

Cite this: *RSC Adv.*, 2014, 4, 48228

# 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  $U^{Dz}$  that bears a diazirine moiety on the nucleobase. RNA probes containing  $U^{Dz}$  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  $U^{Dz}TP$  was synthesized and its capacity to act as a substrate for the T7 RNA polymerase was tested in transcription assays.  $U^{Dz}TP$  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.

Received 14th August 2014  
Accepted 22nd September 2014

DOI: 10.1039/c4ra08682a

[www.rsc.org/advances](http://www.rsc.org/advances)

## 1. Introduction

In recent years it has become clear that the roles played by non-coding RNAs (ncRNAs) are diverse and complex.<sup>1–3</sup> Examples of ncRNAs include not only the classical transfer RNA (tRNA)<sup>4</sup> and ribosomal RNA (rRNA)<sup>5</sup> involved in translation, but small interfering RNA (siRNA),<sup>6</sup> micro RNA (miRNA),<sup>6,7</sup> small nucleolar RNA (snoRNA),<sup>8</sup> small nuclear RNA (snRNA)<sup>8</sup> and numerous others that do not fit into a distinct category. Each of these classes of ncRNAs fulfils specific functions in all three domains of life (*i.e.* Eukarya, Bacteria, and Archaea) during gene expression, early development,<sup>9</sup> and cell maintenance, with an ever-growing list of diseases<sup>10–13</sup> related to unusual ncRNA behavior hinting at a crucial involvement of these RNAs for correct cellular function.<sup>14</sup> Difficulties in identifying the function of these ncRNAs is often connected to complications in differentiating mRNAs from ncRNAs in bioinformatic approaches,<sup>15</sup> alongside problems generating enough material with sufficient separation to distinguish unique ncRNAs in experimental procedures.<sup>16</sup>

RNA pathways are extensively regulated by proteins called RNA binding proteins (RBPs)<sup>17–19</sup> and *vice versa*,<sup>20</sup> and together RNA and RBPs function as ribonucleoprotein (RNP) complexes, the best known of which are the spliceosome<sup>21</sup> and ribosome.<sup>5</sup> 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,<sup>22–24</sup> X-ray crystallography<sup>25–27</sup> and cross-linking<sup>28–34</sup> 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,<sup>35–37</sup> or by inducing the cross-link photochemically with UV light.<sup>38,39</sup>

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*.<sup>40</sup> 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.<sup>41</sup> Such modifications have included halides,<sup>28,42</sup> inclusion of sulfur atoms,<sup>43–45</sup> and the addition of

Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland. E-mail: [hollenstein@dcb.unibe.ch](mailto:hollenstein@dcb.unibe.ch); [leumann@dcb.unibe.ch](mailto:leumann@dcb.unibe.ch); Fax: +41 31 631 3422; Tel: +41 31 631 4355

† Electronic supplementary information (ESI) available: Experimental details of the synthesis of  $U^{Dz}$  and  $U^{Dz}TP$ , cross-linking experiments, transcription assays, and additional gels. See DOI: 10.1039/c4ra08682a

more sophisticated moieties such as azides,<sup>46,47</sup> and benzophenones.<sup>48</sup> In this context, diazirines<sup>49–52</sup> have a particular appeal over their simpler counterparts, as they form carbenes upon irradiation with long wavelength UV light,<sup>53,54</sup> that then insert into adjacent covalent bonds.<sup>55,56</sup> Particularly,  $\pi$ -stabilized carbenes,<sup>57,58</sup> such as generated by exposure of 3-trifluoromethyl-3-phenyldiazirine<sup>54,59</sup> to 365 nm irradiation, have been used in a wide range of applications.<sup>56,60</sup> Examples include: cross-linking of DNA–protein interactions;<sup>49–52,61,62</sup> photoactive amino acids and peptides for substrate identification;<sup>63–65</sup> and diazirine modified sugars to capture glycoprotein binding events.<sup>66</sup> Further to this, a recent study of commonly used cross-linkers have identified the trifluoromethyldiazirine moiety superior to benzophenone and phenylazide derivatives, as it produces simple and accurate cross-linking profiles of amyloid nanostructures.<sup>67</sup>

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,<sup>68</sup> it involved coupling the photoactive moiety to a 2-thiocytidine residue already incorporated into a tRNA anticodon 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 (N\*TPs) have been previously synthesized as photoactive substrates for polymerases; although the modifications have generally been made on the DNA level, rather than RNA.<sup>69</sup> Photo-labelling of viral proteins during transcription with 5-azidouridine triphosphate<sup>70</sup> has been described, but the azido functionality is incompatible with phosphoramidite chemistry, limiting its applicability. In addition, N\*TPs have efficiently served in *in vitro* selection experiments for the generation of modified aptamers and ribozymes,<sup>71</sup> 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 **U<sup>Dz</sup>** 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.<sup>72</sup> 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 (**U<sup>Dz</sup>**) was designed to be incorporated into RNA probes using standard phosphoramidite chemistry (Fig. 1). The DNA analogue **dU<sup>Dz</sup>** 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 photolyase in the presence of other proteins.<sup>48,62</sup>

