A Chemical Probe Based on the PreQ1 Metabolite Enables Transcriptome-wide Mapping of Binding Sites

Sumirtha Balaratnam
National Cancer Institute
Curran Rhodes
National Cancer Institute
Desta Bume
National Cancer Institute
Colleen Connelly
National Cancer Institute https://orcid.org/0000-0003-1671-3243
Christopher Lai
National Cancer Institute
James Kelley
National Cancer Institute
Kamyar Yazdani
National Cancer Institute
Phil Homan
National Cancer Institute
Danny Incarnato
University of Groningen https://orcid.org/0000-0003-3944-2327
Tomoyuki Numata
National Heart Lung and Blood Institute
Jay Schneekloth (schneeklothjs@mail.nih.gov)
National Cancer Institute https://orcid.org/0000-0001-7459-783X

Article

Keywords: mapping, RNA, binding

Posted Date: March 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-322220/v1

Version of Record: A version of this preprint was published at Nature Communications on October 6th, 2021. See the published version at https://doi.org/10.1038/s41467-021-25973-x.

Abstract

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 riboswitch aptamer covalently. For the most active probe, a diazirine-based photocrosslinker, 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 small molecule. In addition, a small number of RNA targets in endogenous human transcripts were found to bind specifically to PreQ1, 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 interacts with human RNAs.

Introduction

Riboswitches are naturally occurring RNA sequences that influence bacterial gene expression by binding directly to small molecules[1–3]. Because they can bind tightly 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 hypermodified nucleotide 7-aminomethyl-7-deazaguanine (known as PreQ1)[7]. Upon recognition of PreQ1, the RNA changes conformation and induces transcription termination, resulting in altered gene expression. 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 elements[8]. 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)[9]. PreQ1 has also been utilized in a series of intriguing studies aimed at enzymatic tagging of complex RNAs[10,11]. Due 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 applications[12]. A full understanding of the interactions of PreQ1 with the human transcriptome is needed to better understand how useful PreQ1 would be as a chemical biology probe. 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[13,14], controlling gene expression[15], imaging RNAs[16,17], tagging RNAs with functional handles, and demonstrating target engagement for RNA-binding small molecules[18–21]. 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[22–24]. Understanding the chemical features that impact RNA crosslinking efficiency would influence probe design, as well 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 chloride[25], potentially representing a challenge in designing probes. 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 PreQ1 aptamer represents an ideal system for this study due to its relatively small size and role in governing gene expression in bacteria. For example, the PreQ1 aptamer from *Bacillus subtilis* (*Bs*) consists of just 34 nucleotides and has an equilibrium dissociation constant of 2 nM for PreQ1[7]. In addition to this tight affinity, the *Bs* aptamer has been studied extensively using both X-Ray crystallography and NMR, providing detailed information about the mode of ligand binding[26–28]. Other studies on this system include how it responds to other natural metabolites, engineering to remodel ligand specificity[12,29,30], and investigations for inhibitor design[31]. PreQ1 riboswitches from other organisms have been studied in detail as well, revealing multiple distinct classes of independently evolved riboswitches that function by different mechanisms[7]. In fact, PreQ1 riboswitches are among the most frequently occurring riboswitches in Nature[32]. This example of convergent evolution demonstrates remarkably high diversity of RNA sequences that recognize PreQ1 with high affinity and selectivity. Given the extensive work reported on the PreQ1 aptamer, its potential for use in chemical biology and synthetic biology applications, the role of PreQ1 in influencing gene expression, and frequency of natural aptamers for PreQ1, efforts to probe the binding of this metabolite to human RNAs are warranted.

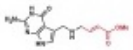
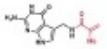
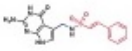
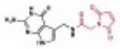
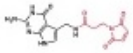
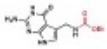
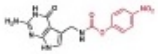
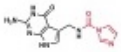
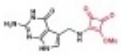
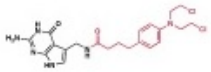
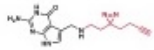
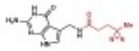
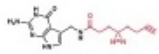
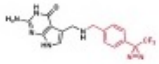
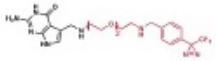
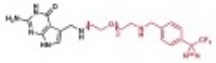
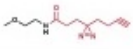
Here, we report the design and evaluation of a series of covalent probes for the PreQ1 riboswitch. In total, a series of 16 probes are reported, including both electrophilic and photochemically activated crosslinkers. Gel-based and mass spectrometry experiments are conducted to quantify crosslinking 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 And Discussion

To design covalent probes targeting the PreQ1 riboswitch aptamer, 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[26–29, 31]. 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[26]. 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[33].

