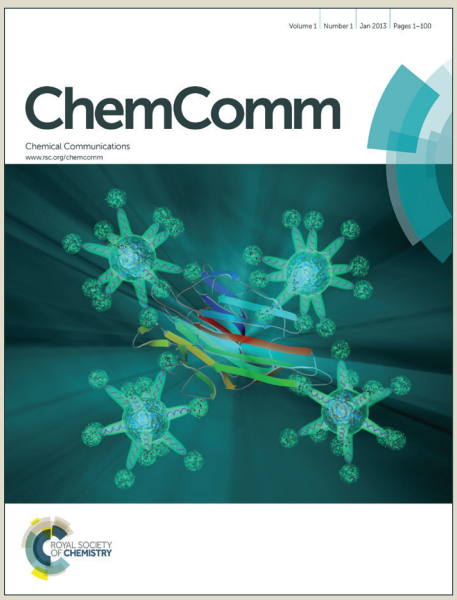
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

ChemComm

RSCPublishing

COMMUNICATION

Reverse transcription through a bulky triazole linkage in RNA: implications for RNA sequencing

Cite this: DOI: 10.1039/x0xx00000x

Xiong Chen[a], Afaf H. El-Sagheer[b,c] and Tom Brown[c*]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

A triazole linkage is formed in RNA by untemplated strain-promoted or CuAAC chemical ligation of 3′-azide and 5′-cyclooctyne oligonucleotides under denaturing conditions. Reverse transcriptase reads through these artificial linkages with omission of one nucleotide. These surprising results have implications for RNA isolation, amplification, sequencing and a variety of biological applications.

Small RNAs account for about 1% of the total RNA population in eukaryotic cells and include microRNAs (miRNAs), small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs).[1] They are crucial in regulating many biological processes including transcription and translation. Analysis of small RNAs can be carried out by high-throughput next generation sequencing (HTS) techniques.[2, 3] In order to generate RNA sequencing libraries, the isolated native RNAs are normally reverse transcribed into DNA then amplified by PCR. For this procedure the RNAs need to be ligated to adaptors of known sequence at their 3′-end and 5′-ends. Ligation of the adaptors can lead to sequencing bias[4-7] due to RNA secondary structures, which drastically reduce ligation efficiency. The resultant distorted representation of the RNA population greatly affects the reliability of studies to measure RNA expression levels.[7, 8] If T4-mediated ligation could be changed to chemical ligation under RNA-denaturing conditions the problem might be overcome. The choice of ligation chemistry is important; copper-catalysed alkyne-azide cycloaddition (CuAAC)[9] and strain-promoted copper-free[10] click ligation (SPAAC) should be considered, as these reactions are robust and can be performed under conditions in which RNA secondary structures do not exist, e.g. in salt-free water. In order to apply click ligation to RNA sequencing and related applications, a method is required to add the alkyne or the azide function to the 3′-end of the native RNA pool. Such a method has recently been developed to efficiently add a 2′-azido NTP to RNA.[11] In addition, the resultant triazole linkage has to be compatible with the reverse transcriptase (RT) enzyme used to produce complementary DNA (cDNA). This has not yet been achieved, but we have previously shown that a specific triazole mimic of the DNA phosphodiester backbone of the kind depicted in Figure 1A can be read through by DNA and RNA polymerases. The biocompatibility of this artificial linkage has been established in *E. coli* [9] and is the first example of a non-natural DNA linkage being functional in mammalian cells.[12] RNA strands can also be ligated by the CuAAC reaction to produce catalytically active ribozymes.[13] With these developments in mind we decided to use this linkage to ligate RNA adaptors to the 3′-end of RNA for applications involving reverse transcription. We now report that RT can read through the above triazole backbone with the omission of one nucleotide. Surprisingly, a much bulkier linkage produced by copper-free click ligation between a strained cyclooctyne and an azide is also read through by reverse transcriptase, as described below, offering the future possibility of an improved RNA isolation and amplification workflow for next generation sequencing (Figure 2).