### 2.2 Synthesis of the phosphoramidite **U<sup>Dz</sup>** and nucleoside triphosphate **U<sup>Dz</sup>TP**

The synthesis of phosphoramidite **U<sup>Dz</sup>** followed the synthetic pathway developed for the corresponding DNA building block.<sup>62</sup> Briefly, diazirine **1** (ref. 54, 73 and 74) was coupled to 4-pentyn-1-ol in the presence of NaH, before Sonogashira coupling to the protected nucleoside **3** (which was prepared *via* standard DMT protection of 5-iodouridine) to generate **4**, a common precursor to both **U<sup>Dz</sup>** and **U<sup>Dz</sup>TP** (Scheme 1 and ESI†). Indeed, in order to generate phosphoramidite **U<sup>Dz</sup>**, **4** was selectively protected at the 2'-OH with TBDMS before phosphitylation under standard conditions to yield **U<sup>Dz</sup>** (56% over 2 steps; Scheme 2).

Moreover, **U<sup>Dz</sup>TP** could be accessed by first bis-acetylating compound **4** (71%) and then removing the DMT protecting

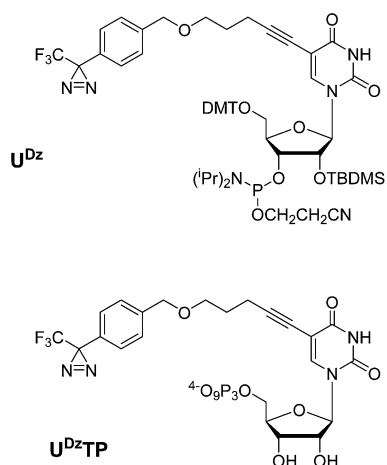
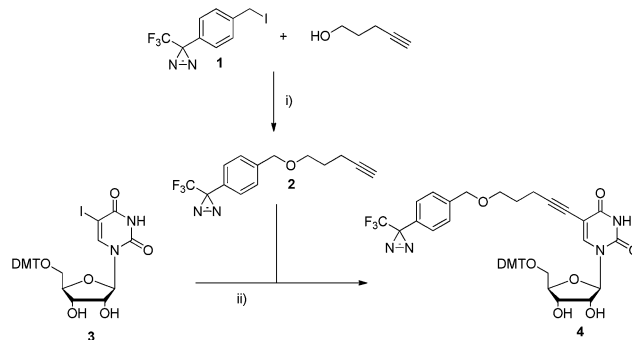
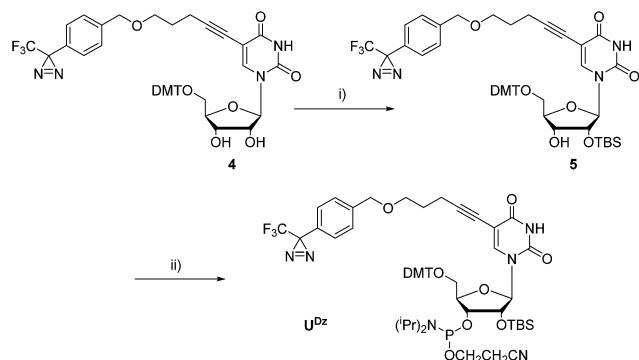


Fig. 1 Chemical structures of the photoactive uridine analogues **U<sup>Dz</sup>** and **U<sup>Dz</sup>TP**.

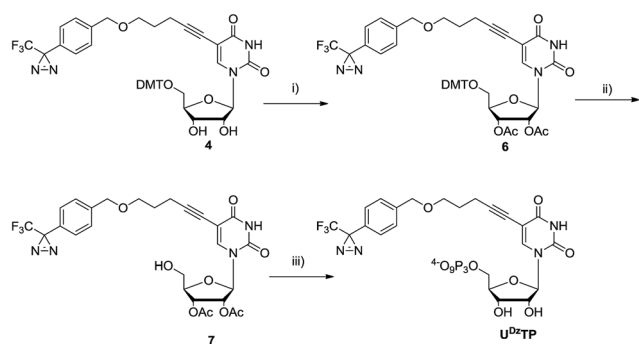


Scheme 1 Synthesis of intermediate **4**. Reagents and conditions were as follows: (i) 4-pentyn-1-ol, NaH, THF, 0 °C, 2 h, 68%; (ii) Pd(PPh<sub>3</sub>)<sub>2</sub>Cl<sub>2</sub>, Et<sub>3</sub>N, CuI, DMF, r.t., o.n., 84%.





**Scheme 2** Synthesis of phosphoramidite  $U^{Dz}$ . Reagents and conditions were as follows: (i) TBDMS-Cl, pyridine,  $AgNO_3$ , THF, r.t., o.n., 72%; (ii) CEP-Cl, DIPEA, THF, r.t., 1 h, 78%.



