The synthesis and application of a diazirine-modified uridine analogue for investigating RNA–protein interactions†

Christine C. Smith, Marcel Hollenstein* and Christian J. Leumann*

The roles played by many ncRNAs remain largely unknown. Similarly, relatively little is known about the RNA binding proteins involved in processing ncRNA. Identification of new RNA/RNA binding protein (RBP) interactions may pave the way to gain a better understanding of the complex events occurring within cells during gene expression and ncRNA biogenesis. The development of chemical tools for the isolation of RBPs is of paramount importance. In this context, we report on the synthesis of the uridine phosphoramidite UDz that bears a diazirine moiety on the nucleobase. RNA probes containing UDz units were irradiated in the presence of single-stranded DNA binding protein (SSB), which is also known to bind ssRNAs, and shown to efficiently (15% yield) and selectively cross-link to the protein. The corresponding diazirine-modified uridine triphosphate UDzTP was synthesized and its capacity to act as a substrate for the T7 RNA polymerase was tested in transcription assays. UDzTP was accepted with a maximum yield of 38% for a 26mer RNA containing a single incorporation and 28% yield for triple consecutive incorporations. Thus, this uridine analogue represents a convenient biochemical tool for the identification of RNA binding proteins and unraveling the role and function played by ncRNAs.

RNA pathways are extensively regulated by proteins called RNA binding proteins (RBPs)27–39 and vice versa,28 and together RNA and RBPs function as ribonucleoprotein (RNP) complexes, the best known of which are the spliceosome31 and ribosome.5 The inherent association of RNA with proteins has led to the development of experimental approaches for identifying ncRNA and RBP function based on the interactions between the two groups of biomacromolecules. Computational analysis,25–24 X-ray crystallography25–27 and cross-linking methods of isolation have all been employed to investigate RNA/RBP interactions. One advantage of cross-linking is that no prior knowledge of an interaction between a protein and RNA is required. It may also be used to identify networks of RNA and proteins involved in forming RNP complexes. Cross-linking between oligonucleotides and proteins may be achieved by the addition of external cross-linking reagents,35–37 or by inducing the cross-link photochemically with UV light.38,39

The advantage of photochemical cross-linking is that the natural binding is preserved, which may be disrupted by the addition of external reagents. Unfortunately, the wavelengths of UV light required to induce cross-linking between nucleic acids and proteins (∼254 nm) are actually damaging to the molecules and thus have limited applicability in vivo.40 The introduction of photoreactive groups into nucleic acids, allows for the modulation of the wavelength of UV light needed to achieve reasonable yields of cross-linking and minimize the potential damage caused to biomolecules.33 Such modifications have included halides,38 inclusion of sulfur atoms,39,42 and the addition of

Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland. E-mail: hollenstein@dcb.unibe.ch; leumann@dcb.unibe.ch; Fax: +41 31 631 4322; Tel: +41 31 631 4355

† Electronic supplementary information (ESI) available: Experimental details of the synthesis of UDz and UDzTP, cross-linking experiments, transcription assays, and additional gels. See DOI: 10.1039/c4ra08682a
more sophisticated moieties such as azides, and benzo-phenones. In this context, diazirines have a particular appeal over their simpler counterparts, as they form carbenes upon irradiation with long wavelength UV light, that then insert into adjacent covalent bonds. Particularly, π-stabilized carbenes, such as generated by exposure of 3-trifluoromethyl-3-phenyldiazirine to 365 nm irradiation, have been used in a wide range of applications. Examples include: cross-linking of DNA–protein interactions, photoactive amino acids and peptides for substrate identification, and diazirine modified sugars to capture glycoprotein binding events. Further to this, a recent study of commonly used cross-linkers has identified the trifluoromethylldiazirine moiety superior to benzophenone and phenylazide derivatives, as it produces simple and accurate cross-linking profiles of amyloid nanostructures.