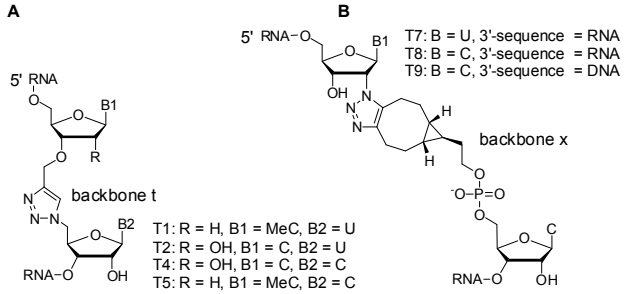


Figure 1. Structures of triazole linkages used in this study. When MeC is used in place of C this is purely for synthetic convenience

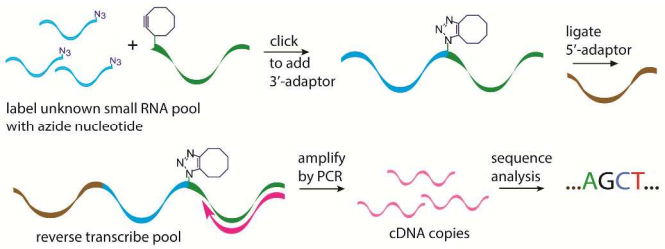


Figure 2. SPAAC ligation as a component of a proposed workflow for next generation RNA sequencing

We carried out a series of studies using a synthetic version of miR-155, a micro-RNA which plays a key role in the function of the immune system.[14] An RNA reverse transcription template with the internal MedC-triazole-U linkage MedCtU (T1 in Figure 1A) was synthesized using the CuAAC reaction. Templated click ligation of 3′-propargyl-5-MedC oligonucleotide (K1) (Table S1) and the 5′-azido-U oligonucleotide (Z1) using a complementary splint (S1) afforded T1 in good yield. An additional template (T2) was synthesised in the same way using 3′-propargyl C instead of 3′-propargyl 5-MedC (Figure S1). To study the compatibility of these backbones with reverse transcriptase (RT) enzymes, M-MLV RT and M-MLV RT (RNase H-) were used. The latter is engineered for reduced RNase H activity and is used to minimize RNA degradation.[15]

Five DNA primers of varying lengths (P1 to P5) were used with the two RNA triazole templates T1 and T2. These primers generated reverse transcription starting points at various loci; before the triazole linkage (-7, P1), (-3, P2), next to the triazole (+0, P3), one nucleotide after (+1, P4) and four nucleotides after (+4, P5). In the last two cases the triazole linkage is bridged by the bound primer. Reverse transcription of the triazole templates T1 and T2 stopped around the triazole site when primers P1, P2 and P3 were used, but gave the full length product with the bridged primers P4 and P5. Addition of Mn^{2+} to the Mg^{2+}-free buffer allowed the enzyme to read through the triazole[16] even with the non-bridged primers. This gave a product that lacked a single deoxyguanosine, as shown by polyacrylamide gel-electrophoresis (PAGE) and mass spectrometry (MS) (Figures S2 and S3). This omission is due to the enzyme not recognising the presence of the template C base adjacent to the triazole linkage. The inability of RT to read through the triazole linkage efficiently in the presence of Mg^{2+} could be partly due to instability in the duplex formed between the triazole template and the primer.[17] Therefore we changed the dinucleotide around the triazole linkage from CtU to give a more stable template/primer duplex. Two new templates T4 (CtC) and T5 (MedCtC) were synthesized by CuAAC click ligation and encouragingly M-MLV RT read through both triazole linkages in Mg^{2+} buffer with bridged and non-bridged primers (Figure 3). This indicates that duplex stabilisation allows the enzyme to copy the RNA template accurately with bridged primers, but with omission of one nucleotide after the triazole linkage with non-bridged primers (Figure S4).

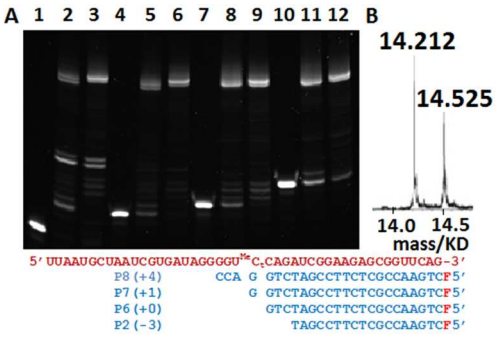


Figure 3. A) Bridged and non-bridged primers read through triazole linkage. 20% PAGE for reverse transcription of T5 (^{Me}CtC, lanes 2, 5, 8, 11) and control T6 (lanes 3, 6, 9, 12) using M-MuLV Reverse Transcriptase in Mg^{2+} buffer at 37 °C for 2 h. Lane 1, 2 and 3; primer P2(-3) and reverse transcription using it, lane 4, 5 and 6; primer P6 (+0) and reverse transcription using it; lane 7, 8 and 9; primer P7 (+1) and reverse transcription using it; lane 10, 11 and 12; primer P8 (+4) and reverse transcription using it. B) Mass spectrum of reverse transcription product using template T5 and primer P7. Calc. Mass: 14212 (M); 14525 (M + dA).

