manner^{36,37}.

The reactivity of each of these probes towards PreQ₁-I class riboswitch aptamers was evaluated using mass spectrometry (e.g., MALDI-TOF) and denaturing PAGE experiments. Initial in vitro analysis of the reaction mixtures between the *Tt*-preQ₁-RS aptamers and the electrophilic probes showed some evidence of higher molecular weight species corresponding to potential adducts in low yields. However, these species did not correlate to a specific product. Furthermore, when these reaction mixtures were subjected to denaturing PAGE analysis, no evidence was seen for specifically modified RNA (Supplementary information Fig. S1). Efforts to optimize the reaction conditions by altering reaction temperature or buffer conditions did not affect the outcome. Based on these findings, we speculate that these electrophiles may simply not be reactive enough to yield high percentages of covalently modified RNAs, or that the electrophiles could not achieve the proper geometry for covalent modification. Similarly, others have reported that modification of the amine at the C7 position of PreQ₁ with various aminoalkyl chains results in decreased activity in an in vivo *E. coli* reporter system³⁸. To overcome this challenge, we turned our attention towards the synthesis and evaluation of more reactive, diazirine-containing photoaffinity crosslinkers.

Design and optimization of photocrosslinking covalent probes.

Similar to electrophilic probes, photoaffinity probes are commonly employed for labeling studies with proteins^{39,40}. Diazirine functionalized nucleosides have been used previously to map nucleic acid-ligand interactions or conjugate dyes to aptamers^{19–22}. However, typically only one probe was reported, and the effect of linker structure or crosslinking group was not broadly evaluated. In order to assess how photoaffinity linker structure influences crosslinking efficiency, a series of photoaffinity probes were designed and synthesized (Fig. 1 and Table 1). Additional details for the synthesis and characterization of these compounds are provided in the supplementary information (Section 3: synthesis and characterization). Included in this series are probes varied in crosslinker structure (stabilized and unstabilized diazirines), as well as linker length^{41,42}. The *Bs*-preQ₁-RS was treated with 50-fold excess of probe in the presence of riboswitch buffer containing 2 mM MgCl₂, irradiated with 365 nm light sources for 15 minutes, and PAGE and MALDI MS analysis was used to assess the formation of crosslinks. Gratifyingly, the expected adducts were observed in all cases, however only trace crosslinking was observed with 14 in MALDI. Crosslinking efficiencies were above 30% for unstabilized probes 11, 12, and 13, and approximately 20% with stabilized probes 15 and 16 (Supplementary information Fig. S2).

Effect of linker length and composition on photocrosslinking efficiency. The behavior of compounds 14–16 highlight the importance of the linker structure and length in preserving the native interaction between PreQ₁ and the riboswitch as well as enabling crosslinking. Crosslinking was not observed for 14 in denaturing PAGE analysis, potentially due to steric clashes that could occur between the phenyl ring in the linker and the RNA that prevent the compound from inserting fully into the binding

pocket. The results for compounds 15 and 16 support this conclusion as some crosslinking can be restored when the phenyl ring is spaced further away from the PreQ₁ core by PEG chains. Overall, we showed that both compounds 11 and 13 can modify the aptamer with high efficiency (>30%) under these conditions. We chose to proceed with compound 11 because it retained the positive charge in the natural PreQ₁ ligand and contained an alkyne handle for further modification.

We next evaluated the effects of crosslinking 11 to the PreQ₁ aptamer as a function of both time and concentration. In both denaturing PAGE and MALDI-TOF assays, clear dose and time-dependent crosslinking was observed (Figs. 2a, b and S3). Importantly, even at higher doses or longer times of irradiation, a single crosslinked species was the predominant product, indicating a highly specific crosslinking event.

Competitive crosslinking experiments demonstrate selectivity in vitro. With an optimized probe in hand, we next explored whether the photoaffinity labeling with compound 11 achieves proximity induced site-specificity or if the observed adducts were due to indiscriminate reactivity of diazirine with the RNA. Consequently, we performed labeling experiments in the presence of increasing concentrations of unmodified PreQ₁. PreQ₁ showed a dose-dependent inhibition of crosslinking of 11, with complete inhibition of crosslinking observed at higher concentrations (Fig. 2c). To further establish the selectivity of 11, we evaluated the labeling efficiency in the presence of up to tenfold excess of tRNA with respect to the aptamer. No significant decrease in the labeling efficiency was observed under these conditions, indicating that 11 retains selectivity for the aptamer in the presence of excess tRNA (Fig. 2d). We also performed labeling experiments with other structured RNAs such as microRNA 21 (miR-21) and SAM II riboswitch with compound 11. No detectable high molecular weight (slow migrating) crosslink product was observed in denaturing PAGE gels under these conditions (Fig. 2e).

Mass spectrometry and crystallography establish mode of interaction with 11. In an effort to probe the site-specificity of the crosslink further, the photocrosslinked RNA was subjected to a nucleoside digestion using NEB Nucleoside Digestion Mix[®]. The resulting reaction mixture was analyzed by Orbitrap-Liquid chromatography–mass spectrometry (LC/MS) and we identified guanosine-compound 11 adduct as a primary modified species (Fig. 3a, b). Moreover, further fragmentation and MS/MS analysis of guanosine adduct using a higher-energy collisional dissociation (HCD) and collision-induced dissociation (CID) techniques showed either a loss of the sugar or the PreQ₁ portion of the probe (Fig. 3c–e). Based on these fragmentation patterns, we conclude that the photoaffinity probe exclusively reacts with guanosine on the nucleobase rather than on the sugar. These findings are consistent with proximity-induced alkylation of guanines located near the binding site of the ligand as observed in crystal structures.

To further validate the binding of 11 to the PreQ₁ aptamer, we performed X-ray crystallography with the PreQ₁ aptamer from *T. tengcongensis* (*Tt*-preQ₁-RS), using a previously reported method²⁹. Efforts to determine the X-ray crystal structure of 11 photocrosslinked to the *Tt*-PreQ₁-RS aptamer were unsuccessful. Co-crystals were obtained using wild type (WT) and an abasic mutant at positions 13 and 14 (ab₁₃₋₁₄) without photocrosslinking, and the complex structures were solved by the molecular replacement method at 2.80 and 1.57 Å resolution respectively (Fig. 4a, b and Fig. S4, and Table S1). Positions A13 and A14 of the *Tt*-preQ₁-RS are important for stabilization of the expression platform³¹. However, when we co-crystallized the wild-type *Tt*-

Table 1 Structures and crosslinking efficiencies for electrophilic and photoaffinity probes used for proximity induced covalent modification of the *Tt*-preQ₁-RS aptamer (compounds 1-10) or *Bs*-preQ₁-RS aptamer (compounds 11-16). Error bars indicate standard deviation, *n* = 3. (Molar ratio between the aptamer and probe is 1:50).

Compound ID	Compound structure	% Crosslink efficiency (PAGE)
1		Not observed
2		Not observed
3		Not observed
4		Not observed
5		Not observed
6		Not observed
7		Not observed
8		Not observed
9		Not observed
10		Not observed
11		Not observed
12		31.3 ± 2.9
13		49.2 ± 3.2
14		39.9 ± 5.3
15		Trace
16		Trace