Table 1. Structures and crosslinking efficiencies for electrophilic and photoaffinity probes used for proximity induced covalent modification of the PreQ1 aptamer

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		31.3 ± 2.9
12		49.2 ± 3.2
13		39.9 ± 5.3
14		Trace
15		22.0 ± 2.2
16		19.4 ± 4.4
17		Not Determined

The reactivity of each of these probes towards PreQ1-riboswitch aptamers was evaluated using mass spectrometry (e.g. MALDI-TOF) and denaturing PAGE experiments. Initial in vitro analysis of the reaction mixtures between PreQ1 riboswitch 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 Figure 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 to achieve alkylation. To overcome this challenge, we turned our attention towards the synthesis and evaluation of more reactive, diazirine-containing photoaffinity crosslinkers.

Similar to electrophilic probes, photoaffinity probes are commonly employed for labeling studies with proteins[34,35]. Diazirine functionalized nucleosides have been used previously to map nucleic acid-ligand interactions or conjugate dyes to aptamers[18–21]. However, typically only one probe was reported, and the effect of linker structure or crosslinking group was not broadly evaluated. In order to determine whether photoaffinity linker structure influences crosslinking efficiency, a series of photoaffinity probes were designed and synthesized (Scheme 1, Table 1). Additional details for the synthesis and characterization of these compounds are provided in the supplementary information (Figure S11). Included in this series are probes varied in crosslinker structure (stabilized and unstabilized diazirines) as well as linker length[36,37]. The *Staphylococcus saprophyticus* PreQ1 aptamer (Ss-PreQ1-RS) was treated with 50-fold excess of probe in the presence of 1mM $MgCl_2$, irradiated with 365 nm light sources for 15 minutes, and PAGE analysis was used to assess the formation of crosslinks. Gratifyingly, the expected adducts was observed in all cases, however efficiency appeared to be low for **14**. In agreement with mass spectral data, only trace crosslinking was observed with **14**. Crosslinking efficiencies were above 30% for unstabilized probes **11**, **12** and **13**, and approximately 20% with stabilized probes **15** and **16** (Supplementary information Figure S2).

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. No crosslinking was detected for **14** 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 time and dose-dependent crosslinking was observed (Fig. 1A, 1B 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.

With an optimized probe in hand, we next explored whether the photoaffinity labeling with compound **11** achieves proximity induced site-specificity or 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. 1C). 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. 1D). We also performed labelling 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. 1E).

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-LC/MS and we identified guanosine-compound **11** adduct as a primary modified species (Fig. 2A and 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. 2C, D and 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 near the binding site of the ligand in crystal structures[26].

To further validate the binding of **11** to the PreQ1 aptamer, we performed X-ray crystallography with the PreQ1 aptamer from *Thermoanaerobacter tengcongensis* (*Tt*), using a previously reported method[31]. Co-crystals were obtained using wild type (WT) and an abasic mutant at positions 13 and 14 (ab_13–14), and the complex structures were solved by the molecular replacement method at 2.80 and 1.57 Å resolution respectively (Fig. 3A and B, Figure S4, and Table S1). These co-crystal structures exhibit a similar mode of binding to **11**. Because of its high-resolution structure, here we describe the interactions between ab_13–14 and **11** in detail (Fig. 3A and B). From inspection of the structure, it is clear that the PreQ1 portion of **11** binds to the *Tt* aptamer in an analogous mode to the native ligand, making near identical contacts[30]. The sidechain of the probe projects out of the binding pocket toward the solvent and is proximal to two guanosines that it could potentially modify upon irradiation. Thus, it is clear that modification of the exocyclic amine of PreQ1 does not dramatically alter its mode of binding to RNA.

Next, single round transcription termination assays were performed to assess the function of **11**. We used a previously reported protocol to assess the activity of both the cognate ligand and the photocrosslinking probe[31,38]. Gratifyingly, the photocrosslinking probe induced transcriptional termination to a degree similar to the unmodified compound. The unmodified ligand achieved a maximal termination efficiency of ~ 85% with a EC_{50} of 7.4 nM, while the photocrosslinker achieved a maximal termination efficiency of ~ 65% with a EC_{50} of 7.1 μM (Fig. 3C, Figure S5), demonstrating that the probe functions similar to the cognate ligand albeit requiring a higher concentration. Since the termination assays were performed without irradiation, these experiments confirm that the presence of the photocrosslinker sidechain does not ablate the inhibitory activity of the ligand.

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[16,18–21]. To do this, we synthesized biotin-conjugated **11** (bio-**11**) using copper catalyzed click chemistry with biotin-PEG_3-azide and performed photocrosslinking experiments under our optimized reactions conditions. MALDI-TOF analysis of the reaction mixture showed the expected molecular weight adducts (Supplementary information Figure S6). Concurrently, the photocrosslinked Ss-PreQ1-RS aptamer was also subjected to copper catalyzed click reaction with biotin-PEG-azide and Cy5-picolyl azide and the expected click product was observed in both experiments by mass spectrometry and denaturing PAGE analysis (Supplementary information Figure S7 A and B and S8).