Copper-free click ligation has potential advantages over the CuAAC reaction as it will not lead to degradation of RNA in biological media. In this context the ring-strain promoted alkyne-azide [3+2] cycloaddition (SPAAC) reaction has recently been applied to DNA strand ligation,[10, 18] so we decided to investigate its use in templates for reverse transcription. We chose bicyclo[6.1.0]nonyne (BCN) as the strained alkyne[19] because the resultant triazole linker is less bulky than the alternatives, and is therefore more likely to be accommodated by the RT enzyme. Two RNA templates T7 and T8 with internal UxC and CxC were synthesised by SPAAC ligation of 5′-BCN oligonucleotide O1 with 2′-azide oligonucleotides Z2 or Z3 to produce 5′- 2′ linkages in the RNA backbone (Figure 1B). The key building blocks for the synthesis of the 3′-azide labelled RNA strand are 5′-DMT-2′-azido-2′-dU and 5′-DMT-2′-azido-2′-dC.[20] These were synthesised, coupled to succinylated aminoalkyl solid supports and used in the synthesis of 2′-azide oligonucleotides. To synthesise the cyclooctyne oligonucleotide O1, BCN phosphoramidite was added to the 5′-terminus during solid-phase synthesis. This was reacted with the 2′-azide oligonucleotides in water to generate the RNA templates T7 and T8 in excellent yield (Figure 4A). For ease of purification, and to mimic ligation of biologically-derived RNA, no template was used; instead concentrated reaction mixtures were employed to ensure good ligation yields.

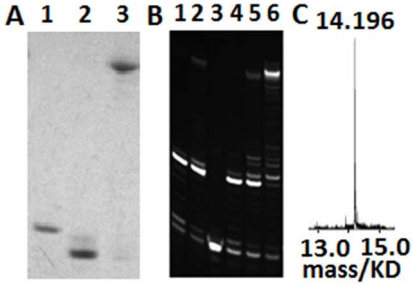


Figure 4. Copper-free RNA ligation and reverse transcription. A) 20% PAGE for non-templated copper free click ligation to synthesize (T8). Lane 1; 2′-azide oligo Z3, lane 2; alkyne oligo O1, lane 3; crude reaction mixture. B) 20% PAGE for reverse transcription of T8 and T9. Lane 1 and 2; reverse transcription of template T9 at 2h and 18h, lane 3; primer P1 (-7); lane 4, 5 and 6 reverse transcription of template T8 at 10 min, 2h and 18h. All reactions in Mg^{2+} buffer. C) Mass spectrum of RT products from template T8 and primer P7 after incubation at 37 °C overnight, Calc. mass: 14196 (M – dG + dA).

Because the BCN/azide triazole linker generated in the SPAAC reaction is much bulkier than the linker generated by CuAAC, and is a 5′- 2′ linkage rather than 5′- 3′, it was expected to hinder reverse transcription. Consequently, Click BCN backbones were reverse transcribed for an extended period (overnight) and M-MLV RT (RNase H-) was used to avoid RNA degradation. First the template with a CxC linkage (T8) was investigated as it would be expected to give a more stable duplex than the UxC linkage (T7). After 2h incubation at 37°C, both PAGE and MS showed that the RT enzyme paused at the triazole site (CxC), but then copied through during overnight incubation (Figure 4B, lanes 5 and 6 and S5). The 2h incubation gave two products; “stopped before triazole” and “full-length minus deoxyguanosine” (“-dG”). After overnight incubation, the “stopped before triazole” product disappeared, the primer extended to the end of the template and produced the “full-length -dG + dA” product (Figure 4C). This extra dA was most likely added to the 3′-end of the RT product as is commonly observed with polymerase enzymes. Changing the divalent cation in the buffer from Mg^{2+} to Mn^{2+} accelerated the