**Scheme 3** Synthesis of triphosphate  $U^{DzTP}$ . Reagents and conditions were as follows: (i)  $Ac_2O$ , pyridine, DMAP, 0 °C to r.t., 2 h, 71%; (ii) 1% DCAA,  $CH_2Cl_2$ , 0 °C, 1.5 h, 48%; (iii) (1) 2-chloro-1,3,2-benzodioxaphosphorin-4-one, pyridine, dioxane; (2)  $(nBu_3NH)_2H_2P_2O_7$ ,  $nBu_3N$ , DMF; (3)  $I_2$ , pyridine,  $H_2O$ ; (4)  $NH_3$  (aq.),  $H_2O$ , 21% (over 4 steps).

group under acidic conditions (Scheme 3). The suitably protected intermediate **7** was then converted to the corresponding triphosphate  $U^{DzTP}$  by application of the Ludwig-Eckstein protocol<sup>75,76</sup> and was obtained in 21% after a thorough semi-preparative RP-HPLC purification.<sup>76</sup>

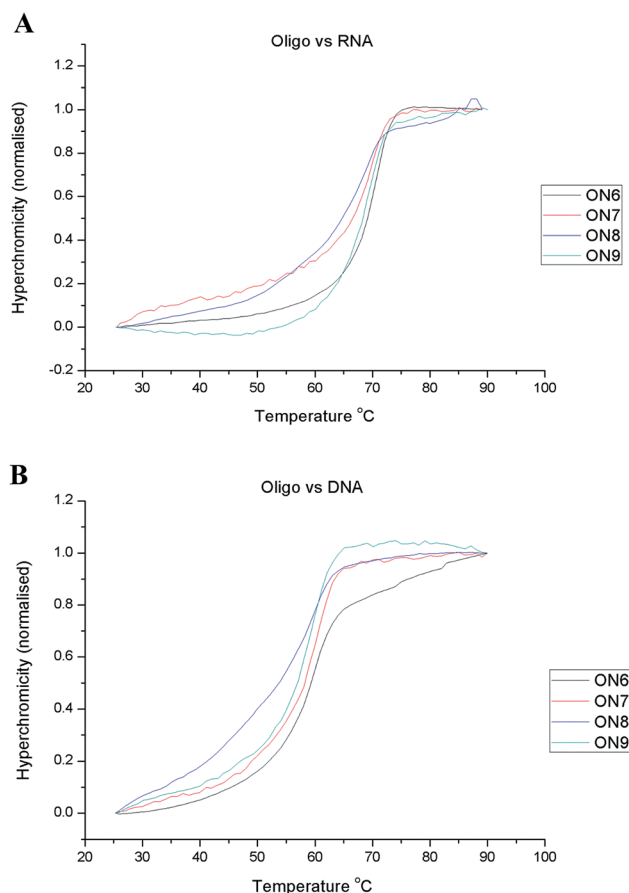
### 2.3 Oligonucleotide synthesis and thermal stability

The 5'-biotinylated 21mer RNA probes were designed for capture on streptavidin magnetic beads after cross-linking and contained single  $U^{Dz}$  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,<sup>77–79</sup> and therefore modifications located at the C5 position of pyrimidines are well-

**Table 1** Biotinylated oligonucleotide sequences used for UV thermal denaturation experiments and cross-linking experiments (X indicates the position of the  $U^{Dz}$  modification)

	Sequence (5' → 3')
ON1	Biotin-CCUAUGUCACUGAAACUAUCC
ON2	Biotin-CCUAUGUCACUGAAACXAUCC
ON3	Biotin-CCUAUGUCACXGAAACUAUCC
ON4	Biotin-CCXAUGUCACUGAAACUAUCC



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

accommodated in the major groove and do not interfere with duplex formation and stability.<sup>80,81</sup>

### 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).<sup>82</sup> Although SSB is a DNA binding protein with a 10-fold lower affinity for ssRNA compared to ssDNA,<sup>83,84</sup> 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.<sup>85</sup>



**Table 2**  $T_m$  values of duplexes of ON1–4 with complementary RNA and DNA ( $c(\text{duplex}) = 1.0 \mu\text{M}$  in 10 mM  $\text{NaH}_2\text{PO}_4$ , 150 mM NaCl, pH 7.0)

	$T_m$ vs. RNA ( $^\circ\text{C}$ ) ( $\Delta T_m$ ( $^\circ\text{C}$ ) per mod.)	$T_m$ vs. DNA ( $^\circ\text{C}$ ) ( $\Delta T_m$ ( $^\circ\text{C}$ ) per mod.)
ON1	$69.2 \pm 1.5$	$59.6 \pm 0.6$
ON2	$69.6 \pm 1.4$ (+0.4)	$60.0 \pm 0.5$ (+0.4)
ON3	$67.8 \pm 1.5$ (−1.4)	$60.1 \pm 0.5$ (+0.5)
ON4	$69.1 \pm 0.5$ (−0.1)	$59.5 \pm 0.6$ (−0.1)

**Table 3** Quantitation (in percentage) of cross-linking of RNA ON2–4 with the SSB protein

	Percentage of cross-linked RNA (relative to free RNA)	
	Irradiated with SSB	Irradiated RNA
ON2	$14.0 \pm 0.7$	$3.4 \pm 2.5$
ON3	$16.2 \pm 2.2$	$5.1 \pm 3.0$
ON4	$16.3 \pm 1.6$	$2.4 \pm 1.8$

The probes and negative controls were irradiated ( $\lambda = 375$  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,<sup>86</sup> followed by phosphorimaging (Fig. 3). The irradiation conditions were chosen so that the diazirine will photolyze into reactive carbenes, which in turn can insert irreversibly into nearby CH bonds of the target protein.<sup>48,62</sup> 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 5) 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.