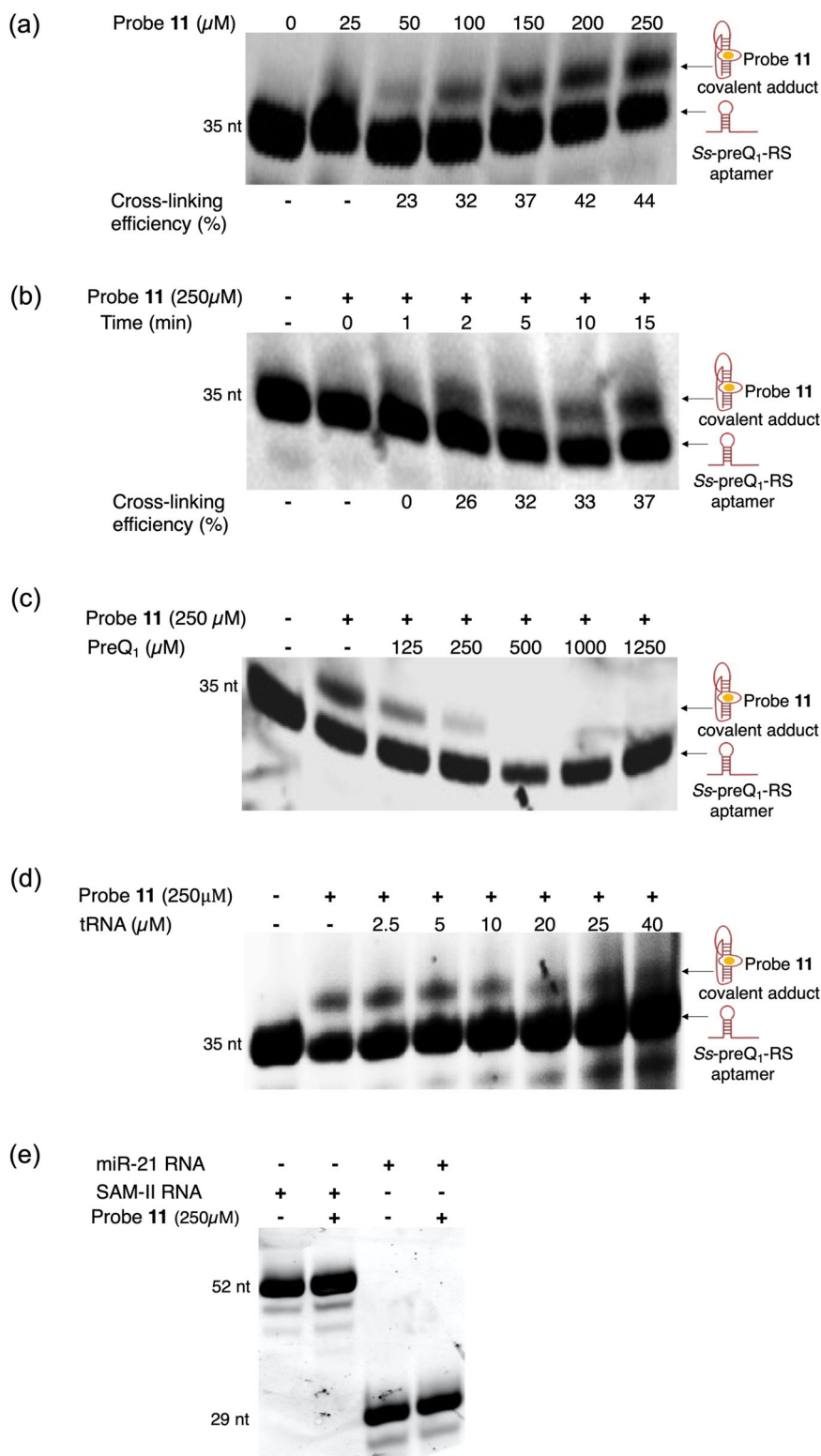
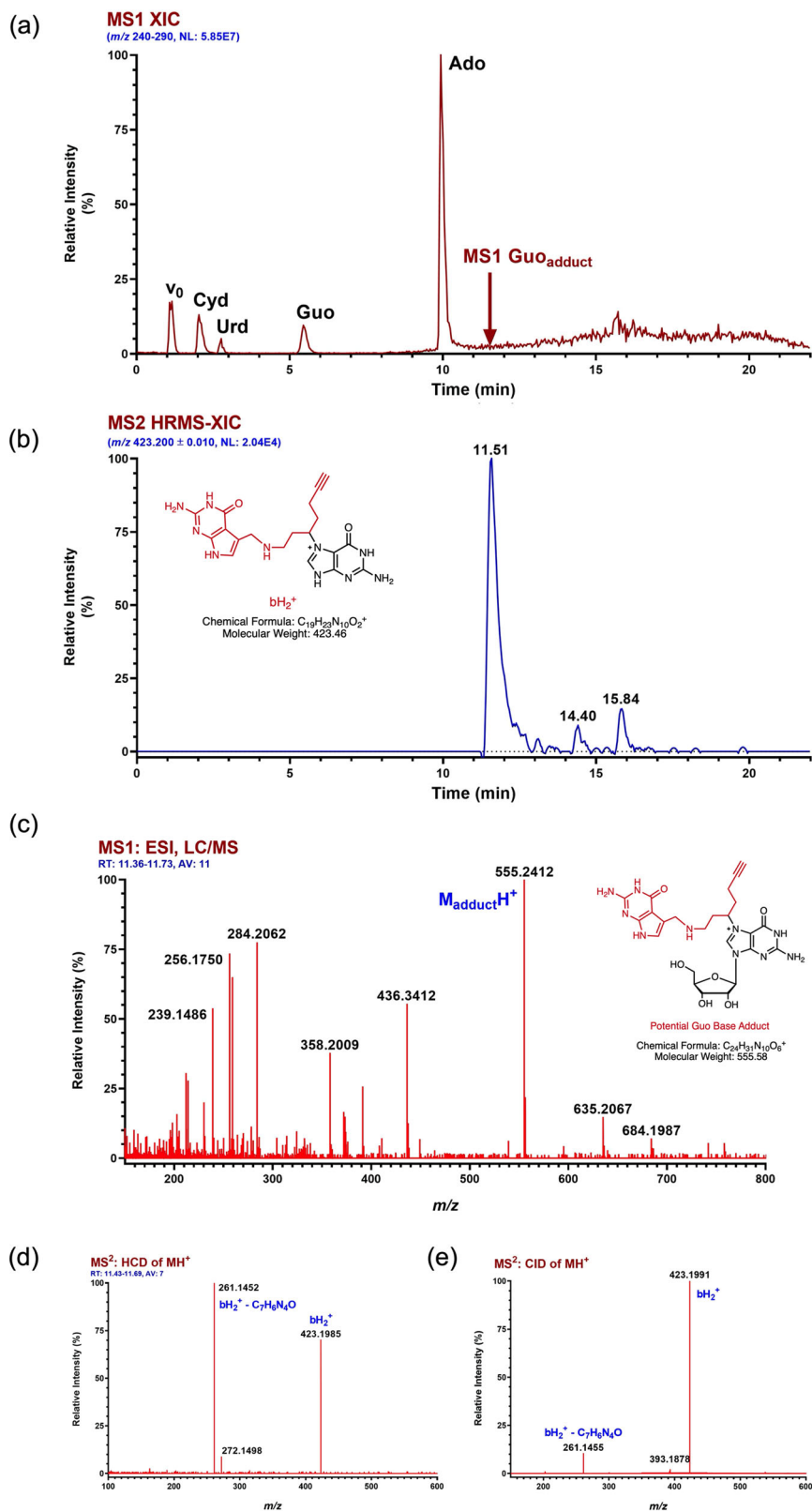


Fig. 2 Biochemical optimization of photocrosslinking and competition experiments to evaluate the specificity of compound 11 for the Ss-preQ₁-RS aptamer. **a** Dose-dependent crosslinking efficiency of compound 11. **b** Time-dependent crosslinking efficiency of compound 11. Competition experiments in the presence of **c** PreQ₁ and **d** tRNA. **e** PAGE analysis to assess crossreactivity of compound 11 with other structured RNAs shows no quantifiable modification of either miR-21 RNA or SAM-II RNA. Values for crosslinking efficiency are reported as the mean \pm standard deviation of three replicate experiments. Source data are provided as a Source Data file.



sidechain does not ablate the inhibitory activity of the ligand or negatively impact selectivity.

Probe 11 retains selectivity in cell lysates and total human RNA. Having demonstrated the selectivity of **11** towards the PreQ₁ aptamer in vitro and confirmed that the

photocrosslinking probe binds to the RNA like the unmodified ligand, we next explored the potential applications of ligand-based strategy for labeling applications, including fluorophore- and biotin-labeling^{17,19-22}. To do this, we synthesized biotin-conjugated **11** (Bio-**11**) using copper catalyzed click chemistry with biotin-PEG₃-azide and performed photocrosslinking experiments with the Ss-preQ₁-RS under our optimized reaction

Fig. 3 Probe 11 selectively crosslinks to guanosine bases in the PreQ₁ aptamer. High-resolution accurate mass (HRAM) and positive ion LC/MS/MS analysis of the nucleoside digest resulting from adducted *Ss*-preQ₁-RS (RNA 35-mer). **a** MS1 extracted ion chromatogram (XIC) for *m/z* 240–290 showing the response for unmodified RNA nucleosides. HPLC void volume is indicated by v0 and the elution time of the major RNA probe adduct is indicated by a red arrow. **b** MS2 (CID of Guo-adduct MH⁺) HRAM XIC for bH2⁺, the ion indicating that adduct formation occurs on the base. Extracted ion mass tolerance is ±10 mDa. The ion structure represents one of several possibilities for adduct and charge location. **c** Averaged and background-subtracted high-resolution MS1 spectrum for the major Guo-adduct located through HRAM XIC analysis of the theoretical MH⁺ (*m/z* 555.2423) for potential adducts. A plausible structure of the adduct is illustrated. **d** Full scan high-resolution MS/MS spectrum for HCD of MH⁺ of the Guo-adduct eluting at 11.51 min in **a** & **b**. **e** Full-range high-resolution MS/MS spectrum for CID of the same Guo-adduct eluting at 11.51 min in **a** & **b**.

conditions. MALDI-TOF analysis of the reaction mixture showed the expected molecular weight adducts (Supplementary information Fig. S8). Concurrently, the photocrosslinked *Ss*-preQ₁-RS aptamer was also subjected to copper catalyzed click reaction with biotin-PEG₃-azide and Cy5-picoyl azide and the expected click product was observed in both experiments by mass spectrometry and denaturing PAGE analysis (Supplementary information Figs. S9 A and B and S10).

Bio-11 was prepared and added to MCF-7 cell lysates that had been supplemented with *Ss*-preQ₁-RS aptamer. After irradiation, the lysates were analyzed by denaturing PAGE. As shown in Fig. 5a, when the gel was stained with SYBR gold stain, a crosslinked higher molecular weight band was observed. In order to confirm the presence of the biotinylated, crosslinked product, RNA samples from the gel were then transferred to a positively charged nylon transfer membrane and the blot was incubated with streptavidin-HRP overnight. As shown in Fig. 5a, biotin modified RNA bands were only observed when the lysates had been photoirradiated in the presence of probe Bio-11.

Next, we demonstrated that 11 could be used for both the photocrosslinking and biotin labeling in a complex environment. We isolated total cellular RNA from MCF-7 cells, added 1 μM of *Bs*-preQ₁-RS (a 70 nt construct containing both the aptamer and expression platforms), and treated with compound 11 in riboswitch buffer. After incubation, samples were irradiated at 365 nm for 15 minutes. Subsequently, we performed copper catalyzed click reaction with TAMRA azide and ran the samples on denaturing PAGE. In the absence of the MCF-7 total RNA, a strong TAMRA fluorescent band was observed at the expected molecular weight indicating that the photocrosslinking and click reactions were successful (Fig. 5b, lane 2). Additionally, only weak fluorescent signals were observed when the *Bs*-preQ₁-RS aptamer was not added into the mixture of 11 and MCF-7 total RNA (Fig. 5b, lane 3). In contrast, as shown in lane 4 of Fig. 5b, compound 11 was able to selectively modify the PreQ₁ riboswitch even in the presence of other cellular RNAs.

Consistent with earlier observations with the *Ss*-preQ₁-RS aptamer, photocrosslinking efficiency is significantly reduced in the presence of free PreQ₁ competitor (Fig. 5b, lanes 5 & 6). This observed competition between 11 and PreQ₁ demonstrates the importance of proximity induced site-specific modification within the binding pocket. Additionally, no fluorescent band was observed when the reaction mixture was not irradiated with UV, confirming that irradiation of 11 is required for detection of the aptamer with this method (Fig. 5b, lane 7). In comparison, when compound 17, a small molecule containing alkyne-diazirine fragment that lacks the PreQ₁ targeting portion, was used to treat total RNA under similar conditions, no detectable labeling of the riboswitch was observed, again confirming that the recognition event between the PreQ₁ scaffold of 11 and the aptamer is critical for diazirines to be useful as photocrosslinking partners (Supplementary information Fig. S11).