The use of diazirines to identify DNA binding protein interactions hints at the possible application of this chemistry to identify RNA binding proteins. Although 3-trifluoromethyl-3-phenyldiazirine has been used to cross-link RNA previously, it involved coupling the photoactive moiety to a 2-thioctydine residue already incorporated into a tRNA anticond loop before cross-linking it to the ribosome. This strategy has shortcomings as it can lead to poor efficiencies of cross-linking, and the external addition requires accessibility to the modified base before cross-linking can be achieved, thus limiting its application. Moreover, modified RNA triphosphates (NTPs) have been previously synthesized as photoactive substrates for polymerases; although the modifications have generally been made on the DNA level, rather than RNA. Photo-labelling of viral proteins during transcription with 5-azidouridine triphosphate has been described, but the azido functionality is incompatible with phosphoramidite chemistry, limiting its applicability. In addition, NTPs have efficiently served in in vitro selection experiments for the generation of modified aptamers and ribozymes, and this methodology could be exploited to detect previously unknown interactions between RNA and proteins and polymerases.

Here, we report on the synthesis of the RNA building block bearing a diazirine moiety on the nucleobase (Fig. 1) with the intent of developing a new chemical tool for the investigation of the interaction of RNAs with RBPs.

2. Results and discussion

2.1 Design of the probe

We postulated that by incorporating the diazirine moiety within an RNA probe, through attachment at the C5-position of the uridine nucleobase, the photoactive RNA probe would be compatible with a wider range of applications. The choice of the location of the diazirine photoreactive group was dictated by the fact that appending modifications at the C5-position of pyrimidines minimizes structural perturbation of duplexes. In addition, the photoreactive group was connected to the nucleobase via a pentynol linker arm that was designed so as to minimize undesired self-binding to the RNA probe. Therefore, a diazirine-containing uridine analogue (UDz) was designed to be incorporated into RNA probes using standard phosphoramidite chemistry (Fig. 1). The DNA analogue dUDz has been previously synthesized by Carell et al. and incorporated into DNA strands opposite a UV induced pyrimidine lesion and was able to selectively cross-link the DNA lesion repair enzyme DNA polytase in the presence of other proteins.

2.2 Synthesis of the phosphoramidite UDz and nucleoside triphosphate UDzTP

The synthesis of phosphoramidite UDz followed the synthetic pathway developed for the corresponding DNA building block. Briefly, diazirine 1 (ref. 48, 53 and 74) was coupled to 4-pentyn-1-ol in the presence of NaH, before Sonogashira coupling to the 2-thiocytidine residue already incorporated into a tRNA anticond loop before cross-linking it to the ribosome. This strategy has shortcomings as it can lead to poor efficiencies of cross-linking, and the external addition requires accessibility to the modified base before cross-linking can be achieved, thus limiting its application. Moreover, modified RNA triphosphates (NTPs) have been previously synthesized as photoactive substrates for polymerases; although the modifications have generally been made on the DNA level, rather than RNA. Photo-labelling of viral proteins during transcription with 5-azidouridine triphosphate has been described, but the azido functionality is incompatible with phosphoramidite chemistry, limiting its applicability. In addition, NTPs have efficiently served in in vitro selection experiments for the generation of modified aptamers and ribozymes, and this methodology could be exploited to detect previously unknown interactions between RNA and proteins and polymerases.

![Chemical structures of the photoactive uridine analogues UDz and UDzTP.](image)

Scheme 1 Synthesis of intermediate 4. Reagents and conditions were as follows: (i) 4-pentyn-1-ol, NaH, THF, 0 °C, 2 h, 68%; (ii) Pd(PPh3)2Cl2, Et3N, Cul, DMF, r.t., o.n., 84%.

Fig. 1 Chemical structures of the photoactive uridine analogues UDz and UDzTP.
The suitably protected intermediate 7 was then converted to the corresponding triphosphate $\text{UD}_2\text{TP}$ by application of the Ludwig-Eckstein protocol$^{75,76}$ and was obtained in 21% after a thorough semi-preparative RP-HPLC purification.$^{76}$

**Scheme 3** Synthesis of triphosphate $\text{UD}_2\text{TP}$. Reagents and conditions were as follows: (i) Ac$_2$O, pyridine, DMAP, 0 °C to r.t., 2 h, 71%; (ii) 1% DCAA, CH$_2$Cl$_2$, 0 °C, 1.5 h, 48%; (iii) (1) 2-chloro-1,3,2-benzodioxaphosphorin-4-one, pyridine, dioxane; (2) (nBu$_3$NH)$_2$H$_2$P$_2$O$_7$, nBu$_3$N, DMF; (3) I$_2$, pyridine, H$_2$O; (4) NH$_3$ (aq.), H$_2$O, 21% (over 4 steps).