Bio-**11** was prepared and added to MCF-7 cell lysates that had been supplemented with Ss-PreQ1 aptamer. After irradiation, the lysates were analyzed by denaturing PAGE. As shown in Fig. 4A, when the gel was stained with SYBR gold stain, a crosslinked higher molecular weight band could be 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. 4A, 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 1μM 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. 4B, 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. 4B, lane 3). In contrast, as shown in lane 4 of Fig. 4B, 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. 4B, 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. 4B, 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 Figure S9).

Chem-CLIP experiments were also performed to further probe the selectivity of **11** for the full-length BS-PreQ1 aptamer in the presence of MCF-7 derived total cellular RNA[33]. After UV crosslinking in the presence of **11**, total RNA was click labeled with biotin-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_2fold-change > 2.5, adjusted p-value = $2.9e^{-21}$(Fig. 5A)). As evident in Fig. 5B 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.

o further investigate cross-reactivity of 11 with other 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_2fold-change > 0.95, -log_{10}P > 4) (Fig. 6A and supplementary information S10). 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. Due 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[39,40]. 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)[41]. 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 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 probes[42] (Fig. 6B, C and 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 competable 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 mM) (Fig. 6D. 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.

Conclusion

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 guanines in the aptamer, two of which are proximal to the binding site, this observation confirms a highly specific crosslinking event[26]. 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[36,37]. 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[43,44]).

Importantly, 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. 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 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**.

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[9,45,46]. 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[47,48]. Here, we demonstrate that PreQ1 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)[40]. Additionally, the other two enriched genes, TERC and RMRP are both highly conserved short, structured noncoding RNAs that are known to play roles in disease pathogenesis[49–51]. 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 hTERT[52]. Telomerase activity is low in the majority of healthy somatic cells, however, it is significantly increased in various cancers[49], and TERC has been implicated as a potential therapeutic target in many of these cancers[53]. 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 diazirine 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 a high selectivity between **11** and the PreQ1 aptamer in cell lysates and total RNA, indicating the potential to use this system in chemical biology applications. Finally, mapping PreQ1 binding events in human transcriptomes revealed several specific binding events to regulatory RNAs. Further investigation of these and other interactions may prove useful to interrogate how metabolites interact with human RNAs and provide insights into new therapeutic strategies.

Declarations

Acknowledgments

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.

Funding

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.

References

1. Breaker, R.R. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4**(2012).
2. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24 (2013).
3. Roth, A. & Breaker, R.R. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* **78**, 305-34 (2009).
4. Howe, J.A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672-7 (2015).
5. 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).
6. 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).
7. 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).
8. Batey, R.T. Riboswitches: still a lot of undiscovered country. *RNA* **21**, 560-3 (2015).
9. Wang, X. et al. Queuosine modification protects cognate tRNAs against ribonuclease cleavage. *RNA* **24**, 1305-1313 (2018).
10. 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).
11. 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).
12. Wu, M.C. et al. Rational Re-engineering of a Transcriptional Silencing PreQ1 Riboswitch. *J Am Chem Soc* **137**, 9015-21 (2015).
13. 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).
14. Tijerina, P., Mohr, S. & Russell, R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc* **2**, 2608-23 (2007).
15. Kadina, A., Kietrys, A.M. & Kool, E.T. RNA Cloaking by Reversible Acylation. *Angew Chem Int Ed Engl* **57**, 3059-3063 (2018).
16. 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).
17. Zhou, C.Y., Alexander, S.C. & Devaraj, N.K. Fluorescent turn-on probes for wash-free mRNA imaging. *Chem Sci* **8**, 7169-7173 (2017).
18. Wang, J., Schultz, P.G. & Johnson, K.A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc Natl Acad Sci U S A* **115**, E4604-E4612 (2018).