Chem-CLIP sequencing experiments confirm target engagement in total human RNA. Chem-CLIP experiments were also

performed to further probe the selectivity of 11 for the full-length *Bs*-preQ₁-RS aptamer in the presence of MCF-7 derived total cellular RNA³⁶. After UV crosslinking in the presence of 11, total RNA was click labeled with biotin-PEG₃-azide and enriched with streptavidin-conjugated magnetic beads. Enriched RNA was then reverse transcribed, and samples were submitted for next-generation sequencing. Compound 17, which lacks the PreQ₁ targeting warhead, was again used as a negative control. In samples where *Bs*-preQ₁-RS was spiked into total RNA, differential gene expression analysis between 11 and 17 treated samples revealed that the aptamer sequence was significantly enriched by compound 11 (\log_2 fold-change > 2.5, adjusted p -value = $2.93e^{-21}$ (Fig. 6a)). As evident in Fig. 6b and c, no other RNAs were significantly enriched when comparing 11 to 17 in this analysis. This high level of selectivity makes this compound-aptamer an ideal pair for chemical biology applications that require specific labeling in complex environments.

Chem-CLIP and c-Chem-CLIP sequencing experiments to quantitatively assess selectivity of 11 in human transcriptomes.

To further investigate crossreactivity of 11 with other human cellular RNAs, a Chem-CLIP experiment was also performed without exogenous *Bs*-preQ₁-RS. In the absence of *Bs*-preQ₁-RS, 11 selectively enriched 16 transcripts including TERC, HIST1H3F and HIST2H2BF (cutoff: \log_2 fold-change > 0.95, $-\log_{10}P > 4$) (Fig. 7a and Supplementary Information S12). No genes were significantly depleted when the same cutoffs are applied. Interestingly, 14 out of 16 of the most significantly enriched hits belong to the histone gene family. Owing to the high level of sequence conservation between these genes, it is possible that their mRNA products share a common structure (i.e., an aptamer-like domain) that is specifically being recognized by 11. For example, histone family mRNAs have a conserved, stable stem loop at their 3' end in place of the typical polyadenylated tail that forms essential interactions with stem-loop proteins required for histone mRNA processing and translation^{44,45}. The two other hits in this group, TERC and RMRP, both encode structured non-coding RNAs that form ribonucleoprotein complexes with the human telomerase catalytic subunit (hTERT)⁴⁶. The lack of enrichment of these genes in the presence of *Bs*-preQ₁-RS, which presumably sequesters the majority of 11, suggests that these could represent specific interactions (Fig. 5a).

Competitive quantitative reverse transcription PCR (RT-qPCR) experiments (c-Chem-CLIP) were also performed with TERC, HIST1H2F and HIST2H2BF to validate individual enrichment events and assess biochemical competition with unlabeled probes⁴⁷ (Fig. 7b–d). By RT-qPCR, TERC was enriched approximately 1.5-fold in 11 treated samples compared to the negative control (17), in good agreement with sequencing data. Furthermore, this enrichment was competitive with free PreQ₁ ligand, which lacks the diazirine-alkyne crosslinking warhead. Similar results were observed with HIST1H3F and HIST2H2BF. Enrichment was also observed for HIST2H2BF, however, competition with PreQ₁ was only evident at the highest concentration of competitor (100 μM)

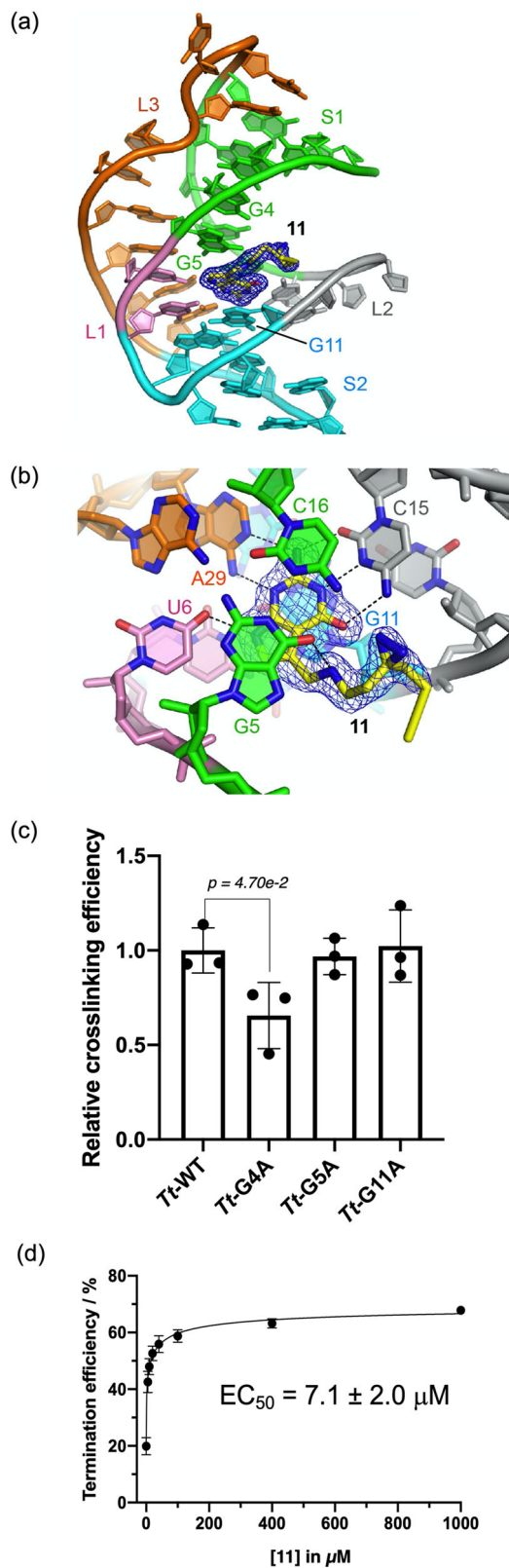


Fig. 4 An X-ray co-crystal structure and functional assessment of **11.** **a** Co-crystal structure showing interaction of compound **11** with the *Tt*-preQ₁-RS aptamer (ab_13-14) solved at 1.57 Å resolution. S1, S2, L1, L2, and L3 are colored green, cyan, pink, gray, and orange, respectively. The mF_o - DF_c electron density maps for the compound are colored blue and contoured at 3.0 σ. **b** Detail of ligand:RNA interaction showing conservation of contacts seen with unmodified PreQ₁. The nucleotides that interact with the compound are labeled, and hydrogen bonds are indicated as dotted lines. Residues at the site of the interaction are labeled to show the proximity of G5 and G11 to the diazirine. The mF_o - DF_c electron density maps for the compound are also shown as in **a**. **c** Bar graph representing the photocrosslinking efficiency of wild-type *Tt*-preQ₁-RS aptamer (*Tt*-WT) and three different mutant aptamers (*Tt*-G4A, *Tt*-G5A, and *Tt*-G11A) to Probe **11**. **d** Single-round transcription termination assay results for compound **11**. Quantification of transcription termination efficiency with increasing concentrations of **11** was calculated based on the band intensity of read-through transcription product (RT) and terminated transcription product (T). Data are presented as the mean ± SEM (n = 3) of three independent experiments. Statistical significance was calculated by two-tailed t-test analysis. Source data are provided as a Source Data file.

Discussion

Herein, we employ a structure-informed strategy to design and characterize covalent probes of the PreQ₁ riboswitch aptamer. Initial attempts to design electrophilic probes were unsuccessful, however photocrosslinking probes reacted with efficiency approaching 50% modification. We applied multiple different biophysical and biochemical assays to assess the reactivity and site-specificity of the photoaffinity probe towards the target RNA in crosslinking experiments. Digestion of the crosslinked aptamer and analysis by mass spectrometry revealed that the photocrosslinker only reacts with guanine bases, with no crosslinks observed to the sugars. Given that there are only 3 conserved guanines in the class-I preQ₁ aptamer (Bs, Tt and Ss), two of which are proximal to the binding site, this observation confirms a highly specific crosslinking event. The carbene-based reactivity typically associated with diazirines would be expected to predominate in stabilized diazirine probes such as compounds **14–17** and to alkylate multiple chemically diverse sites due to high reactivity^{41,42}. However, here we observe unstabilized diazirines as having higher crosslinking efficiency and specific reactivity, suggesting that crosslinking may occur through an alkylation event via a rearranged diazo-intermediate (as observed elsewhere^{48,49}). Based on the high reactivity of this intermediate it is likely that **11** can form adducts with multiple different atoms in multiple different nucleobases in the binding pocket, and more than one product is possible. However, we observed that mutation of G4 to adenosine resulted in a roughly 40% reduction in crosslinking, suggesting that this residue is the majority site of modification.

Importantly, binding studies, X-ray crystallography and transcription termination assays revealed that labeling PreQ₁ with the crosslinking sidechain does not ablate binding or function and preserves the majority of the contacts that the compound makes with the aptamer. Despite a 50-fold reduction in affinity, the specificity of this interaction enables selective photocrosslinking of **11** to a PreQ₁ riboswitch in cell lysates, which further illustrates the robustness of this approach to selectively target the RNA of interest under biologically relevant conditions. Thus, this PreQ₁/aptamer pair may find substantial utility in chemical biology applications for targeting RNA due to the ease of manipulation of the PreQ₁ scaffold and robustness of the proximity induced crosslinking. This straightforward approach relies exclusively on

(Fig. 7d). Together, the results of these sequencing and qPCR experiments indicate that these RNAs selectively bind to this ligand. Further investigation is underway to map these crosslinks and fully characterize these interactions. Extension of this c-Chem-CLIP crosslinking pull-down strategy to live cells in an effort to confirm the relevance of these potential interactions with **11** is currently underway.