reverse transcription reaction, which was nearly completed in 2h instead of overnight (Figure 5A lanes 5 and 6) and the (full length - dG + dA) product was obtained. The shortest P1 (-7) primer worked more efficiently than the bridged primer P4 (+1), indicating that it is easier for the enzyme to pass the triazole-BCN linkage when it has a "running start" from a stretch of unmodified RNA than to start on top of the triazole-BCN link. This inefficiency of bridged primer extension was not observed when using the RNA template with the biocompatible CtC triazole linkage (T5), for which the bridged primer worked as efficiently as the -7 primer. This difference in behaviour can be explained by the length and bulkiness of the BCN triazole backbone. Template T7 with the sequence "UxC" instead of "CxC" was also investigated. It was anticipated that it would be copied inefficiently, as it would be expected to give a less stable duplex than the corresponding CxC template. Indeed, reverse transcription of this backbone in Mg^{2+} buffer read through the triazole-BCN linkage with difficulty. A reaction buffer containing 3 mM Mg^{2+} and 3 mM Mn^{2+} improved the situation, but reverse transcription produced a mixture of products including one and two nucleotide deletions. Both DNA[21] and RNA oligonucleotides have been used as 3´-adaptors for RNA ligation.[22] However, M-MLV RT reads through RNA more efficiently than DNA. To evaluate if this is also the situation with BCN-triazole-modified templates, T9 with a DNA adaptor instead of RNA adaptor was synthesized. Compared to template T8 which contains an RNA adaptor, reverse transcription of T9 was much less efficient (Figure 4B lanes 1 and 2). Addition of Mn^{2+} to the buffer greatly improved the situation but reverse transcription of T8 was still faster and finished in 2h compared to 16h for T9.

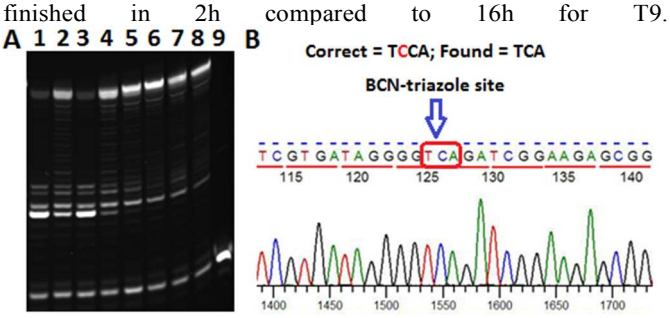


Figure 5. A) Reverse transcription of T8 in presence of Mg^{2+} or Mn^{2+} ions for 2h and 7h respectively. 20% PAGE. Lane 1 and 2; 3 mM Mg^{2+}, lane 3 and 4; 10 mM Mg^{2+}, lane 5 and 6; 3 mM Mn^{2+}, lane 7 and 8; 10 mM Mn^{2+}, lane 9; primer P1 (-7). B) Sanger sequencing analysis of the DNA. The cytidine (C) in red at the TCxCA site in the template is the one that is not copied by RT.

Sanger sequencing of reverse transcribed DNA showed that the RT products from the BCN-triazole templates formed by copper-free click chemistry failed to copy one of the two bases around the modified linkage (Figure 5B and S7). However, this would not be an issue in RNA sequencing applications; the identity of this base would be known, as it would be added to the 3´-end of the native RNA pool using an azide-modified NTP and Yeast poly-A polymerase (PAP).[11]

In conclusion, using bridged primers the M-MLV reverse transcriptase enzyme accurately copies an RNA template containing biocompatible triazole linkage 1A. When templates containing this or the BCN-triazole linkage 1B are used with normal primers, one nucleotide is omitted. The use of the SPAAC reaction to ligate RNA strands is significant because it should enable small natural RNAs to be linked to synthetic adaptors in non-templated copper-free RNA-denaturing conditions without degradation (Figure 2). The ability of RT to read through bulky linkages in RNA is likely to be a general property extending to other modified nucleic acid backbones. These findings could have applications in RNA isolation, sequence analysis and other biological applications.

Acknowledgments

This research was funded by the BBSRC sLOLA grant BB/J001694/1 "Extending the boundaries of nucleic acid chemistry." ATDBio provided funding for Xiong Chen.

Notes and references

[a] School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ. [b] Chemistry Branch, Dept. of Science and Mathematics, Faculty of Petroleum and Mining Engineering, Suez University, Suez, 43721, Egypt.[c] Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Road, Oxford, OX1 3TA, UK. *e.mail: tom.brown@chem.ox.ac.uk.