19. Mortison, J.D. et al. Tetracyclines Modify Translation by Targeting Key Human rRNA Substructures. *Cell Chem Biol* **25**, 1506-1518.e13 (2018).
20. Suresh, B.M. et al. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc Natl Acad Sci U S A* **117**, 33197-33203 (2020).
21. Mukherjee, H. et al. PEARL-seq: A Photoaffinity Platform for the Analysis of Small Molecule-RNA Interactions. *ACS Chem Biol* **15**, 2374-2381 (2020).
22. Hargrove, A.E. Small molecule-RNA targeting: starting with the fundamentals. *Chem Commun (Camb)* **56**, 14744-14756 (2020).
23. 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).
24. Meyer, S.M. et al. Small molecule recognition of disease-relevant RNA structures. *Chem Soc Rev* **49**, 7167-7199 (2020).
25. Baker, J.L. et al. Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride. *Science* **335**, 233-235 (2012).
26. 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).
27. Santner, T., Rieder, U., Kreutz, C. & Micura, R. Pseudoknot preorganization of the preQ1 class I riboswitch. *J Am Chem Soc* **134**, 11928-31 (2012).
28. Eichhorn, C.D., Kang, M. & Feigon, J. Structure and function of preQ. *Biochim Biophys Acta* **1839**, 939-950 (2014).
29. Dutta, D. & Wedekind, J.E. Nucleobase mutants of a bacterial preQ. *J Biol Chem* **295**, 2555-2567 (2020).
30. 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).
31. Connelly, C.M. et al. Synthetic ligands for PreQ. *Nat Commun* **10**, 1501 (2019).
32. McCown, P.J., Corbino, K.A., Stav, S., Sherlock, M.E. & Breaker, R.R. Riboswitch diversity and distribution. *RNA* **23**, 995-1011 (2017).
33. 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).
34. 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. *Angewandte Chemie International Edition* **52**, 8551-8556 (2013).
35. 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).
36. 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).

37. Dubinsky, L., Krom, B.P. & Meijler, M.M. Diazirine based photoaffinity labeling. *Bioorg Med Chem* **20**, 554-70 (2012).
38. Artsimovitch, I. & Henkin, T.M. In vitro approaches to analysis of transcription termination. *Methods* **47**, 37-43 (2009).
39. 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 Research* **23**, 654-662 (1995).
40. Dávila López, M. & Samuelsson, T. Early evolution of histone mRNA 3' end processing. *RNA* **14**, 1-10 (2008).
41. Maida, Y. et al. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature* **461**, 230-235 (2009).
42. 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).
43. Chee, G.L., Yalowich, J.C., Bodner, A., Wu, X. & Hasinoff, B.B. A diazirine-based photoaffinity etoposide probe for labeling topoisomerase II. *Bioorg Med Chem* **18**, 830-8 (2010).
44. West, A. et al. Labeling Preferences of Diazirines with Protein Biomolecules. (ChemRxiv, 2020).
45. 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).
46. Vinayak, M. & Pathak, C. Queuosine modification of tRNA: its divergent role in cellular machinery. *Biosci Rep* **30**, 135-48 (2009).
47. 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).
48. Zhang, J. et al. tRNA Queuosine Modification Enzyme Modulates the Growth and Microbiome Recruitment to Breast Tumors. *Cancers (Basel)* **12**(2020).
49. Artandi, S.E. & Depinho, R.A. Telomeres and telomerase in cancer. *Carcinogenesis* **31**, 9-18 (2010).
50. Baena-Del Valle, J.A. et al. MYC drives overexpression of telomerase RNA (hTR/TERC) in prostate cancer. *The Journal of Pathology* **244**, 11-24 (2018).
51. Cherkaoui Jaouad, I. et al. Novel Mutation and Structural RNA Analysis of the Noncoding RNase *MRP* Gene in Cartilage-Hair Hypoplasia. *Molecular Syndromology* **6**, 77-82 (2015).
52. Zhang, Q., Kim, N.-K. & Feigon, J. Architecture of human telomerase RNA. *Proceedings of the National Academy of Sciences* **108**, 20325-20332 (2011).
53. 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 Medicine* **8**(2016).