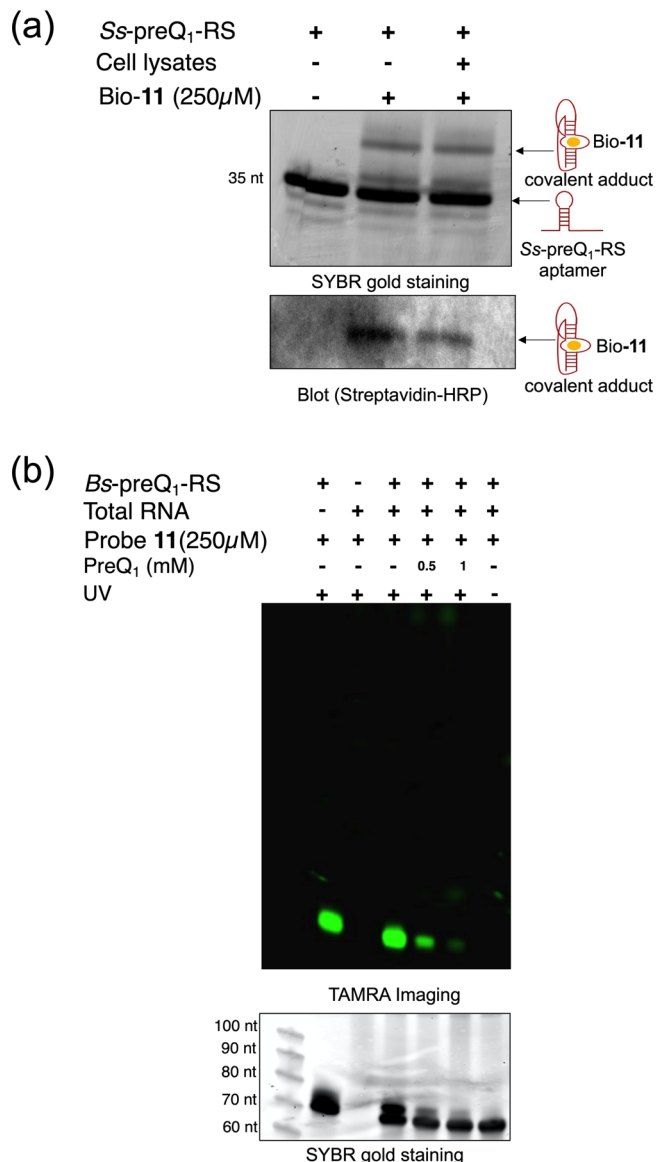


Fig. 5 Diazirine probe **11** selectively labels the PreQ₁ aptamer in cell lysates. **a** In denaturing PAGE experiments Bio-**11** selectively crosslinks the Ss-preQ₁-RS aptamer in the presence of cell lysates. The crosslinked product is observed by SYBR gold staining and streptavidin-HRP detection. **b** Competitive photocrosslinking experiments with Bs-preQ₁-RS full-length aptamer (70 nts) spiked into MCF-7 total RNA and imaged by TAMRA labeling (top) and SYBR gold stain (bottom). Source data are provided as a Source Data file.

chemical recognition to achieve site-specific modification of an RNA. By fully characterizing this specific recognition event, we were additionally able to probe the transcriptome for potential PreQ₁ binding sites on human RNAs. Through Chem-CLIP and competitive qPCR experiments in total cellular RNA, we were able to show that **11** crosslinked to the aptamer with exceptional selectivity. In the absence of aptamer, **11** selectively enriched 16 RNAs when compared to the negative control probe **17**. The high level of structural similarity between PreQ₁ and guanine is particularly intriguing in this regard, as it is possible that guanine can also specifically interact with structures in these RNAs. In this way, evaluation of bacterial riboswitch ligands like PreQ₁ can shed light on potential existing RNA aptamers for human metabolites.

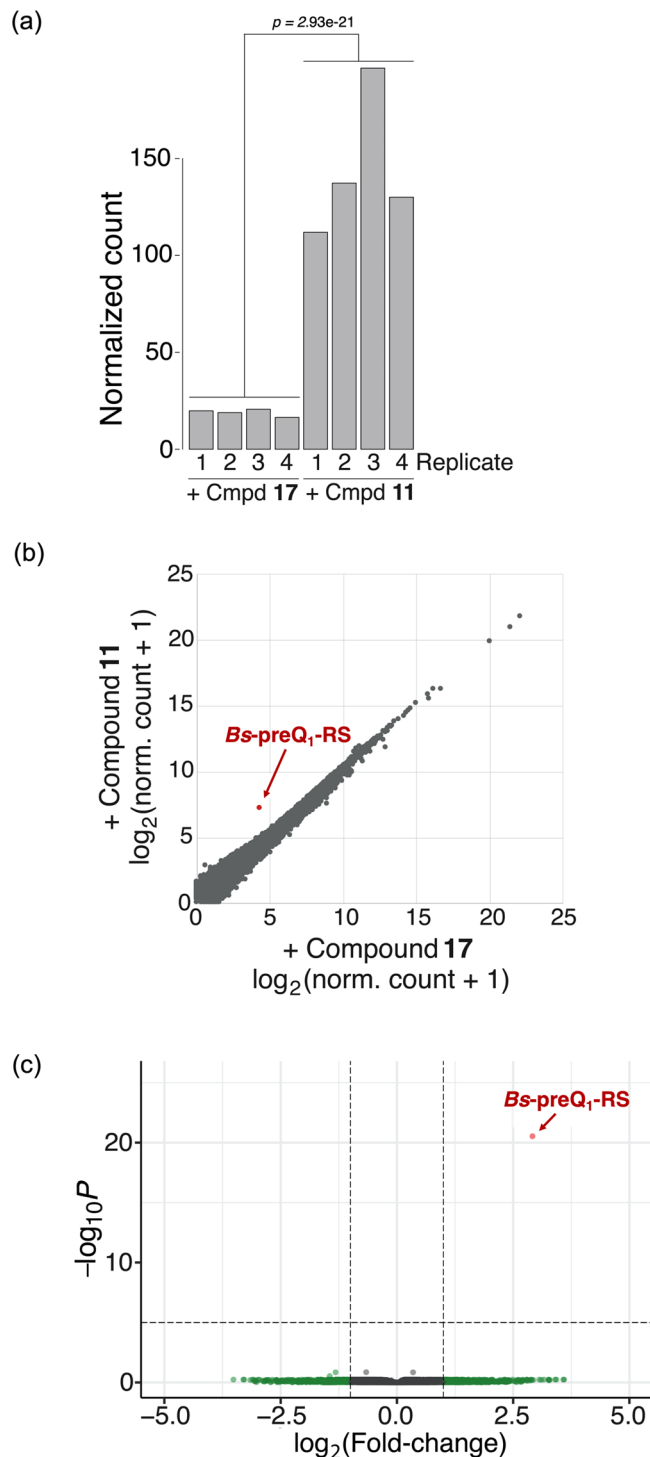
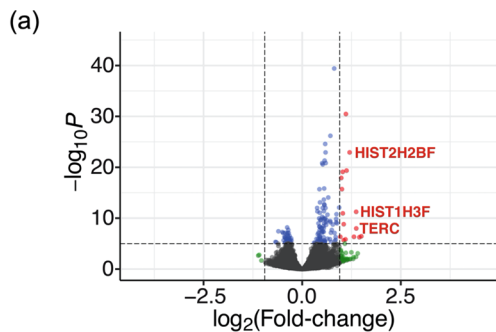
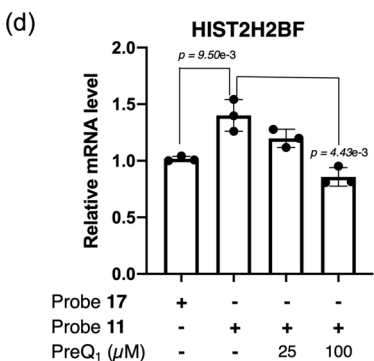
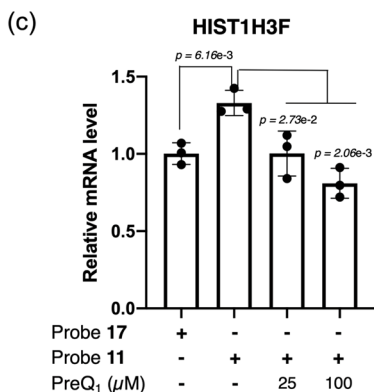
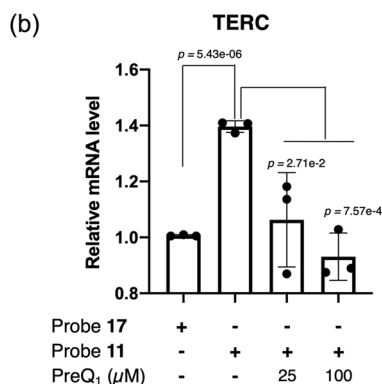


Fig. 6 Selective enrichment of Bs-preQ₁-RS aptamer by **11** in total human RNA with Bs-preQ₁-RS aptamer spiked in. **a** Comparison of the normalized counts of the Bs-preQ₁-RS aptamer across **11** and **17** treated samples shows significantly higher counts in all four replicates of **11** treated samples. Statistical significance was calculated by two-tailed t-test analysis. **b** Correlation plot showing differential gene enrichment for compounds **11** and **17**. The gene entry corresponding to the Bs-preQ₁-RS is labeled in red. **c** Volcano plot from differential expression analysis between **11** and **17**. The gene entry corresponding to the Bs-preQ₁-RS is labeled in red. p = adjusted p -value. All analyses performed on four independent replicate samples treated with **11** or **17**.

**Fig. 7 The enrichment of endogenous human transcripts with 11. a**

Volcano plot of differential expression (DeSeq2) analysis between compound **11** and **17** treated samples ($n = 4$ for **11** and $n = 3$ for **17**). $p =$ adjusted p -value. Histograms generated through RT-qPCR representing the relative level of **b** TERC, **c** HIST1H3F, and **d** HIST2H2BF in **11** and **17** treated samples. Data are presented as the mean \pm SEM ($n = 3$) of three independent experiments. Statistical significance was calculated by two-tailed t -test analysis. Source data are provided as a Source Data file.