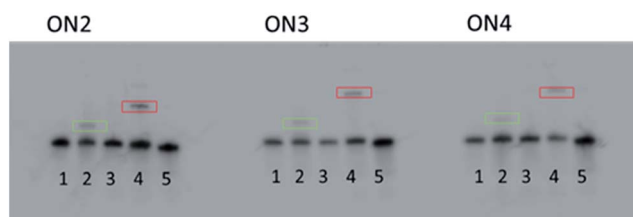
As a second confirmation of the presence of RNA cross-linked to the protein, samples of ON1 and ON4 were irradiated in the presence of SSB, before purification with a size exclusion spin column (30 kDa cut off), then 3'-end  $^{32}\text{P}$ -radiolabelled. The resulting samples were eluted on 10% tricine SDS PAGE, alongside the ON4 sample prepared in the first experiment (see ESI Fig. S45†). The radioactive band corresponding to the RNA–SSB crosslinks in both samples with ON4 eluted at the same rate, and the selective nature of radiolabelling using an RNA ligase confirms the presence of RNA in

the modified samples, with no such corresponding band in the natural control, or in the absence of irradiation.

The yields for cross-linking to SSB obtained by quantification of the respective bands can be found in Table 3. The yield of cross-linking for each modified strand of RNA was consistently around 15%, which compares favorably to the crosslinking yields obtained for DNA probes equipped with diazirine groups and DNA photolyases.<sup>62</sup> The main side reaction competing with the carbene CH bond insertion in this experiment is the insertion of  $\text{H}_2\text{O}$ . The insertion of  $\text{H}_2\text{O}$  is unavoidable in the unbound RNA, thus, only the probes in contact with the protein at the moment of excitation are cross-linked. The advantage of the carbene being quenched by small molecules is that the resulting RNA will not interfere with data obtained for cross-linking to macromolecules as they will be resolved easily under electrophoretic conditions, giving specific cross-linking profiles for a mixture of components. Isolation with streptavidin beads was never achieved. Perhaps the bulk of the protein prevented the biotin from interacting with the beads.

## 2.5 Transcription assays

Nucleoside triphosphates represent a convenient vector for the introduction of chemical diversity into nucleic acids.<sup>87–89</sup> In the context of ribonucleoside triphosphates ( $\text{N}^*\text{TPs}$ ), the modifications are usually anchored at the 2'-OH of the sugar moiety and only rarely at the level of the nucleobase.<sup>89</sup> Thus,  $\text{U}^{\text{Dz}}\text{TP}$  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<sup>90,91</sup> and has been used to incorporate  $\text{N}^*\text{TPs}$  including 2'-fluoro- 2'-amino-NTPs;<sup>92</sup> unnatural base pairs;<sup>93,94</sup> stabilized RNA analogues;<sup>95</sup> photoactive monomers such as 8-azido ATP,<sup>96</sup> 5-bromo and 5-iodo UTP;<sup>28,97</sup> and other 5'-modified UTP analogues.<sup>98,99</sup> Matsuda *et al.* used a 43 base-pair dsDNA template containing a minimal T7 promoter followed by the sequence of the transcripts to generate 26 mer products containing 4'-thiorRNA nucleotides.<sup>95</sup> This dsDNA template was the basis for the design of the system used to test the incorporation of  $\text{U}^{\text{Dz}}\text{TP}$  into RNA oligonucleotides (Fig. 4). Indeed, the templates generate 26mer RNA transcripts containing a single (Product **P1**) or triple consecutive (Product **P2**) incorporation of  $\text{U}^{\text{Dz}}$  nucleotides. Moreover, in order to assess the efficiency of



**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.





**T1:** 5'-TAA TAC GAC TCA CTA TAG GGA GAC CAT GCA CAA CCG GAA ACC C-3'  
**T1':** 3'-ATT ATG CTG AGT GAT ATC CCT CTG GTA CGT GTT GGC CTT TGG G-5'

T7 RNA polymerase  
 0.4 mM natural NTPs  
 0.4 mM U<sup>Dz</sup>TP  
 [ $\alpha$ -<sup>32</sup>P]ATP

Product **P1** 5'-pppGGG AGA CCA U<sup>Dz</sup>GC ACA ACC GGA AAC CC-3'

**T2:** 5'-TAA TAC GAC TCA CTA TAG GGA GAC CAA GCA CAA CCG GTT TCC C-3'  
**T2':** 3'-ATT ATG CTG AGT GAT ATC CCT CTG GTT CGT GTT GGC CAA AGG G-5'

T7 RNA polymerase  
 0.4 mM natural NTPs  
 0.4 mM U<sup>Dz</sup>TP  
 [ $\alpha$ -<sup>32</sup>P]ATP

Product **P2** 5'-pppGGG AGA CCA AGC ACA ACC GGU<sup>Dz</sup>U<sup>Dz</sup>U<sup>Dz</sup>C CC-3'

**Fig. 4** Incorporation of U<sup>Dz</sup>TP into RNA *via* transcription assays using T7 RNA polymerase and dsDNA templates **T1/T1'** and **T2/T2'**. Italicized letters represent the T7 promoter sequence.