Figures

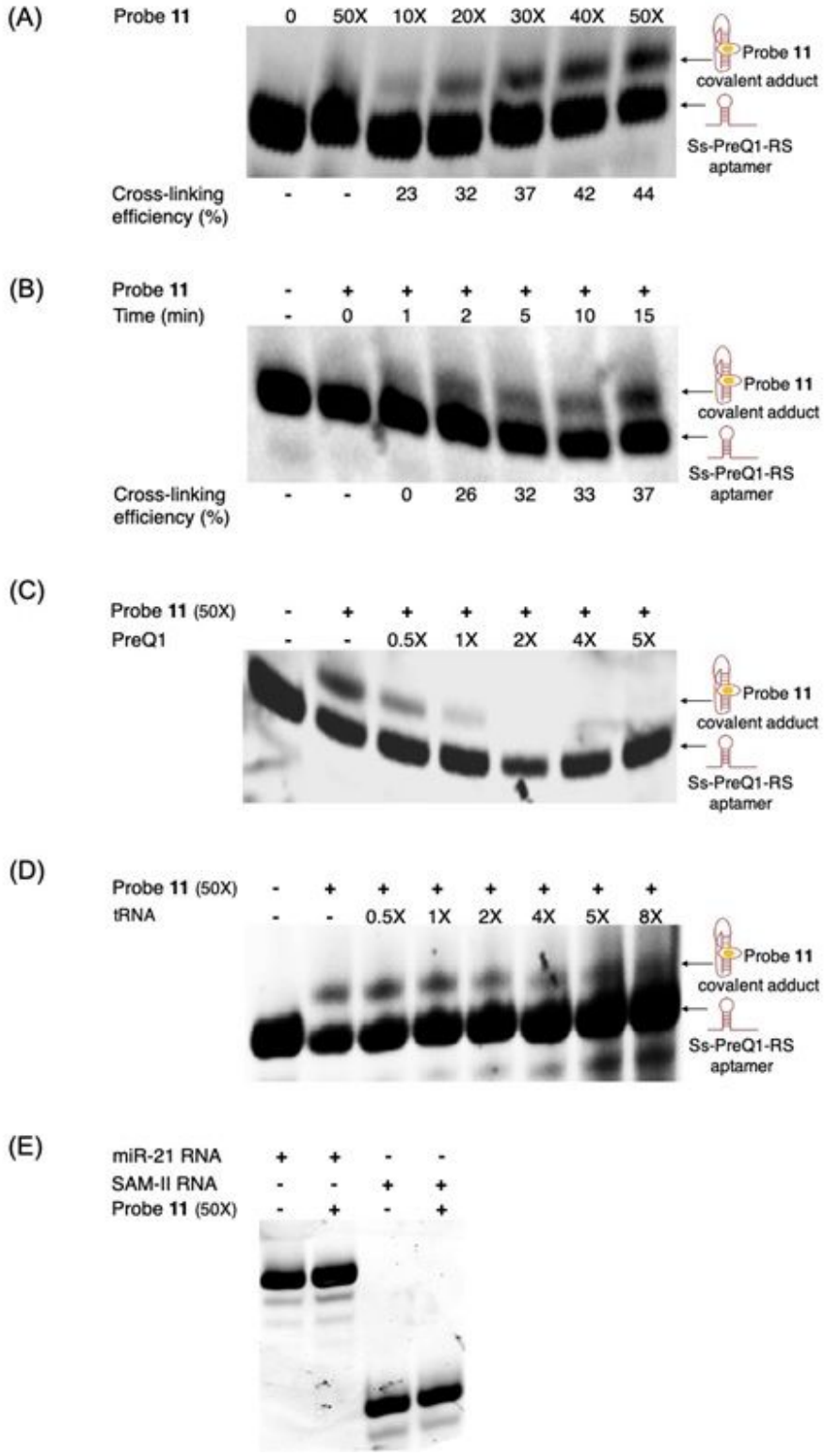


Figure 1

Biochemical optimization of photocrosslinking and competition experiments to evaluate the specificity of compound 11 for the PreQ1 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.