Humans lack the biosynthetic machinery to produce PreQ₁ and related metabolites such as queuine and queuosine, and they must be acquired from diet or commensal bacteria such as *B. subtilis*. However, they are important metabolites that specifically modify certain human tRNAs to promote protein folding^{10,50,51}. Moreover, queuosine modification occurs at the wobble position of the tRNA with GUN anticodons and this modification

regulates the translation efficiency⁵². The specific role that these metabolites play in humans remains under investigation, though there is at least some evidence that they may also have anti-cancer activity in addition to their regulatory roles in commensal bacteria^{53,54}. Here, we demonstrate that **11** also has specific interactions with important regulatory human RNAs, including TERC, HIST1H3F, and HIST2H2BF, all of which have been shown previously to form stable RNA structures. 14 of the 16 enriched genes were for histone mRNAs, which is intriguing due to the previously reported shared, stable stem-loop motif in the 3' UTR of histone mRNAs that plays an important regulatory role in their expression through interactions with stem-loop binding protein (SLBP)⁴⁵. Additionally, the other two enriched genes, TERC and RMRP are both highly conserved short, structured non-coding RNAs that are known to play roles in disease pathogenesis^{55–57}. The structure of vertebrate TERC has been studied extensively and this RNA has been shown to form multiple stable structures including a pseudoknot domain that is required for binding to hTERT⁵⁸. Telomerase activity is low in the majority of healthy somatic cells, however, it is significantly increased in various cancers⁵⁵, and TERC has been implicated as a potential therapeutic target in many of these cancers⁵⁹. While the biological significance of the proposed interaction between **11** and TERC is currently unknown, small molecule probes that can selectively recognize this RNA would be valuable to researchers studying telomerase biology. Overall, the observation that simple metabolites have specific binding interactions with important regulatory RNAs such as these is intriguing and will require more in-depth investigation to evaluate potential regulatory roles.

This work highlights the importance of design considerations when generating chemical probes to crosslink RNA. We designed 16 probes, but only a small number crosslinked with sufficient efficiency to be broadly useful. Thus, care should be taken to design probes that avoid false negative events in photocrosslinking applications, particularly if measuring selectivity in a complex system is the goal. We find that unstabilized diazine probes modify RNA with the highest efficiency, likely through an alkylation mechanism, and that linker structure has a high impact. Through in-depth biophysical analysis and Chem-CLIP studies, we demonstrate high selectivity between **11** and three class-I PreQ₁ aptamers in cell lysates and total RNA, indicating potential uses for this system in chemical biology applications. Further, probe **11** targets both minimal aptamer domains and full riboswitches that include both transcriptional and translational expression platforms. Finally, mapping PreQ₁ binding events in human transcriptomes revealed several specific binding events to regulatory RNAs. Further investigation of these and other interactions, including cell-based photocrosslinking approaches may prove useful to interrogate how related metabolites like guanine interact with human RNAs.

Methods

General information. Unless otherwise stated, all reagents and solvents were purchased from commercial vendors and used without further purification. All ¹H, ¹³C, and ¹⁹F NMR spectra were collected on a 500, 126, and 471 MHz NMR

spectrometers in (CD₃)₂SO, respectively. The ¹H, ¹³C, and ¹⁹F NMR chemical shifts are given in parts per million (δ) with respect to an internal tetramethylsilane (TMS, δ = 0.00 ppm) standard and/or 3-chlorobenzotrifluoride (δ = -64.2 ppm relative to CFCl₃). NMR data are processed, annotated, and reported in the following format: chemical shift (multiplicity (s = singlet, brs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), coupling constants (Hz), Integration) using MestReNova. The large-scale synthesis of 7-aminomethyl-7-deazaguanine (PreQ₁) was carried out according to previously reported protocol⁶⁰.

Oligonucleotides. Deprotected and HPLC purified riboswitch aptamers or full-length riboswitch were purchased from Horizon Discovery Inc. or IDT with the following sequences and used without any further purifications.

Tt-preQ₁-RS: 5'-CUGGGUCGAGUAACCCAGUUAACAAAACAAG-3'
Cy5-Tt-preQ₁-RS: 5'-Cy5-
 CUGGGUCGAGUAACCCAGUUAACAAAACAAG-3'
Tt-preQ₁-RS(G4A): 5'-
 CUGAGUCGAGUAACCCAGUUAACAAAACAAG-3'
Tt-preQ₁-RS(G5A): 5'-
 CUGGAUCGAGUAACCCAGUUAACAAAACAAG-3'
Tt-preQ₁-RS(G11A): 5'-
 CUGGGUCGAAUAACCCAGUUAACAAAACAAG-3'
Tt-preQ₁-RS (ab 13_14): 5'- CUGGGUCGAGU-rab-rab-
 CCCAGUUAACAAAACAAG-3' (rab means abasic site)
Bs-preQ₁-RS: 5'-AGAGGUUCUAGCUACACCCUCUAUAAAAACUAA-3'
Cy5-Bs-preQ₁-RS: 5'-Cy5-
 AGAGGUUCUAGCUACACCCUCUAUAAAAACUAA-3'
Bs-preQ₁-RS full length (includes both the aptamer domain and the expression platform): 5'-
 GCGGGAGAGGUUCUAGCUACACCCUCUAUAAAAACUAGGACGAGCU-
 GAUCCUUGGAUACGGCCUUUU-3'
Ss-preQ₁-RS: 5'-AGAGGUUCUAGCUGAUACCCUCUAUAAAAACUAA-3'
Cy5-Ss-preQ₁-RS: 5'-Cy5-
 AGAGGUUCUAGCUGAUACCCUCUAUAAAAACUAA-3'
SAM-II-WT-RS: 5'-
 UCGCGUGAUUUACCGUAUUGCAAGCGCGUGAUAAAUGU
 AGCUAAAAAGGG
miR21-hp: 5'-GGGUAGCUGUUGAUUCUCAUGGCAACCC-3'

Covalent crosslinking with electrophile containing compounds 1-10. The Tt-preQ₁-RS aptamer (see above) was diluted to 10 μM in riboswitch buffer (50 mM Tris (pH 7.5), 100 mM KCl, 2 mM MgCl₂) and was folded by heating at 75 °C in a heat block for 5 mins, and was allowed to cool to RT over the course of 1 h. For compound treatment, 10 μL of total volume was prepared by adding 9.5 μL of the refolded Tt-preQ₁-RS to 0.5 μL of DMSO containing compounds 1-10 (500 μM final concentration, 50:1 molar ratio of crosslinking small molecule/RNA). The solutions were incubated at 40 °C for 2 h. After incubation, the samples were mixed with denaturing loading dye and heated at 95 °C for 5 min. Equal volumes of the samples were then subjected to 17% denaturing PAGE.

UV photocrosslinking experiments. For the initial screening of compounds 11-16 binding to Bs-preQ₁-RS aptamer, the Bs-preQ₁-RS (5 μM, 50 μL) in riboswitch buffer (50 mM Tris (pH 7.5), 100 mM KCl, 2 mM MgCl₂) was heated to 75 °C for 5 min and allowed to cool to room temperature over 1 h. After folding, the RNA was treated with compounds 11-16 (250 μM final concentration in 5% DMSO/riboswitch buffer) and incubated for 30 min in the dark at room temperature. Note that it is advised to keep the DMSO level at or below 5% in the reaction mixture. After pre-incubation, the reaction mixture was irradiated with 365 nm light in a SPECTROLINKER XL-1000 UV Crosslinker for 15 min. Multiple crosslinking methods were explored; however, optimal results were obtained by transferring the solution to a clear 96-well plate and placing the uncovered wells as close as possible to the UV light source in the photocrosslinker. Subsequently, the reaction mixture was analyzed by MALDI-TOF and/or denaturing PAGE.

For the dose-dependent crosslinking experiments, the Ss-preQ₁-RS aptamer (5 μM) was folded in riboswitch buffer and incubated with varying concentration of probe 11 (25 μM, 50 μM, 100 μM, 150 μM, 200 μM, and 250 μM) at room temperature for 30 min. Then samples were UV photocrosslinked as described above and analyzed by MALDI-TOF and/or denaturing PAGE.

For the time-dependent crosslinking experiments, the Ss-preQ₁-RS aptamer (5 μM) was folded in riboswitch buffer and incubated with probe 11 (250 μM) at room temperature for 30 min. Then samples were UV crosslinked as described above but with different length exposure to UV light (2, 5, 10, and 15 min). After the UV crosslinking, samples were analyzed by MALDI-TOF and/or denaturing PAGE.

For the competition experiments with PreQ₁ and tRNA, the Ss-preQ₁-RS aptamer (5 μM) was folded in riboswitch buffer and incubated with probe 11 (250 μM) at room temperature for 30 min. For competition experiments with PreQ₁, samples were prepared with free PreQ₁ ligand with increasing concentration from 125 μM to 1250 μM and added into the solution containing compound 11 and the Ss-preQ₁-RS to allow for competition before UV crosslinking. In the case

of tRNA competition, samples were prepared with tRNA with increasing concentration from 2.5 μM to 40 μM concentration and titrated into the solution containing compound 11 and the Ss-preQ₁-RS allow for competition before UV crosslinking. Samples were UV crosslinked for 15 min. After UV crosslinking, samples were analyzed by MALDI-TOF and/or denaturing PAGE.

Photocrosslinking experiments for other RNAs such as SAM-II and miR-21 were performed as described above. The molar ratio between the probe 11 and SAM-II or miR-21 was maintained at 50:1 ratio and samples were UV crosslinked for 15 min. Then samples were analyzed by PAGE.

For photocrosslinking experiments of 11 to Tt-preQ₁-RS mutants, the same protocol was used as described above. 10 μM of Tt-preQ₁-RS wild type or mutants were folded in riboswitch buffer and incubated with probe 11 (500 μM) at room temperature for 30 min. Samples were then UV crosslinked for 15 min. After photocrosslinking, samples were analyzed by PAGE.