the incorporation of the modified U<sup>Dz</sup>TP *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

**Table 4** Relative transcription efficiency with dsDNA templates

	Product <b>P1</b>	Product <b>P2</b>
Efficiency compared to natural transcript (%)	38	28

products could be observed when U<sup>Dz</sup>TP was included (data not shown). On the other hand, optimized reaction conditions, *i.e.* increased concentrations of U<sup>Dz</sup>TP and longer reaction times, (*vide infra*), led to the appearance of bands corresponding to full length products when U<sup>Dz</sup>TP 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 U<sup>Dz</sup>TP in a row (Fig. 5B) without stalling, clearly showing that this N\*TP is a rather good substrate for the polymerase.

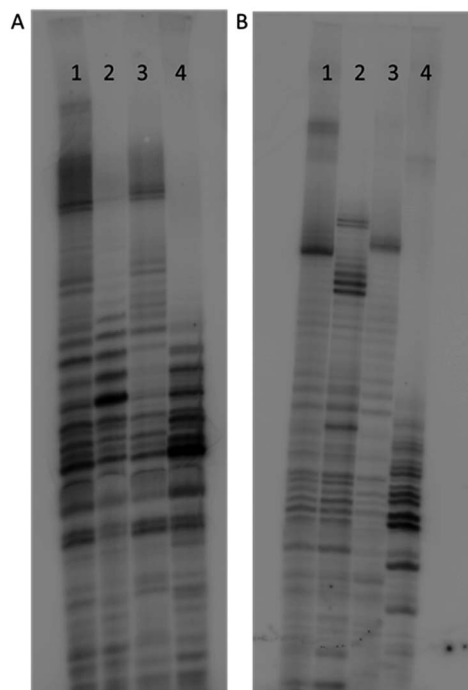
However, the transcription efficiencies are rather depleted when compared to those obtained with the unmodified UTP (Table 4) hinting at a slowing of transcription due to the presence of the N\*TP. This is confirmed by the fact that the reaction times had to be extended (to at least 8 and 24 hours for 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 *et al.* which should lead to six incorporations of U<sup>Dz</sup>TP (data not shown). There also seems to be a position dependence on the ability of the enzyme to incorporate U<sup>Dz</sup>TP, 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 *via* phosphoramidite chemistry (U<sup>Dz</sup>) or by enzymatic polymerization (U<sup>Dz</sup>TP).

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



**Fig. 5** Gel images of 20% PAGE of transcription assays using (A) dsDNA template **T1/T1'** or (B) template **T2/T2'**. Lane 1: Natural NTPs. Lane 2: Negative control. Lane 3: U<sup>Dz</sup>TP. Lane 4: ATP and GTP only. General procedure: 40 pmol of template DNA, 0.4 mM of natural NTPs, 0.6 mM U<sup>Dz</sup>TP, 100 units of T7 RNA polymerase, 0.6 units of PPIase, 0.5  $\mu$ l RNasin, and 10  $\mu$ Ci of  $\alpha$ -<sup>32</sup>P-ATP in the transcription optimized buffer. Reactions were incubated at 37 °C: template 1, for at least 8 hours; template 2 for at least 24 hours; under 20  $\mu$ l mineral oil.



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 H<sub>2</sub>O, 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.<sup>84</sup> 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 U<sup>Dz</sup>TP 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,<sup>98</sup> amines, thiols,<sup>79,100</sup> and azides<sup>101</sup> 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 U<sup>Dz</sup>TP on the catalytic efficiency ( $k_{cat}/K_M$ ) is currently under way and will shed some light on this phenomenon. Finally, the substrate acceptance of the non-canonical U<sup>Dz</sup>TP analogue might be further improved by using mutant RNA polymerases.<sup>102</sup>

Overall, the good substrate acceptance of triphosphate U<sup>Dz</sup>TP and the photo-crosslinking ability of the probes containing the uridine analogue U<sup>Dz</sup> 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 MgCl<sub>2</sub>, 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 U<sup>Dz</sup>TP, 100 units of T7 RNA polymerase, 0.6 units of PPIase, 0.5 µl RNasin, and 10 µCi of α-<sup>32</sup>P-ATP were then added at 0 °C to the transcription optimized buffer composed of 40 mM Tris-HCl (pH 7.9), 6 mM MgCl<sub>2</sub>, 10 mM NaCl, 2 mM spermidine, 10 mM DTT and 0.05% of the polysorbate detergent Tween, and made up to a total volume of 20 µl with ultra-pure H<sub>2</sub>O. 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

U <sup>Dz</sup>	Diazirine modified uridine phosphoramidite
U <sup>Dz</sup> TP	Diazirine modified uridine triphosphate;
N*TP	Modified nucleoside triphosphate
Tris	Tris(hydroxymethyl)aminomethane
TBDMS	<i>tert</i> -Butyldimethylsilyl
RP-HPLC	Reversed phase high-performance liquid chromatography
<i>T</i> <sub>m</sub>	Melting temperature
ON	Oligonucleotide
T	Template
P	Product
DMF	Dimethyl formamide
<i>n</i> Bu <sub>3</sub> N	<i>n</i> -Tributylamine
DCAA	Dichloroacetic acid
DIPEA	<i>N,N</i> -Diisopropylethylamine
CEP-Cl	2'-Cyanoethyl <i>N,N</i> -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

- M. Morlando, M. Ballarino, A. Fatica and I. Bozzoni, *ChemMedChem*, 2014, **9**, 505–510.