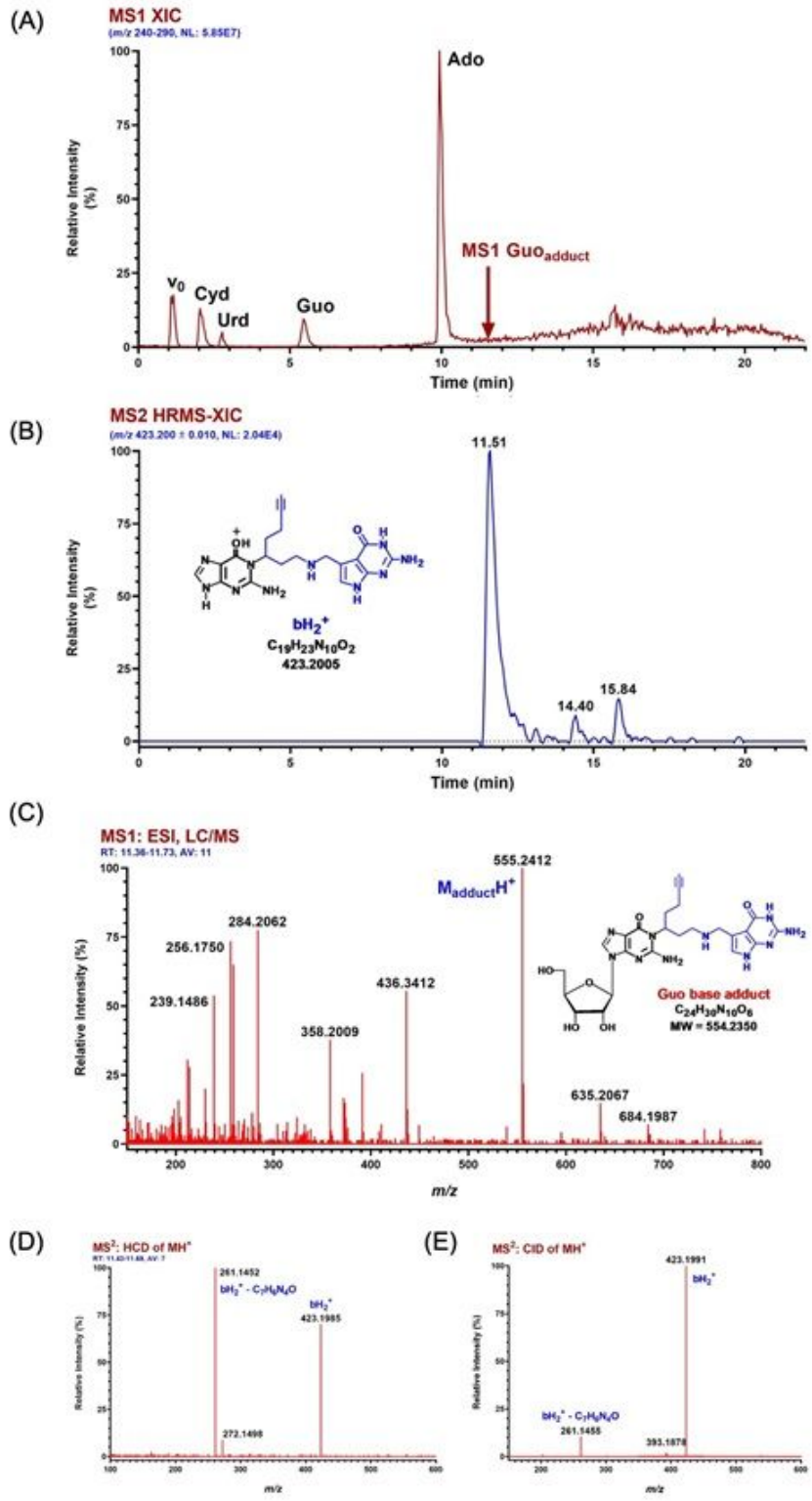


Figure 2

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 SsPreQ1-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.