Denaturing PAGE analysis (general procedure). For PAGE analysis, RNA samples were mixed with equal volume of denaturing loading dye and heated at 95 °C for 5 min. Equal volumes of the samples were then subjected to 17% PAGE in the presence of 7 M urea. The gel was stained in 1X SYBR® Gold Nucleic Acid Gel Stain (Invitrogen) for 15 min at room temperature. The gel images were obtained by scanning the gel on an AMERSHAM TYPHOON Biomolecular Imager. Image J software was used for processing. Crosslinking efficiency was calculated by quantifying the higher molecular weight band intensity (crosslink adduct band). Values reported are the average of three replicates.

Orbitrap-LC/MS analysis and digestion. The Ss-PreQ₁-RS aptamer (50 μM, 100 μL) was folded in riboswitch buffer by heating at 75 °C for 5 min and followed by cooling to room temperature over 1 h. Then the folded RNA was mixed with probe 11 (500 μM) and incubated for 30 min in the dark at room temperature. After pre-incubation, the reaction mixture was UV photocrosslinked for 15 min as describe above. Then sample mixtures (contains a mixture of photocrosslinked and unmodified RNA) were assessed for extent of crosslinking by negative ion MALDI operated in linear mode over a 5-20 kDa range. Next, the mixtures were analyzed by negative ion electrospray ionization (ESI)-LC/MS on an LTQ-Orbitrap-XL LC/MSn system at 30,000 resolutions. High-resolution deconvolution of oligonucleotide anion charge-states was carried out using the Xtract program⁶¹. Crosslinked RNA was subjected to enzymatic digestion to generate individual nucleosides and was analyzed by LC/MSn to determine where crosslinks were occurring and to identify possible structures. For the enzymatic digestion study, photocrosslinked RNA was transferred to a PCR tube and treated with NEB Nucleoside Digestion Mix® Reaction Buffer (10 μL, 10x) and Nucleoside Digestion Mix (5 μL) and incubated at 37 °C for 1 h. Then, the reaction mixture was analyzed by MS and MS/MS using both CID and HCD. A narrow-bore C18 column system, eluted with CH₃OH/H₂O gradients tailored to the compounds of interest, was used for separations. High-resolution CID and HCD MS/MS of the guanosine adduct MH⁺ indicated that adduction occurred exclusively on the guanine base.

Click reaction on photocrosslinked RNA for labeling with fluorophores and biotin-azides. Biotin or fluorophore-azide probes were crosslinked to aptamer RNA according to an established copper catalyzed click chemistry protocol⁶². A solution of each 10 μM Ss-preQ₁-RS aptamer was folded in the riboswitch buffer. Then the aptamers were incubated with 500 μM of Probe 11 and photocrosslinked as described above. Note that this crude solution contains a mixture of photocrosslinked and unmodified RNA (approximately 1:1.5 photocrosslinked to unmodified). To the above solution, either Biotin-PEG₃-azide or Cy5-picoly-azide (10 μL of 10 mM solution in DMSO), THPTA/CuSO₄ (25 μL of premixed solution of THPTA and CuSO₄ for 10 min in 2:1 ratio using 200 mM and 100 mM solutions in H₂O, respectively), and L-Sodium ascorbate (7.5 μL of 100 mM in H₂O) were added and the reaction mixture was stirred for 1 h. The resulting adducts were analyzed by MALDI-TOF and denaturing PAGE.

Single-round transcription termination assay. The transcription termination assays were carried out as previously reported²⁹. The DNA plasmid containing λPR promoter and 26-nt C-less sequence followed by the *Staphylococcus saprophyticus* (Ss) preQ₁ riboswitch and its downstream sequence cloned into pIDTSMART-AMP was purchased from Integrated DNA Technologies. The DNA template was amplified by PCR from the plasmid using forward and reverse primers, and then was gel-extracted after agarose gel electrophoresis for purification. Halted transcription complexes were prepared in a solution containing 1 μM GTP, 5 μM ATP, 5 μM UTP, 100 μM ApU, [α-³²P] GTP, 75 nM DNA template, 0.0167 U/μL *Escherichia coli* RNA polymerase holoenzyme (New England BioLabs) in 1 × transcription buffer (20 mM Tris-HCl, pH 8.0, 2 mM NaCl, 1 mM MgCl₂, 4% glycerol, 0.1 mM DTT, and 0.1 mM EDTA), and incubated at 37 °C for 15 min. A DNA oligonucleotide complementary to the 26-nt C-less sequence was added to the reactions at 1.1 μM final concentration in 1 × transcription buffer and incubated at room temperature for 5 min, in order to prevent undesired non-specific interactions between the 26-nt C-less sequence and riboswitch. Elongation was restarted by combining 9 μL of halted transcription complex, 3 μL of the compound either PreQ₁ or probe 11 (0-5 mM compound and 25% DMSO in 1 × transcription

buffer), and 3 μL of NTPs mix (200 μM ATP, 200 μM CTP, 200 μM GTP, 200 μM UTP, 100 $\mu\text{g}/\text{mL}$ heparin, and 250 mM KCl in 1 \times transcription buffer), and incubated at 37 $^{\circ}\text{C}$ for 20 min. To remove the DNA template, 0.5 U of RQ1 RNase-Free DNase (Promega) was added to the reactions and incubated at 37 $^{\circ}\text{C}$ for 10 min. The reactions were stopped by adding equal volume of loading dye (8 M urea, 20% sucrose, 0.05% bromophenol blue, and 0.05% xylene cyanol in 2 \times TBE). The reaction mixture was separated by 8% denaturing PAGE and visualized by phosphorimager. Full gel images are provided as Supplementary Figures. The band intensity was analyzed by ImageQuant software (GE Healthcare). Termination efficiency was calculated with dividing the intensity for the terminated RNA band by those for the total (terminated and antiterminated) RNAs.

Microscale thermophoresis (MST)-binding measurements for PreQ₁ and 11.

Fluorescently labeled *Bs*-preQ₁-RS aptamer and *Tt*-preQ₁-RS aptamer were prepared in 50 mM Tris-HCl (pH 7.5), 100 mM KCl, and 2 mM MgCl₂. Both preQ₁ riboswitch aptamers were annealed by heating to 75 $^{\circ}\text{C}$ for 5 min and then cooled to room temperature over 1 h. A twofold dilution series of PreQ₁ or probe 11 was prepared in 100 nM Cy5-labeled aptamer and 5% DMSO, with the final concentrations of PreQ₁ or 11 ranging from 500 μM to 0.0153 μM . Samples were incubated at room temperature for at least 15 min. Following incubation, the samples were added to premium coated capillaries (NanoTemper Technologies) and subsequently subjected to MST analysis. MST experiments were conducted in triplicate on a Monolith NT.115 system (NanoTemper Technologies). The results were analyzed by TJump analysis, and the values obtained were normalized and plotted against the PreQ₁ or 11 concentrations. The apparent dissociation constant was then determined using a single-site model to fit the curve.

X-ray crystallography experiments and data collection. For X-ray crystallography, we used the preQ₁ riboswitch aptamer domain from *Thermoanaerobacter tengcongensis* (Tt). The wild-type aptamer and abasic mutant (ab13_14), in which the nucleobases at positions 13 and 14 were removed, were employed for crystallization. The crystals of *Tt*-preQ₁-RS aptamer domain in complex with 11 were prepared at 20 $^{\circ}\text{C}$ and cryoprotected by the same methods as reported previously⁴. X-ray diffraction data were collected at the beamlines BL32XU and BL45XU of the SPring-8 (Hyogo, Japan) with the aid of an automatic data-collection system ZOO⁶³ and BL-1A of the Photon Factory (Ibaraki, Japan). Diffraction data were integrated and scaled with the programs KAMO⁶⁴ and XDS⁶⁵. Data processing statistics are summarized in Table S1.

Structure determination and refinement. The co-crystal structures were solved by the molecular replacement method with the program PHASER⁶⁶, using the previously determined *Tt*-preQ₁-RS aptamer domain structure (PDB ID: 6E1W) as the search model. The solutions were subjected to simulated annealing to uncouple the working and free *R* values in the refinement process, as well as energy minimization, restrained isotropic B-factor, and TLS refinement with PHENIX⁶⁷. The resulting electron density maps revealed the locations of the compound. The atomic models were then improved by iterative cycles of refinement with PHENIX and manual rebuilding with COOT⁶⁸. The current co-crystal structures of ab13_14 and the wild-type *Tt*-preQ₁-RS aptamer domain in complex with 11 were refined to an $R_{\text{work}}/R_{\text{free}}$ of 0.190/0.208 and 0.193/0.235 at 1.57 and 2.80 \AA resolution, respectively.

Cell culture and RNA isolation. MCF-7 cells were grown in T75 flasks in Dulbecco's modified Eagle's medium (DMEM) with high glucose supplemented with 10% fetal bovine serum and 1% antibiotics (streptomycin and penicillin) at 37 $^{\circ}\text{C}$ in 5% CO₂ in a humidified incubator according to ATCC's recommendations. On the day of the experiment (~80% confluency) the cells were harvested by trypsinization, washed and resuspended in 1x PBS. Total RNA was isolated from MCF-7 cells using the RNeasy mini kit (Qiagen, Cat #: 74104) with on column DNase digestion according to the manufacture protocol. The concentration of eluted RNA was measured using a nanodrop.

Photocrosslinking experiments with biotinylated probe 11 (Bio-11) in the presence of cell lysates.