2.3 Oligonucleotide synthesis and thermal stability

The 5'-biotinylated 21mer RNA probes were designed for capture on streptavidin magnetic beads after cross-linking and contained single $\text{UD}_2$ units at different locations within the sequence (ONs 2–4 in Table 1 and ESI†).

Moreover, the sequence composition was chosen so as to exclude the formation of dimers and tertiary structures. UV thermal melting measurements for each sequence versus both the RNA and DNA complements were recorded to assess whether the modification caused any disturbance in the duplex stability (Fig. 2). The impact of the modification on the duplex stability with both the complementary DNA and RNA strands seems to be negligible ($\Delta T_m$ in the range of $-1.4$ to $+0.5$ °C per modification; Table 2). A similar behavior has been observed for other C5-modified uridine analogues,$^{77-79}$ and therefore modifications located at the C5 position of pyrimidines are well-accommodated in the major groove and do not interfere with duplex formation and stability.$^{80,81}$

**Fig. 2** UV thermal melting experiments for ON1-4 versus RNA (A) and DNA (B).

<table>
<thead>
<tr>
<th>ON</th>
<th>Biotin-CCUAUGUCACUGAAACUAUCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ON2</td>
<td>Biotin-CCUAUGUCACUGAAACXAUCC</td>
</tr>
<tr>
<td>ON3</td>
<td>Biotin-CCUAUGUCACXGAAACUAUCC</td>
</tr>
<tr>
<td>ON4</td>
<td>Biotin-CCXAUGUCACUGAAACUAUCC</td>
</tr>
</tbody>
</table>

2.4 Cross-linking experiments

Initial irradiation experiments with the modified oligonucleotides were performed in the presence of the single stranded DNA binding protein (SSB).$^{82}$ Although SSB is a DNA binding protein with a 10-fold lower affinity for ssRNA compared to ssDNA,$^{83,84}$ it was deemed sufficient for a “proof of concept” experiment. ON1-4 were 3'-end $^{32}$P-radiolabelled with T4 RNA ligase for visualization on the gel and to avoid photobleaching of potential fluorescent tags.$^{85}$
The probes and negative controls were irradiated (λ = 375 nm for 30 minutes, in pH 7.8 buffer) in the presence of a slight excess of the SSB protein, and eluted on a 10% tricine SDS PAGE, followed by phosphorimaging (Fig. 3). The irradiation conditions were chosen so that the diazirine will photolyze into reactive carbones, which in turn can insert irreversibly into nearby CH bonds of the target protein. The appearance of a slower running radioactive band (highlighted in red) in Lane 4 of each gel, is consistent with cross-linking of the modified RNA after irradiation. Indeed, the appearance of this band was not observed in the natural control (Lane 3) or in the absence of irradiation (Lane 2). Moreover, the appearance of a faint, faster running band in Lane 2 (highlighted in green), indicates that a higher molecular weight species than the free RNA is formed under irradiation conditions, in the absence of the protein.

We hypothesize that interstrand cross-linking might occur by insertion of the carbene generated under irradiation conditions into CH bonds of an adjacent RNA strand, although care was taken when designing the oligonucleotide sequences to ensure no hairpin formation or dimerization was likely. The disappearance of this band in the presence of the protein underscores the high selectivity and effectiveness of the diazirine photoreactive group incorporated in the RNA sequences.

As a second confirmation of the presence of RNA cross-linked to the protein, samples of ON1 and ON4 were irradiated in the presence of SSB, before purification with a size exclusion spin column (30 kDa cut off), then 3'-end 32P-radiolabelled. The resulting samples were eluted on 10% tricine SDS PAGE, along with the ON4 sample prepared in the absence of SSB. The radioactive band corresponding to the RNA–SSB crosslinks in both samples with ON4 was taken when designing the oligonucleotide sequences to ensure no hairpin formation or dimerization was likely. The radioactive band in Lane 2 (highlighted in green), indicates that a higher molecular weight species than the free RNA is formed under irradiation conditions, in the absence of the protein.