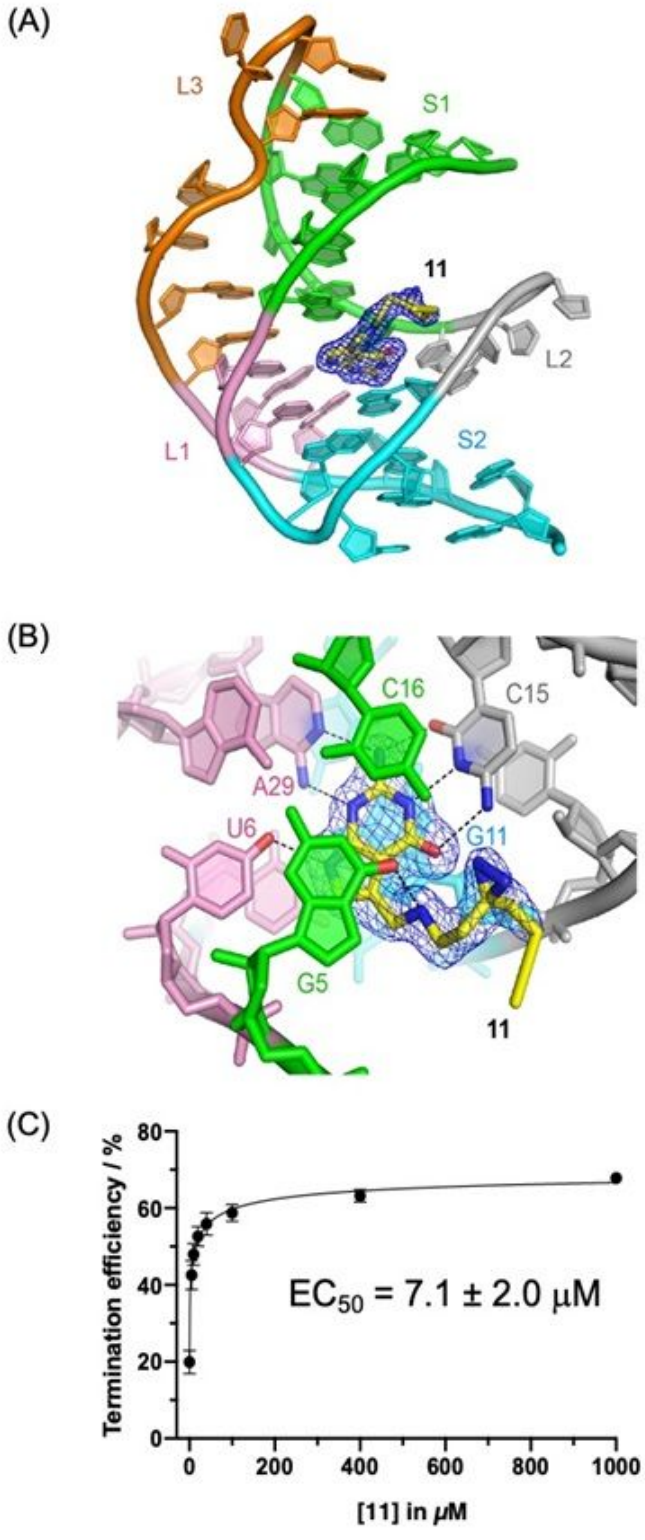


Figure 3

X-Ray co-crystal structure and functional assessment of 11. (A) Co-crystal structure showing interaction of compound 11 with the Tt-PreQ1-riboswitch (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 nucleotides between L3 and S2, which interact with L1, are also in pink. The |Fo| – |Fc| electron density maps for the compound are colored blue and contoured at 3.0 (B) Detail of ligand:RNA interaction showing conservation of contacts seen with unmodified PreQ1. The nucleotides interact with the compound are labeled, and hydrogen bonds are indicated as dotted lines. Residues at the site of the interaction are labelled to show the proximity of G5 and G11 to the diazirine. (C) Termination assay results for compound 11. Quantification of transcription termination efficiency with increasing concentrations of 11. Error bars indicate standard deviation, n = 3.

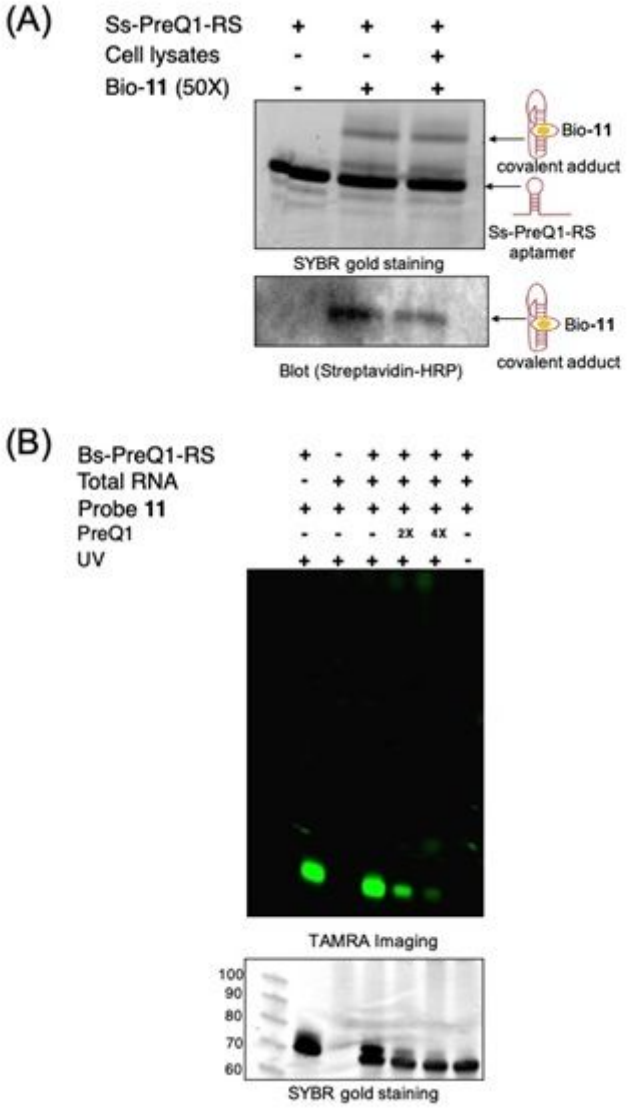


Figure 4

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 (1 uM) spiked into MCF-7 total RNA and imaged by TAMRA labeling (top) and SYBR gold stain (bottom).