Cell lysates were extracted from the MCF-7 cells using RIPA buffer (Santa Cruz). Then 1 μM of folded *Ss*-preQ₁-RS aptamer (34 nucleotide length, see above) was added into the 5 μg of cell lysate and photocrosslinked to Bio-11 (250 μM) as described above. Next, equal volumes of the samples were subjected to 17% denaturing PAGE in the presence of 7 M urea. Meanwhile, pre-cut nylon membrane, whatman filter paper and sponge were soaked in 0.5x TBE buffer. Once the gel was run, the transfer sandwich was prepared, and the sandwich was placed into the BioRad Trans-Blot Cell and filled with pre-chilled 0.5x TBE. The Trans-Blot Cell was run at 80 V for 60 min at 4 $^{\circ}\text{C}$ with continuous stirring. Then the membrane was UV crosslinked at 1200 mJ for 3 min in an UVP crosslinker. Then the membrane was transferred into blocking buffer (ThermoFisher TM pierce starting blocking buffer at room temperature for 1 h in the Shaker. Finally, the membrane was incubated with HRP-conjugated Streptavidin (ThermoFisher N100) in the 1:2000 dilution at room temperature for 2 h and then the biotinylated probe conjugated band was visualized by Western

Blotting Luminol Reagent (Cell signaling Technology #7003) in the ChemiDoc imager.

Photocrosslinking experiments with 11 in the presence of total cellular RNA.

Total RNA isolated from the MCF-7 cells was split into different tubes (5 μg RNA/condition) and was diluted up to 50 μL total volume with riboswitch buffer (50 mM Tris (pH 7.5), 100 mM KCl and 2 mM MgCl₂). One micromolar of the full-length *Bs*-preQ₁-RS (70 nts) was added to the appropriate tubes during the dilution with riboswitch buffer. Before the addition, RNA was folded as described above, by heating at 75 $^{\circ}\text{C}$ for 5 min and allowing to cool to RT over the course of 1 h. Next, 20 mM compounds 11 and 17 were prepared in DMSO and the compounds were diluted to a final concentration of 250 μM with the RNA in riboswitch buffer (5% final DMSO). Compounds 11 or 17 were mixed with RNA samples and incubated in the dark for 30 min to establish an equilibrium. After incubation, the solutions were moved to a clear plastic 96-well microtiter plate (VWR) for UV crosslinking. The control sample (Fig. 6, lane 7), that would not be irradiated, was left in the dark in the original tube. In the case of competition experiments, samples were prepared with the free PreQ₁ ligand (500 μM and 1000 μM concentrations) titrated into the solution containing compound 11 and the total RNA to allow for competition before UV crosslinking. Samples in the plate were irradiated for 15 min at 365 nm using a SPECTROLINKER XL-1000 UV Crosslinker. After UV irradiation the sample mixtures were conjugated to TAMRA azide using the click labeling protocols described above. Finally, RNA was ethanol precipitated after TAMRA conjugation and samples were analyzed by PAGE.

RNA sequencing experiments. Total RNA (5 μg) isolated from MCF-7 cells was diluted up to 50 μL total volume with riboswitch buffer (50 mM Tris (pH 7.5), 100 mM KCl and 2 mM MgCl₂). One micromolar of the full-length *Bs*-preQ₁-RS (70 nts) was added into the tubes during the dilution with riboswitch buffer. Before the addition, *Bs*-preQ₁-RS was folded as described above, by heating at 75 $^{\circ}\text{C}$ for 5 min and allowing to cool to RT over the course of 1 h. 20 mM stock solutions of compounds 11 and 17 were prepared in DMSO and the compounds were diluted to a final concentration of 250 μM with the RNA in riboswitch buffer (5% final concentration of DMSO). The compound and RNA mixtures were incubated at room temperature for 1 h and samples were UV crosslinked as described above. The modified RNA was labeled with biotin-PEG₃-azide using the click protocol described above. After labeling the RNA was subjected to ethanol precipitation to remove excess biotin-PEG₃-azide. The RNA pellets were dissolved in 50 μL of incubation buffer (50 mM Tris pH 7.5, 100 mM KCl, 0.01% Tween-20) and 50 μL of streptavidin linked magnetic beads (NEB) was added to each tube. After 30 min, the beads were pulled down on a magnetic tower and washed three times with 500 μL incubation buffer. Finally, 50 μL of elution buffer (95% formamide/H₂O, 10 mM EDTA) was added to each tube and the tubes were placed on a 95 $^{\circ}\text{C}$ heat block for 4 min. The supernatant containing the eluted RNA was removed on the magnetic tower and the solution was diluted with RNase free water and ethanol precipitated overnight. The concentration of the reconstituted RNA was measured using a Qubit RNA HS Assay Kit (Invitrogen). Two-hundred nanograms of each sample was subsequently used to prepare sequencing libraries with the QiaSeq Stranded Total RNA Library Kit (Qiagen). Completed libraries were submitted to Novogene for sequencing on a HiSeq 2500 (2 \times 150 reads). Four replicates of each treatment condition were prepared and sequenced. We followed the same protocol as describe above but without addition of the *Bs*-preQ₁-RS aptamer to assess crosslinking to other RNAs in MCF7 total RNA. All sequencing experiments were performed with four independent replicates for each condition ($n = 4$). In experiments without aptamer spike-in one replicate of 17 failed QC and was not included in this analysis.

Sequencing data analysis. For the analysis of these pulldown experiments, a reference was built by using the longest isoform for every human RefSeq transcript, plus the sequence of the *Bs*-preQ₁-RS full-length aptamer (see above). Reads were aligned to this reference using Bowtie2 and the "--very-sensitive-local" preset⁶⁹. Resulting BAM files were then processed to TSV count files, containing the number of reads mapping to each transcript. The DESeq2 package⁷⁰ (adjusted p -value < 0.05) was used to identify transcripts significantly enriched by compound 11 in the pulldown with respect to the control (compound 17).

For the sequencing experiments without the aptamer spiked-in, the volcano plot generated by the DESeq2 package was used to identify potential hits that were significantly enriched (FC > 1.5-fold, adjusted p -value < 0.05 when comparing 11 treated samples to 17 treated samples).

Quantitative reverse transcription PCR (RT-qPCR). Samples for these experiments were prepared using the same protocol as described above for the RNA-sequencing experiments (without spiking *Bs*-preQ₁-RS aptamer). However, in this case extra samples were prepared with the free PreQ₁ ligand (25 μM and 100 μM concentrations) titrated into the solution containing compound 11 and the total RNA to allow for competition before UV crosslinking. RNA concentration was determined by using a Qubit RNA HS Assay Kit (Invitrogen), then 200 ng of RNA was used for cDNA synthesis with qScript cDNA SuperMix (Quanta Biosciences) according to the manufacturer's protocol. The 20 μL

reactions were incubated in a thermocycler (Bio-rad) for 5 min at 22 °C, 30 min at 42 °C, 5 min at 85 °C, then held at 4 °C. Then 20 ng of cDNA was subjected to qPCR using a Perfecta SYBR Green Super Mix (Quanta Biosciences) on an Eppendorf Mastercycler RealPlex2 in the presence of the appropriate set of primers. The reactions were incubated at 95 °C for 10 min, followed by 40 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 20 s. Threshold cycles (C_T) is defined as the first cycle that showed a detectable increase in fluorescence due to the formation of PCR products and was used to determine the template amount in each sample. The relative fold change in expression was measured using the Livak method⁷¹ and were normalized to the relative level of RNAs of the control experiment (compound 17). For example, to calculate the $\Delta\Delta(C_T)$ between each gene of interest and the average of the control samples: $\Delta(C_T) = C_T$ (target gene) – C_T (Gapdh); $\Delta\Delta(C_T) = \Delta C_T$ (treatment or knockdown) – ΔC_T (control); fold change = $2^{-\Delta\Delta(C_T)}$. The primers used in these experiments are shown in the Supplementary Information Table S2.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Atomic coordinates and the structure factors of the co-crystal structures of ab13_14 and the wild-type *Tt*-preQ₁-RS aptamer domain in complex with **11** have been deposited in the Protein Data Bank, under the accession codes 7E9E and 7E9I, respectively. Previously published crystals 3Q50, 6E1W and 3B1V were accessed from the PDB and used for reference. Raw fastq files along with processed alignment files and counts files from these sequencing experiments have been deposited on the NCBI's gene expression omnibus (GEO) database. These files can be accessed with the following GEO accession code: GSE168353. The source data underlying Fig. 2A–E, 4C–D, 5A–B, 7B–D and Supplementary Figs. S1, S2, S5, S6A–B, S7A–D and S10 are provided as a Source Data file. All other relevant data are available from the authors upon request Source data are provided with this paper.

Received: 12 March 2021; Accepted: 26 August 2021;