- 2 J. S. Mattick and I. V. Makunin, *Hum. Mol. Genet.*, 2006, **15**, R17–R29.
- 3 A. Huttenhofer, P. Schattner and N. Polacek, *Trends Genet.*, 2005, **21**, 289–297.
- 4 A. Rich and U. L. Rajbhandary, *Annu. Rev. Biochem.*, 1976, **45**, 805–860.
- 5 R. N. Nazar, *IUBMB Life*, 2004, **56**, 457–465.
- 6 R. W. Carthew and E. J. Sontheimer, *Cell*, 2009, **136**, 642–655.
- 7 D. P. Bartel, *Cell*, 2004, **116**, 281–297.
- 8 A. G. Matera, R. M. Terns and M. P. Terns, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 209–220.
- 9 U. A. Orom, T. Derrien, M. Beringer, K. Gumireddy, A. Gardini, G. Bussotti, F. Lai, M. Zytnicki, C. Notredame, Q. Huang, R. Guigo and R. Shiekhattar, *Cell*, 2010, **143**, 46–58.
- 10 M. Esteller, *Nat. Rev. Genet.*, 2011, **12**, 861–874.
- 11 M. A. Varela, T. C. Roberts and M. J. A. Wood, *Neurotherapeutics*, 2013, **10**, 621–631.
- 12 S. D. Rege, T. Geetha, S. R. Pondugula, C. A. Zizza, C. M. Wernette and J. R. Babu, *ISRN Neurol.*, 2013, **2013**, 375852.
- 13 T. A. Cooper, L. L. Wan and G. Dreyfuss, *Cell*, 2009, **136**, 777–793.
- 14 J. E. Wilusz, H. Sunwoo and D. L. Spector, *Genes Dev.*, 2009, **23**, 1494–1504.
- 15 M. E. Dinger, K. C. Pang, T. R. Mercer and J. S. Mattick, *PLoS Comput. Biol.*, 2008, **4**, e1000176.
- 16 A. Huttenhofer and J. Vogel, *Nucleic Acids Res.*, 2006, **34**, 635–646.
- 17 D. J. Hogan, D. P. Riordan, A. P. Gerber, D. Herschlag and P. O. Brown, *PLoS Biol.*, 2008, **6**, 2297–2313.
- 18 T. Glisovic, J. L. Bachorik, J. Yong and G. Dreyfuss, *FEBS Lett.*, 2008, **582**, 1977–1986.
- 19 A. E. Pasquinelli, *Nat. Rev. Genet.*, 2012, **13**, 271–282.
- 20 X. Wang, S. Arai, X. Song, D. Reichart, K. Du, G. Pascual, P. Tempst, M. G. Rosenfeld, C. K. Glass and R. Kurokawa, *Nature*, 2008, **454**, 126–130.
- 21 Z. L. Zhou, L. J. Licklider, S. P. Gygi and R. Reed, *Nature*, 2002, **419**, 182–185.
- 22 S. Jones, D. T. A. Daley, N. M. Luscombe, H. M. Berman and J. M. Thornton, *Nucleic Acids Res.*, 2001, **29**, 943–954.
- 23 M. Terribilini, J.-H. Lee, C. Yan, R. L. Jernigan, V. Honavar and D. Dobbs, *RNA*, 2006, **12**, 1450–1462.
- 24 H. Kim, E. Jeong, S. W. Lee and K. Han, *FEBS Lett.*, 2003, **552**, 231–239.
- 25 T. Ritschel, C. Atmanene, K. Reuter, A. Van Dorselaer, S. Sanglier-Cianferani and G. Klebe, *J. Mol. Biol.*, 2009, **393**, 833–847.
- 26 J. A. Doudna, *Curr. Biol.*, 1999, **9**, R731–R734.
- 27 C. Oubridge, N. Ito, P. R. Evans, C. H. Teo and K. Nagai, *Nature*, 1994, **372**, 432–438.
- 28 M. C. Willis, B. J. Hicke, O. C. Uhlenbeck, T. R. Cech and T. H. Koch, *Science*, 1993, **262**, 1255–1257.
- 29 J. Ule, K. B. Jensen, M. Ruggiu, A. Mele, A. Ule and R. B. Darnell, *Science*, 2003, **302**, 1212–1215.