Electronic Supplementary Information (ESI) available: Chemical synthesis, table of oligonucleotide sequences, other reverse transcriptase reactions, DNA sequencing. See DOI: 10.1039/c000000x/

1. R. W. Carthew and E. J. Sontheimer, Cell, 2009, 136, 642-655.
2. D. C. Koboldt, K. M. Steinberg, D. E. Larson, R. K. Wilson and E. R. Mardis, Cell, 2013, 155, 27-38.
3. L. Mamanova and D. J. Turner, Nature Protocols, 2011, 6, 1736-1747.
4. S. E. V. Linsen, E. de Wit, G. Janssens, S. Heater, L. Chapman, R. K. Parkin, B. Fritz, S. K. Wyman, E. de Bruijn, E. E. Voest, S. Kuersten, M. Tewari and E. Cuppen, Nature Methods, 2009, 6, 474-476.
5. M. Hafner, N. Renwick, M. Brown, A. Mihailovic, D. Holoch, C. Lin, J. T. G. Pena, J. D. Nusbaum, P. Morozov, J. Ludwig, T. Ojo, S. J. Luo, G. Schroth and T. Tuschl, RNA-a Publication of the RNA Society, 2011, 17, 1697-1712.
6. A. D. Jayaprakash, O. Jabado, B. D. Brown and R. Sachidanandam, Nucleic Acids Res., 2011, 39.
7. F. L. Zhuang, R. T. Fuchs, Z. Y. Sun, Y. Zheng and G. B. Robb, Nucleic Acids Res., 2012, 40, 14.
8. K. Sorefan, H. Pais, A. E. Hall, A. Kozomara, S. Griffiths-Jones, V. Moulton and T. Dalmay, Silence, 2012, 3, 4.
9. A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown, Proc. Natl. Acad. Sci. U. S. A., 2011, 108, 11338-11343.
10. M. Shelbourne, A. H. El-Sagheer and T. Brown, Chem. Commun., 2012, 48, 11184-11186.
11. M. L. Winz, A. Samanta, D. Benzinger and A. Jaschke, Nucleic Acids Res., 2012, 40 e78.
12. C. N. Birts, A. P. Sanzone, A. H. El-Sagheer, J. P. Blaydes, T. Brown and A. Tavassoli, Angew. Chem. Int. Edit., 2014, DOI: 10.1002/anie.201308691.
13. A. H. El-Sagheer and T. Brown, Proc. Natl. Acad. Sci. U. S. A., 2010, 107, 15329-15334.
14. A. Rodriguez, E. Vigorito, S. Clare, M. V. Warren, P. Couttet, D. R. Soond, S. van Dongen, R. J. Grocock, P. P. Das, E. A. Miska, D. Vetrie, K. Okkenhaug, A. J. Enright, G. Dougan, M. Turner and A. Bradley, Science, 2007, 316, 608-611.
15. M. L. Kotewicz, C. M. Sampson, J. M. Dalessio and G. F. Gerard, Nucleic Acids Res., 1988, 16, 265-277.
16. E. Scolnick, E. Rands, S. A. Aaronson and G. J. Todaro, Proc. Natl. Acad. Sci. U. S. A., 1970, 67, 1789-&.
17. A. Dallmann, A. H. El-Sagheer, L. Dehmel, C. Mugge, C. Griesinger, N. P. Ernsting and T. Brown, Chem.-Eur. J., 2011, 17, 14714-14717.
18. M. Shelbourne, X. Chen, T. Brown and A. H. El-Sagheer, Chem. Commun., 2011, 47, 6257-6259.
19. J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. Rutjes, J. C. M. van Hest, D. J. Lefeber, P. Friedl and F. L. van Delft, Angew. Chem. Int. Edit., 2010, 49, 9422-9425.
20. M. Aigner, M. Hartl, K. Fauster, J. Steger, K. Bister and R. Micura, ChemBioChem, 2011, 12, 47-51.
21. M. Hafner, P. Landgraf, J. Ludwig, A. Rice, T. Ojo, C. Lin, D. Holoch, C. Lim and T. Tuschl, Methods, 2008, 44, 3-12.
22. R. D. Morin, M. D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A.-L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C. J. Eaves and M. A. Marra, Genome Res., 2008, 18, 610-621.