Published online: 06 October 2021

References

- Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003566 (2012).
- Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17–24 (2013).
- Roth, A. & Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78**, 305–34 (2009).
- Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–7 (2015).
- Warner, K. D. et al. Validating fragment-based drug discovery for biological RNAs: lead fragments bind and remodel the TPP riboswitch specifically. *Chem. Biol.* **21**, 591–5 (2014).
- Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N. & Breaker, R. R. Antibacterial lysine analogs that target lysine riboswitches. *Nat. Chem. Biol.* **3**, 44–9 (2007).
- McCown, P. J., Liang, J. J., Weinberg, Z. & Breaker, R. R. Structural, functional, and taxonomic diversity of three preQ₁ riboswitch classes. *Chem. Biol.* **21**, 880–889 (2014).
- Roth, A. et al. A riboswitch selective for the queuosine precursor preQ₁ contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol.* **14**, 308–17 (2007).
- Batey, R. T. Riboswitches: still a lot of undiscovered country. *RNA* **21**, 560–3 (2015).
- Wang, X. et al. Queuosine modification protects cognate tRNAs against ribonuclease cleavage. *RNA* **24**, 1305–1313 (2018).
- Alexander, S. C., Busby, K. N., Cole, C. M., Zhou, C. Y. & Devaraj, N. K. Site-specific covalent labeling of RNA by enzymatic transglycosylation. *J. Am. Chem. Soc.* **137**, 12756–9 (2015).
- Zhang, D., Zhou, C. Y., Busby, K. N., Alexander, S. C. & Devaraj, N. K. Light-activated control of translation by enzymatic covalent mRNA labeling. *Angew. Chem. Int. Ed. Engl.* **57**, 2822–2826 (2018).
- Wu, M. C. et al. Rational re-engineering of a transcriptional silencing PreQ₁ riboswitch. *J. Am. Chem. Soc.* **137**, 9015–21 (2015).
- Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–6 (2006).
- Tijerina, P., Mohr, S. & Russell, R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat. Protoc.* **2**, 2608–23 (2007).
- Kadina, A., Kietrys, A. M. & Kool, E. T. RNA cloaking by reversible acylation. *Angew. Chem. Int. Ed. Engl.* **57**, 3059–3063 (2018).
- Ayele, T. M., Loya, T., Valdez-Sinon, A. N., Bassell, G. J. & Heemstra, J. M. *Imaging and Tracking mRNA in Live Mammalian Cells Via Fluorogenic Photoaffinity Labeling*. (Cold Spring Harbor Laboratory, 2020).
- Zhou, C. Y., Alexander, S. C. & Devaraj, N. K. Fluorescent turn-on probes for wash-free mRNA imaging. *Chem. Sci.* **8**, 7169–7173 (2017).
- Wang, J., Schultz, P. G. & Johnson, K. A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl Acad. Sci. USA* **115**, E4604–E4612 (2018).
- Mortison, J. D. et al. Tetracyclines modify translation by targeting key human rRNA substructures. *Cell Chem. Biol.* **25**, 1506–1518.e13 (2018).
- Suresh, B. M. et al. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl Acad. Sci. USA* **117**, 33197–33203 (2020).
- Mukherjee, H. et al. PEARL-seq: a photoaffinity platform for the analysis of small molecule-RNA interactions. *ACS Chem. Biol.* **15**, 2374–2381 (2020).
- Hargrove, A. E. Small molecule-RNA targeting: starting with the fundamentals. *Chem. Commun. (Camb.)* **56**, 14744–14756 (2020).
- Connelly, C. M., Moon, M. H. & Schneekloth, J. S. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem. Biol.* **23**, 1077–1090 (2016).
- Meyer, S. M. et al. Small molecule recognition of disease-relevant RNA structures. *Chem. Soc. Rev.* **49**, 7167–7199 (2020).
- Baker, J. L. et al. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**, 233–235 (2012).
- Sarkar, B., Ishii, K. & Tahara, T. Microsecond folding of preQ₁ riboswitch and its biological significance revealed by two-dimensional fluorescence lifetime correlation spectroscopy. *J. Am. Chem. Soc.* **143**, 7968–7978 (2021).
- Klein, D. J., Edwards, T. E. & Ferré-D'Amaré, A. R. Cocystal structure of a class I preQ₁ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* **16**, 343–344 (2009).
- Connelly, C. M. et al. Synthetic ligands for PreQ. *Nat. Commun.* **10**, 1501 (2019).
- Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V. & Wedekind, J. E. Comparison of a preQ₁ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* **286**, 24626–37 (2011).
- Schroeder, G. M. et al. Analysis of a preQ₁-I riboswitch in effector-free and bound states reveals a metabolite-programmed nucleobase-stacking spine that controls gene regulation. *Nucleic Acids Res.* **48**, 8146–8164 (2020).
- Santner, T., Rieder, U., Kreutz, C. & Micura, R. Pseudoknot preorganization of the preQ₁ class I riboswitch. *J. Am. Chem. Soc.* **134**, 11928–31 (2012).
- Eichhorn, C. D., Kang, M. & Feigon, J. Structure and function of preQ. *Biochim. Biophys. Acta* **1839**, 939–950 (2014).
- Dutta, D. & Wedekind, J. E. Nucleobase mutants of a bacterial preQ. *J. Biol. Chem.* **295**, 2555–2567 (2020).
- McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).
- Guan, L. & Disney, M. D. Covalent small-molecule-RNA complex formation enables cellular profiling of small-molecule-RNA interactions. *Angew. Chem. Int. Ed. Engl.* **52**, 10010–3 (2013).
- Behm-Ansmant, I., Helm, M. & Motorin, Y. Use of specific chemical reagents for detection of modified nucleotides in RNA. *J. Nucleic Acids* **2011**, 1–17 (2011).
- Neuner, E., Frener, M., Lusser, A. & Micura, R. Superior cellular activities of azido- over amino-functionalized ligands for engineered preQ₁ riboswitches in *E. coli*. *RNA Biol.* **15**, 1376–1383 (2018).
- Li, Z. et al. Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew. Chem. Int. Ed.* **52**, 8551–8556 (2013).
- Gao, J., Mfuh, A., Amako, Y. & Woo, C. M. Small molecule interactome mapping by photoaffinity labeling reveals binding site hotspots for the NSAIDs. *J. Am. Chem. Soc.* **140**, 4259–4268 (2018).
- Brunner, J., Senn, H. & Richards, F. M. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J. Biol. Chem.* **255**, 3313–8 (1980).
- Dubinsky, L., Krom, B. P. & Meijler, M. M. Diazirine based photoaffinity labeling. *Bioorg. Med. Chem.* **20**, 554–70 (2012).
- Artsimovitch, I. & Henkin, T. M. In vitro approaches to analysis of transcription termination. *Methods* **47**, 37–43 (2009).
- Williams, A. S. & Marzluff, W. F. The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. *Nucleic Acids Res.* **23**, 654–662 (1995).
- Dávila López, M. & Samuelsson, T. Early evolution of histone mRNA 3' end processing. *RNA* **14**, 1–10 (2008).
- Maida, Y. et al. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **461**, 230–235 (2009).

47. Su, Z. et al. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **84**, 239 (2014).
48. Chee, G. L., Yalowich, J. C., Bodner, A., Wu, X. & Hasinoff, B. B. A diazirine-based photoaffinity epitope probe for labeling topoisomerase II. *Bioorg. Med. Chem.* **18**, 830–8 (2010).
49. West, A. et al. Labeling preferences of diazirines with protein biomolecules. *J. Am. Chem. Soc.* **143**, 6691–6700 (2020).
50. Fergus, C., Barnes, D., Alqasem, M. A. & Kelly, V. P. The queuine micronutrient: charting a course from microbe to man. *Nutrients* **7**, 2897–929 (2015).
51. Vinayak, M. & Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci. Rep.* **30**, 135–48 (2009).
52. Tuorto, F. et al. Queuosine-modified tRNAs confer nutritional control of protein translation. *EMBO J.* **37**, e99777 (2018).
53. Xu, D. et al. PreQ0 base, an unusual metabolite with anti-cancer activity from *Streptomyces qinglanensis* 172205. *Anticancer Agents Med. Chem.* **15**, 285–90 (2015).
54. Zhang, J. et al. tRNA Queuosine modification enzyme modulates the growth and microbiome recruitment to breast tumors. *Cancers (Basel)* **12**, 628 (2020).
55. Artandi, S. E. & Depinho, R. A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9–18 (2010).
56. Baena-Del Valle, J. A. et al. MYC drives overexpression of telomerase RNA (hTR/TERC) in prostate cancer. *J. Pathol.* **244**, 11–24 (2018).
57. Cherkaoui Jaouad, I. et al. Novel mutation and structural RNA analysis of the noncoding RNase MRP gene in cartilage-hair hypoplasia. *Mol. Syndromol.* **6**, 77–82 (2015).
58. Zhang, Q., Kim, N.-K. & Feigon, J. Architecture of human telomerase RNA. *Proc. Natl Acad. Sci. USA* **108**, 20325–20332 (2011).
59. Jafri, M. A., Ansari, S. A., Alqahtani, M. H. & Shay, J. W. Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med.* **8**, 69 (2016).
60. Klepper, F., Polborn, K. & Carell, T. Robust synthesis and crystal-structure analysis of 7-cyano-7-deazaguanine (PreQ₀ Base) and 7-(aminomethyl)-7-deazaguanine (PreQ₁ Base). *Helvetica Chim. Acta* **88**, 2610–2616 (2005).
61. Walzthoeni, T. et al. xTract: software for characterizing conformational changes of protein complexes by quantitative cross-linking mass spectrometry. *Nat. Methods* **12**, 1185–1190 (2015).
62. Fantoni, N. Z., El-Sagheer, A. H. & Brown, T. A Hitchhiker's guide to click-chemistry with nucleic acids. *Chem. Rev.* **121**, 7122–7154 (2021).
63. Hirata, K. et al. ZOO: an automatic data-collection system for high-throughput structure analysis in protein microcrystallography. *Acta Crystallogr. D. Struct. Biol.* **75**, 138–150 (2019).
64. Yamashita, K., Hirata, K. & Yamamoto, M. KAMO: towards automated data processing for microcrystals. *Acta Crystallogr. D. Struct. Biol.* **74**, 441–449 (2018).
65. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010).
66. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
67. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010).
68. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–32 (2004).
69. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
71. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* **25**, 402–8 (2001).

Acknowledgements

The synchrotron-radiation experiments at SPring-8 (Hyogo, Japan) were performed on beamlines BL32XU and BL45XU with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) under proposal numbers of 2019B2580 and 2020A2568. We also thank the beam-line staff at BL-1A and BL-17A of the Photon Factory (Ibaraki, Japan) for technical assistance during data collection. This research was supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. Project numbers Z01 BC011585 07 (PI, J. S. Schneekloth, Jr). This work was supported in part by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers 16KK0166 and 20K21281) and a grant from The Uehara Memorial Foundation to T.N. D.I. was supported by the Dutch Research Council (Netherlands Organisation for Scientific Research, NWO) as part of the research programme NWO Open Competitie ENW-XS (project number OCENW.XS3.044), and by the Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen.

Author contributions

S.B. performed photocrosslinking experiments, qPCR, and helped write and edit the paper. C.R. performed photocrosslinking and Chem-CLIP experiments, and helped write and edit the paper. D.D.B. performed chemical synthesis and characterization and mass spectral analyses. C.C. performed photocrosslinking and mass spectral analyses. C.C.L. and J.A.K. performed mass spectral analyses. K.Y., P.J.H., and D.I. performed bioinformatic analyses. T.N. performed X-ray crystallographic analyses and transcription termination assays. J.S.S. conceived the project and helped write and edit the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25973-x>.

Correspondence and requests for materials should be addressed to John S. Schneekloth Jr.

Peer review information *Nature Communications* thanks Joseph Wedekind and other, anonymous, reviewers for their contributions to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2021