We hypothesize that interstrand cross-linking might occur by insertion of the carbene generated under irradiation conditions into CH bonds of an adjacent RNA strand, although care was taken when designing the oligonucleotide sequences to ensure no hairpin formation or dimerization was likely. The disappearance of this band in the presence of the protein underscores the high selectivity and effectiveness of the diazirine photoreactive group incorporated in the RNA sequences.

As a second confirmation of the presence of RNA cross-linked to the protein, samples of ON1 and ON4 were irradiated in the presence of SSB, before purification with a size exclusion spin column (30 kDa cut off), then 3'-end 32P-radiolabelled. The resulting samples were eluted on 10% tricine SDS PAGE, along with the ON4 sample prepared in the absence of SSB. The radioactive band corresponding to the RNA–SSB crosslinks in both samples with ON4 was taken when designing the oligonucleotide sequences to ensure no hairpin formation or dimerization was likely. The appearance of a slower running radioactive band (highlighted in red) in Lane 4 of each gel, is consistent with cross-linking of the modified RNA after irradiation. Indeed, the appearance of this band was not observed in the natural control (Lane 3) or in the absence of irradiation (Lane 2). Moreover, the appearance of a faint, faster running band in Lane 2 (highlighted in green), indicates that a higher molecular weight species than the free RNA is formed under irradiation conditions, in the absence of the protein.

2.5 Transcription assays

Nucleoside triphosphates represent a convenient vector for the introduction of chemical diversity into nucleic acids. In the context of ribonucleoside triphosphates (N*TPs), the modifications are usually anchored at the 2'-OH of the sugar moiety and only rarely at the level of the nucleobase. Thus, UDzTP was synthesized and used in transcription assays in order to assess the compatibility of this heavily-modified uridine analogue with an RNA polymerase. During transcription RNA polymerases use a double stranded DNA template to generate single stranded RNA transcripts. E. coli T7 RNA polymerase is commonly used in molecular biology and has been used to incorporate N*TPs including 2'-fluoro- 2'-amino-NTPs; unnatural base pairs; stabilized RNA analogues; photoactive monomers such as 8-azido ATP, 5-bromo and 5-iodo UTP and other 5'-modified UTP analogues. Matsuda et al. used a 4 base-pair dsDNA template containing a minimal T7 promoter followed by the sequence of the transcripts to generate 26 mer products containing 4'-thioRNA nucleotides. This dsDNA template was the basis for the design of the system used to test the incorporation of U^DzTP into RNA oligonucleotides [Fig. 4]. Indeed, the templates generate 26mer RNA transcripts containing a single [Product P1] or triple consecutive [Product P2] incorporation of U^Dz nucleotides. Moreover, in order to assess the efficiency of

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Tm values of duplexes of ON1–4 with complementary RNA and DNA (c(duplex) = 1.0 μM in 10 mM NaH2PO4, 150 mM NaCl, pH 7.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ON1</td>
</tr>
<tr>
<td>Tm vs. RNA (°C) (∆Tm (°C) per mod.)</td>
<td>69.2 ± 1.5</td>
</tr>
<tr>
<td>Tm vs. DNA (°C) (∆Tm (°C) per mod.)</td>
<td>59.6 ± 0.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Quantitation (in percentage) of cross-linking of RNA ON2–4 with the SSB protein</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Irradiated with SSB</td>
</tr>
<tr>
<td>ON2</td>
<td>14.0 ± 0.7</td>
</tr>
<tr>
<td>ON3</td>
<td>16.2 ± 2.2</td>
</tr>
<tr>
<td>ON4</td>
<td>16.3 ± 1.6</td>
</tr>
</tbody>
</table>

Fig. 3 10% SDS PAGE gels of ON2–3 irradiated in the presence of SSB. Lane 1: ON. Lane 2: ON irradiated. Lane 3: ON + SSB, non-irradiated. Lane 4: ON + SSB, irradiated. Lane 5: ON1 + SSB, irradiated.
Table 4  Relative transcription efficiency with dsDNA templates