- 30 D. D. Licatalosi, A. Mele, J. J. Fak, J. Ule, M. Kayikci, S. W. Chi, T. A. Clark, A. C. Schweitzer, J. E. Blume, X. Wang, J. C. Darnell and R. B. Darnell, *Nature*, 2008, **456**, 464–469.
- 31 J. L. Corden, *Chem. Biol.*, 2010, **17**, 316–318.
- 32 M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr, A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan and T. Tuschl, *Cell*, 2010, **141**, 129–141.
- 33 J. Koenig, K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe and J. Ule, *Nat. Struct. Mol. Biol.*, 2010, **17**, 909–915.
- 34 M. J. Luo and R. Reed, *et al.*, *Current protocols in molecular biology*, ed. Frederick M. Ausubel, 2003, ch. 27, Unit 27 22.
- 35 I. Wower, J. Wower, M. Meinke and R. Brimacombe, *Nucleic Acids Res.*, 1981, **9**, 4285–4302.
- 36 M. Osswald, B. Greuer and R. Brimacombe, *Nucleic Acids Res.*, 1990, **18**, 6755–6760.
- 37 L. L. G. Carrette, E. Gyssels, J. Loncke and A. Maddier, *Org. Biomol. Chem.*, 2014, **12**, 931–935.
- 38 R. Pelle and N. B. Murphy, *Nucleic Acids Res.*, 1993, **21**, 2453–2458.
- 39 B. M. Merrill, K. R. Williams, J. W. Chase and W. H. Konigsberg, *J. Biol. Chem.*, 1984, **259**, 10850–10856.
- 40 G. J. Kantor and D. R. Hull, *Biophys. J.*, 1979, **27**, 359–370.
- 41 M. M. Konarska, *Methods*, 1999, **18**, 22–28.
- 42 M. C. Willis, K. A. Lecuyer, K. M. Meisenheimer, O. C. Uhlenbeck and T. H. Koch, *Nucleic Acids Res.*, 1994, **22**, 4947–4952.
- 43 A. Favre, G. Moreno, M. O. Blondel, J. Kliber, F. Vinzens and C. Salet, *Biochem. Biophys. Res. Commun.*, 1986, **141**, 847–854.
- 44 E. J. Sontheimer, *Mol. Biol. Rep.*, 1994, **20**, 35–44.
- 45 D. Temiakov, P. E. Montesana, K. Ma, A. Mustaev, S. Borukhov and W. T. McAllister, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 14109–14114.
- 46 R. K. Evans, J. D. Johnson and B. E. Haley, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 5382–5386.
- 47 N. Demeshkina, M. Repkova, A. Ven'yaminova, D. Graifer and G. Karpova, *RNA*, 2000, **6**, 1727–1736.
- 48 M. Winnacker, V. Welzmler, R. Strasser and T. Carell, *ChemBioChem*, 2010, **11**, 1345–1349.
- 49 T. Yamaguchi and M. Saneyoshi, *Nucleosides Nucleotides*, 1996, **15**, 607–618.
- 50 T. Yamaguchi, K. Suyama, K. Narita, S. Kohgo, A. Tomikawa and M. Saneyoshi, *Nucleic Acids Res.*, 1997, **25**, 2352–2358.
- 51 M. Liebmann, F. Di Pasquale and A. Marx, *ChemBioChem*, 2006, **7**, 1965–1969.
- 52 U. K. Shigdel, J. Zhang and C. He, *Angew. Chem., Int. Ed. Engl.*, 2008, **47**, 90–93.
- 53 G. F. Bradley, W. B. L. Evans and I. D. R. Stevens, *J. Chem. Soc., Perkin Trans. 2*, 1977, 1214–1220.
- 54 J. Brunner, H. Senn and F. M. Richards, *J. Biol. Chem.*, 1980, **255**, 3313–3318.
- 55 R. A. Moss, *Acc. Chem. Res.*, 2006, **39**, 267–272.
- 56 J. Das, *Chem. Rev.*, 2011, **111**, 4405–4417.