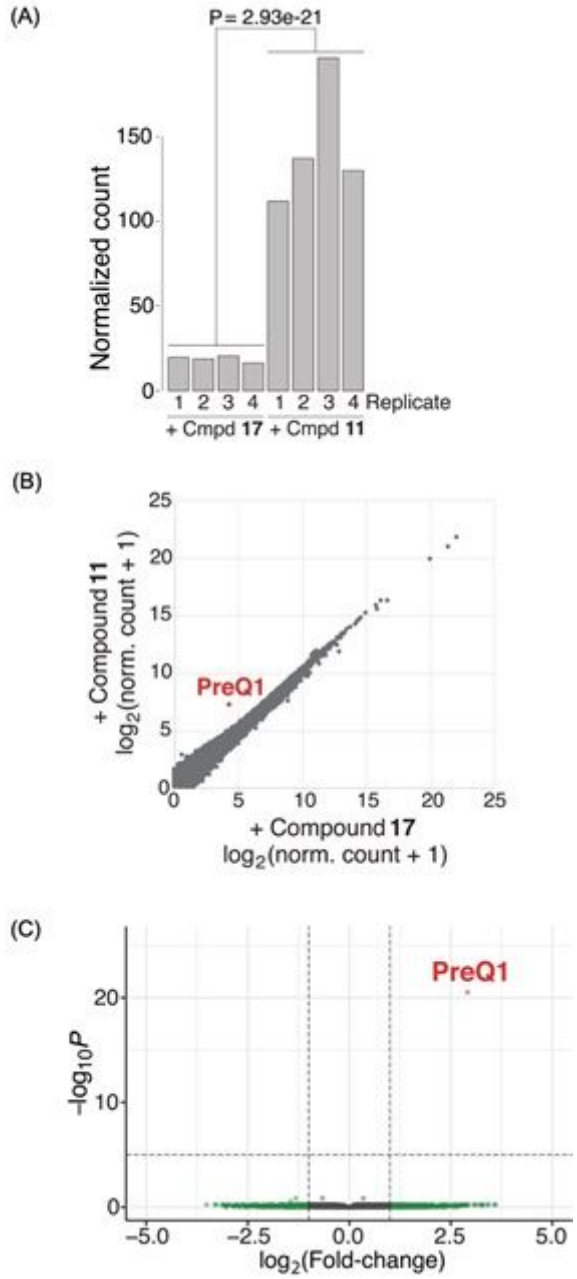


Figure 5

Selective enrichment of Bs-PreQ1 aptamer by 11 in total RNA with the BS-PreQ1 aptamer spiked in confirmed the specificity of the interaction. (A) Comparison of the normalized counts of the BS-PreQ1 aptamer across 11 and 17 treated samples shows significantly higher counts in all 4 replicates of 11 treated samples. P = adjusted p-value. (B) Correlation plot showing differential gene enrichment for compounds 11 and 17. The gene entry corresponding to the BS-PreQ1-RS is labelled in red. (C) Volcano plot from differential expression analysis between 11 and 17. The gene entry corresponding to the BS-PreQ1-RS is labelled in red. P = adjusted p-value. All analyses performed on 4 independent replicate samples treated with 11 or 17

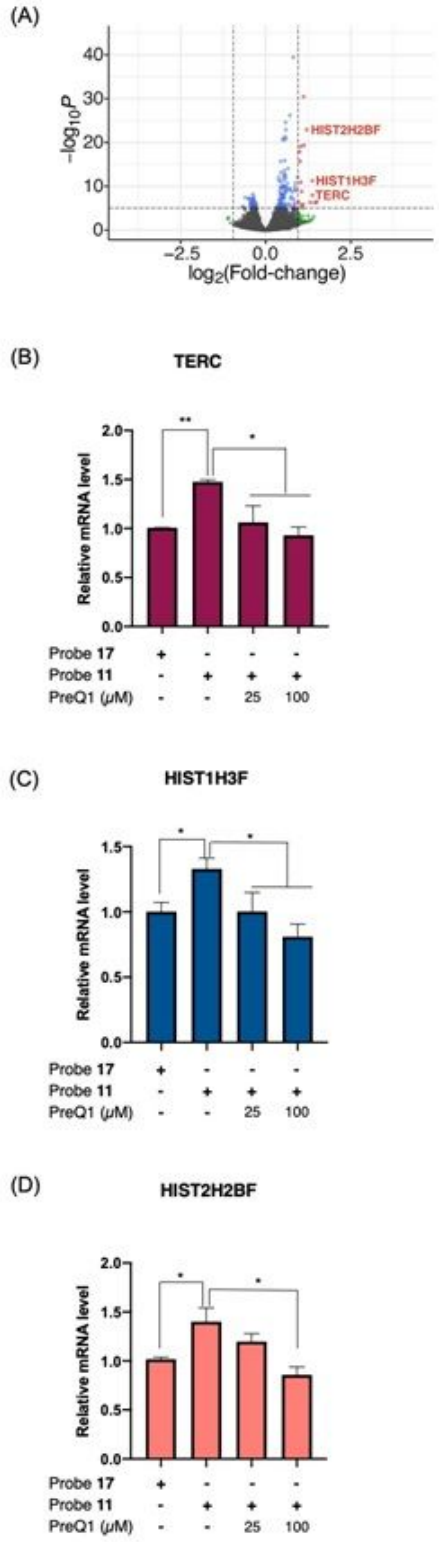


Figure 6

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. The results are presented as the mean ±

SEM (n=3) of three independent experiments. The statistical significance was calculated by t-test analysis. (**p < 0.01 and *p < 0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupportingInfoFINAL.pdf
- Scheme.pdf