<table>
<thead>
<tr>
<th>Product P1</th>
<th>Product P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency compared to natural transcript (%)</td>
<td>38</td>
</tr>
</tbody>
</table>

products could be observed when \( \text{UDzTP} \) was included (data not shown). On the other hand, optimized reaction conditions, \( \text{i.e.} \) increased concentrations of \( \text{UDzTP} \) and longer reaction times, \( \text{(vide infra)} \), led to the appearance of bands corresponding to full length products when \( \text{UDzTP} \) was used as a substitute for UTP (Lane 3 in Fig. 5A and B). This confirms the ability of T7 RNA polymerase to accept this heavily modified N*TP as a substrate. Moreover, the absence in Lane 3 of similar abortive sequences as shown in the negative control experiments (Lanes 2 and 4) is in agreement with this statement and shows that the polymerization reactions proceed with high yields. Interestingly, the T7 RNA polymerase could readily incorporate three modified \( \text{UDzTP} \) in a row (Fig. 5B) without stalling, clearly showing that this N*TP is a rather good substrate for the polymerase.

However, the transcription efficiencies are rather depleted when compared to those obtained with the unmodified UTP (Table 4) hinting at a slowing of transcription due to the presence of the N*TP. This is confirmed by the fact that the reaction times had to be extended (to at least 8 and 24 hours for the dsDNA templates \( T1/T1' \) and \( T2/T2' \), respectively) during optimization to allow the synthesis of the modified RNA. The polymerase was unable to generate the full length transcript with the exact template used by Matsuda \textit{et al.} \textit{which should lead to six incorporations of \( \text{UDzTP} \) (data not shown).} There also seems to be a position dependence on the ability of the enzyme to incorporate \( \text{UDzTP} \), since using a template containing complementary A nucleotides at positions 10, 15 and 21 (template \( T3/T3' \), Table S2, ESI†) only led to abortive sequences (data not shown).

However, a longer template \( T4/T4' \) designed to generate a 52mer RNA product (consisting the sequences of templates \( T1/T1' \) and \( T3/T3' \) appended onto one another Table S2, ESI†) containing 4 modifications dispersed throughout the transcript did generate full length transcripts, in a similar yield (37%) and time scale (26 hours) as the other templates (Fig. S46, ESI†).

3. Conclusions

With the intent of developing a new tool for the investigation of RNA–protein interactions, a photoreactive diazirine-containing uridine analogue was designed so that it could be incorporated into RNA, either synthetically \textit{via} phosphoramidite chemistry \( \text{UDz} \) or by enzymatic polymerization \( \text{UDzTP} \).

The modified RNA oligonucleotides obtained by incorporation of the phosphoramidite building block during solid-phase synthesis served two purposes: (1) assessing the effect of the modification on duplex stability and (2) showing that the inclusion of the photoreactive diazirine moiety into RNAs could serve as a versatile tool for the investigation of RNA–protein

The incorporation of the modified \( \text{UDzTP} \) \textit{in lieu of} its natural counterpart, the transcription efficiencies were measured relative to that using all natural NTPs which was defined as being 100%.

When the standard conditions for T7 RNA polymerase-mediated transcription reactions were used, little full length...
interactions. Not surprisingly, UV-melting experiments revealed that the modification did not cause any disruption of the Watson–Crick base-pairing and was well-tolerated in the major groove of the duplexes. Moreover, the resulting synthetic probes were assessed for cross-linking potential to single-stranded binding protein (SSB) in an irradiation assay. The RNA probes containing the modified uridine analogue were selectively cross-linked to SSB. The selectivity of this assay derives from the main competitor of the CH bond insertion of the protein, being the insertion of H₂O, thus only molecules in direct contact with the modified RNA will be cross-linked, preventing non-specific cross-linking. Even though the SSB mainly served as a paradigm for the development of photo-crosslinking RNA probes, this approach can be extended to investigate on the binding mode of other single-stranded nucleic acid binding proteins. Indeed, the virus-encoded single-stranded DNA binding protein ICP8 was shown to bind to ssRNA with an affinity only ~5 fold lower than that of ssDNA, hinting at a possible implication of ICP8 in the regulation of viral replication. Finally, other yet unknown proteins binding to various forms of coding or non-coding RNAs could be easily targeted by application of this general method.