- 57 P. H. Mueller, N. G. Rondan, K. N. Houk, J. F. Harrison, D. Hooper, B. H. Willen and J. F. Liebman, *J. Am. Chem. Soc.*, 1981, **103**, 5049–5052.
- 58 D. L. S. Brahms and W. P. Dailey, *Chem. Rev.*, 1996, **96**, 1585–1632.
- 59 M. Nassal, *Liebigs Ann. Chem.*, 1983, **1983**, 1510–1523.
- 60 L. Dubinsky, B. P. Krom and M. M. Meijler, *Bioorg. Med. Chem.*, 2012, **20**, 554–570.
- 61 T. Yamaguchi and M. Saneyoshi, *Nucleic Acids Res.*, 1996, **24**, 3364–3369.
- 62 M. Winnacker, S. Breeger, R. Strasser and T. Carell, *ChemBioChem*, 2009, **10**, 109–118.
- 63 Y. Ambroise, F. Pillon, C. Mioskowski, A. Valleix and B. Rousseau, *Eur. J. Org. Chem.*, 2001, 3961–3964.
- 64 H. Nakashima, M. Hashimoto, Y. Sadakane, T. Tomohiro and Y. Hatanaka, *J. Am. Chem. Soc.*, 2006, **128**, 15092–15093.
- 65 D. P. Smith, J. Anderson, J. Plante, A. E. Ashcroft, S. E. Radford, A. J. Wilson and M. J. Parker, *Chem. Commun.*, 2008, 5728–5730.
- 66 M. R. Bond, H. Zhang, P. D. Vu and J. J. Kohler, *Nat. Protoc.*, 2009, **4**, 1044–1063.
- 67 G. W. Preston, S. E. Radford, A. E. Ashcroft and A. J. Wilson, *ACS Chem. Biol.*, 2014, **9**, 761–768.
- 68 T. Doring, P. Mitchell, M. Osswald, D. Bochkariov and R. Brimacombe, *EMBO J.*, 1994, **13**, 2677–2685.
- 69 S. Verma and F. Eckstein, *Annu. Rev. Biochem.*, 1998, **67**, 99–134.
- 70 D. C. Hammond, R. K. Evans and J. A. Lesnaw, *J. Gen. Virol.*, 1992, **73**, 61–66.
- 71 M. Kuwahara and N. Sugimoto, *Molecules*, 2010, **15**, 5423–5444.
- 72 E. T. Kool, *Curr. Opin. Chem. Biol.*, 2000, **4**, 602–608.
- 73 L. B. Shih and H. Bayley, *Anal. Biochem.*, 1985, **144**, 132–141.
- 74 J. H. Robson and J. Reinhart, *J. Am. Chem. Soc.*, 1955, **77**, 498–499.
- 75 J. Ludwig and F. Eckstein, *J. Org. Chem.*, 1989, **54**, 631–635.
- 76 M. Hollenstein, C. C. Smith and M. Ræz, *J. Visualized Exp.*, 2014, e51385.
- 77 M. Terrazas and E. T. Kool, *Nucleic Acids Res.*, 2009, **37**, 346–353.
- 78 M. Terrazas and R. Eritja, *Mol. Diversity*, 2011, **15**, 677–686.
- 79 T. Schoetzau, J. Langner, E. Moyroud, I. Roehl, S. Vonhoff and S. Klusmann, *Bioconjugate Chem.*, 2003, **14**, 919–926.
- 80 P. Čapek, H. Cahová, R. Pohl, M. Hocek, C. Gloeckner and A. Marx, *Chem.–Eur. J.*, 2007, **13**, 6196–6203.
- 81 M. Hollenstein, *Chem.–Eur. J.*, 2012, **18**, 13320–13330.
- 82 A. Sancar, K. R. Williams, J. W. Chase and W. D. Rupp, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 4274–4278.
- 83 J. W. Chase and K. R. Williams, *Annu. Rev. Biochem.*, 1986, **55**, 103–136.
- 84 P. E. Boehmer, *Nucleic Acids Res.*, 2004, **32**, 4576–4584.
- 85 C. I. Richards, S. Choi, J.-C. Hsiang, Y. Antoku, T. Vosch, A. Bongiorno, Y.-L. Tzeng and R. M. Dickson, *J. Am. Chem. Soc.*, 2008, **130**, 5038–5039.
- 86 H. Schaeffer, *Nat. Protoc.*, 2006, **1**, 16–22.
- 87 M. Hocek and M. Fojta, *Org. Biomol. Chem.*, 2008, **6**, 2233–2241.
- 88 M. Hollenstein, *Molecules*, 2012, **17**, 13569–13591.
- 89 L. H. Lauridsen, J. A. Rothnagel and R. N. Veedu, *ChemBioChem*, 2012, **13**, 19–25.
- 90 J. F. Milligan, D. R. Groebe, G. W. Witherell and O. C. Uhlenbeck, *Nucleic Acids Res.*, 1987, **15**, 8783–8798.
- 91 K. E. McGinness and G. F. Joyce, *J. Biol. Chem.*, 2001, **277**, 2987–2991.
- 92 H. Aurup, D. M. Williams and F. Eckstein, *Biochemistry*, 1992, **31**, 9636–9641.
- 93 M. Kimoto, M. Endo, T. Mitsui, T. Okuni, I. Hirao and S. Yokoyama, *Chem. Biol.*, 2004, **11**, 47–55.
- 94 R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama and I. Hirao, *J. Am. Chem. Soc.*, 2005, **127**, 17286–17295.
- 95 Y. Kato, N. Minakawa, Y. Komatsu, H. Kamiya, N. Ogawa, H. Harashima and A. Matsuda, *Nucleic Acids Res.*, 2005, **33**, 2942–2951.
- 96 S. Gopalakrishna, V. Gusti, S. Nair, S. Sahar and R. K. Gaur, *RNA*, 2004, **10**, 1820–1830.
- 97 J. M. Gott, M. C. Willis, T. H. Koch and O. C. Uhlenbeck, *Biochemistry*, 1991, **30**, 6290–6295.
- 98 J. D. Vaught, T. Dewey and B. E. Eaton, *J. Am. Chem. Soc.*, 2004, **126**, 11231–11237.
- 99 A. A. Tanpure and S. G. Srivatsan, *ChemBioChem*, 2014, **15**, 1309–1316.
- 100 N. K. Vaish, A. W. Fraley, J. W. Szostak and L. W. McLaughlin, *Nucleic Acids Res.*, 2000, **28**, 3316–3322.
- 101 H. Rao, A. A. Tanpure, A. A. Sawant and S. G. Srivatsan, *Nat. Protoc.*, 2012, **7**, 1097–1112.
- 102 J. Attwater, S. Tagami, M. Kimoto, K. Butler, E. T. Kool, J. Wengel, P. Herdewijn, I. Hirao and P. Holliger, *Chem. Sci.*, 2013, **4**, 2804–2814.