The ability of T7 RNA polymerase to accept the triphosphate analogue was also investigated in transcription assays, generating modified RNA containing up to four incorporations of the diazirine-modified uridine. The efficient transcription reactions with UDzTP show that bulkier modifications are also tolerated at this particular location. In this context, various investigations have shown that UTP analogues equipped with smaller modifications such as hydrophobic moieties, amines, thiols, and azides acted as good substrates for the T7 RNA polymerase. Thus, the T7 RNA polymerase is rather tolerant to a broad variety of modifications anchored at position C5 of pyrimidine nucleoside triphosphates.

However, the rate of transcription seems to be retarded by the presence of the modified UTP. A kinetic investigation on the effect of the analogue UDzTP on the catalytic efficiency (kcat/Km) is currently under way and will shed some light on this phenomenon. Finally, the substrate acceptance of the non-canonical UDzTP analogue might be further improved by using mutant RNA polymerases.

Overall, the good substrate acceptance of triphosphate UDzTP and the photo-crosslinking ability of the probes containing the uridine analogue UDz bode well for the isolation and identification of RNA binding proteins and ncRNAs as well as understanding their interactions.

4. Experimental section
4.1 Cross-linking experiments
Single-stranded DNA binding protein (SSB) (1 μg, 10 pmol) and oligonucleotide (3 pmol of radiolabelled RNA) were incubated in the irradiation buffer [50 mM Tris-HCl, 10 mM MgCl₂, 10 mM DTT, pH 7.8] in a total volume of 15 μl in a PCR vial at room temperature for at least 30 minutes prior to irradiation. The sample was irradiated at 375 nm for 30 minutes. Loading buffer (70% formamide, 50 mM EDTA, bromophenol blue and xylene cyanol FF (each 0.1%) 15 μl) was then added and heated to 95 °C for 5 minutes. The resulting RNA–protein complexes were then resolved on 10% SDS-PAGE at 10 W for 2 hours. The gel was exposed on an imaging plate at ~18 °C overnight before visualizing on a phosphorimager.

4.2 Transcription assays
General procedure: 40 pmol of ds template DNA were annealed by heating to 95 °C and then gradually cooling down to room temperature. 0.4 mM of natural NTPs, 0.6 mM UDzTP, 100 units of T7 RNA polymerase, 0.6 units of PPiase, 0.5 μl RNasin, and 10 μCi of α-32P-ATP were then added at 0 °C to the transcription optimized buffer composed of 40 mM Tris–HCl (pH 7.9), 6 mM MgCl₂, 10 mM NaCl, 2 mM spermidine, 10 mM DTT and 0.05% of the polyborate detergent Tween, and made up to a total volume of 20 μl with ultra-pure H₂O. Reactions were incubated at 37 °C for different reaction times: template T₁, 8 hours; template T₂ 24 hours; under 20 μl mineral oil. The reactions were quenched by adding of 20 μl of stop solution (70% formamide, 50 mM EDTA, 0.1% bromophenol, 0.1% xylene cyanol) and resolved on 15–20% PAGE.

5. Abbreviations

**UDz** Diazoalkane modified uridine phosphoramidite
**UDzTP** Diazirine modified uridine triphosphate;
**N**TP Modified nucleoside triphosphate
**Tris** Tris(hydroxymethyl)aminomethane
**TBDM** tert-Butyldimethylsilyl
**RP HPLC** Reversed phase high-performance liquid chromatography
**Tm** Melting temperature
**ON** Oligonucleotide
**T** Template
**P** Product
**DMF** Dimethyl formamide
**nBuN** n-Tributylamine
**DCAA** Dichloroacetic acid
**DIPEA** N,N-Diisopropylethylamine
**CEP-Cl** 2′-Cyanoethyl N,N-diisopropylchlorophosphoramidite
**r.t.** Room temperature
**o.n.** Overnight
**DMAP** 4-Dimethylaminopyridine

Acknowledgements

This work was supported by grants from the Swiss National Science Foundation (grant-no.: PZ00P2_144595 to M.H. and grant-no.: 200020-13037 to C.J.L.).

Notes and references


