ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2014, **50**, 1313

Received 11th August 2013,
Accepted 16th October 2013

DOI: 10.1039/c3cc46132g

www.rsc.org/chemcomm

A modified dinucleotide for site-specific RNA-labelling by transcription priming and click chemistry†

Ayan Samanta,‡ André Krause‡ and Andres Jäschke*

An improved strategy for RNA labelling using an alkyne-carrying dinucleotide is reported. This involves near-quantitative priming by phage RNA-polymerases followed by conjugation of different labels using click chemistry. Moreover, these transcripts bear a ligation compatible 5′-end, and thus through ligation the terminal label can be transformed to an internal one.

Modified RNAs have proven to be indispensable in modern life-science research. While solid-phase RNA synthesis allows for the site-selective introduction of a variety of chemical modifications into short RNAs, the same task poses a major challenge for large RNAs.[1] Large unmodified RNAs can be conveniently prepared by *in vitro* run-off transcription using various bacteriophage RNA polymerases (RNAPs). Due to the relaxed substrate tolerance of these polymerases at the initiation step it is possible to selectively modify the 5′-end of a transcript using modified guanosine or adenosine analogues that are functionalised through the 5′-monophosphate.[2] These analogues can only be incorporated at the start of transcription because they lack triphosphate – hence termed as "initiator nucleotides". However, this approach suffers from two major drawbacks: (1) it requires the *de novo* synthesis of the complete initiator molecule and optimisation for its enzymatic incorporation for each different initiator, and (2) all currently known initiators generate RNA transcripts whose 5′-ends are blocked and cannot be ligated further. Moreover, as the transcriptional activities of different RNAPs vary substantially,[3] it is beneficial to have a single optimised labelling-protocol that is compatible with all commonly used polymerases. Therefore it is desirable to have an initiator nucleotide that can be incorporated into transcripts by a polymerase of choice under conditions that are optimised only once, bears a convertible residue for further postsynthetic conjugation with a variety of different labels and renders the 5′-end of the transcript accessible to ligation. To this end, the copper-catalysed azide–alkyne cycloaddition (CuAAC) seemed appropriate as a conjugation strategy due to the bioorthogonal nature of the reaction and the small size of the reactive groups.[4]

Based on this idea, we report here the synthesis and enzymatic incorporation of a "clickable" dinucleotide (OdUpG, where O represents octadiynyl) (Fig. 1) for labelling RNA at the 5′-end employing T3, SP6, or T7 RNAPs under the control of their cognate class III promoters. This initiator has been found to be incorporated by all three polymerases in a near quantitative manner in several instances. Additionally the applicability of this labelling strategy has been demonstrated by preparing various RNA-conjugates from a single transcript as well as by labelling long RNAs. Finally, the ligation compatibility of the primed transcript has been tested. Although the transcription-priming strategy is primarily a RNA 5′-labelling technique, this initiator renders it applicable to site-specific, internal modification of RNA after ligation (Fig. 1).

The dinucleotide initiator was synthesised following the standard phosphoramidite coupling strategy (ESI†) with an overall yield of 27%. Although all four nucleosides can be used for attachment of the alkyne functionality, uridine seemed appropriate as a first attempt due to the commercial availability of the building block and lack of a nucleobase protection strategy during phosphoramidite coupling. Moreover, the deoxy-sugar was chosen for its higher coupling efficiency compared to the ribo-analogue during phosphoramidite coupling. The choice of guanosine as a priming nucleotide was inevitable due to the widespread popularity of phage class-III promoters in molecular biology.

The incorporation of OdUpG by RNAPs was optimised at two levels – (a) at varying GTP concentrations (Fig. 2A, C and E) and (b) at varying NTP concentrations (Fig. 2B, D and F). While a large excess of initiator over GTP will result in a high incorporation of the dinucleotide into the transcripts, it will hamper the overall transcription yield since GTP is required during elongation too. The best NTP concentration was screened always with the optimal GTP concentration determined from the previous experiment. Since the incorporation of the dinucleotide itself

Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Im Neuenheimer Feld 364, Heidelberg 69120, Germany. E-mail: jaeschke@uni-hd.de; Fax: +49 6221 54 6430; Tel: +49 6221 54 4851

† Electronic supplementary information (ESI) available: Full experimental section. See DOI: 10.1039/c3cc46132g

‡ Contributed equally.

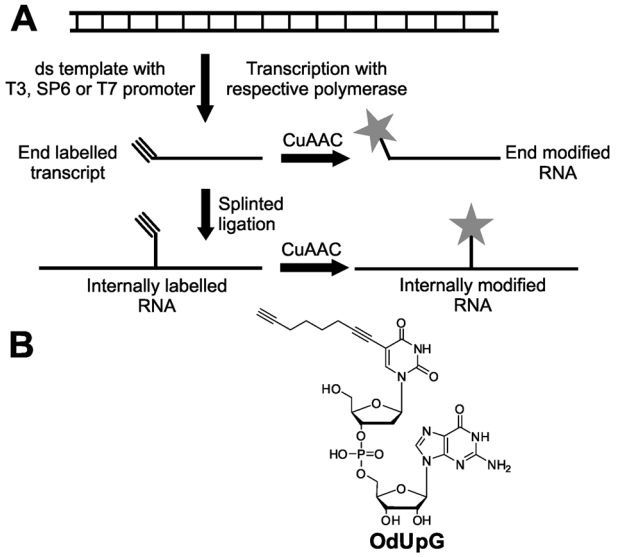

Fig. 1 Schematic representation of the RNA-labelling approach (A), chemical structure of the dinucleotide initiator (B).

induced only a slight shift in the electrophoretic mobility that was indistinguishable from the un-initiated $n + 1$ transcript, the appended alkyne function was subsequently conjugated to a biotinyl residue by CuAAC using biotin azide, thus enabling an unambiguous determination of the labelling efficiency by streptavidin electrophoretic mobility shift assay (Strep-EMSA). GMP-priming was used as a control in all of these experiments (Fig. 2, the first lane in every gel picture).

The labelling efficiencies and relative transcription yields are summarised in Fig. 2 and 3. Based on the utilisation of the initiator at varying GTP and NTP concentrations, the polymerases can be categorised into two broader groups, one consisting of T7 and SP6 RNAPs and the other of T3 RNAP. The relative transcription yield increased with increasing GTP concentrations (at least for the tested concentration range) for T7 and SP6 RNAPs whereas it decreased for T3 RNAP above a certain threshold concentration (0.4 mM) (Fig. 3A). The labelling efficiency decreased for all three RNAPs with increasing GTP concentrations, SP6 being affected the most while T7 the least (Fig. 3A). Furthermore, within the tested NTP concentration range, the transcription yield decreased only very slightly for T7 and SP6 RNAPs with increasing NTP concentrations whereas the decrease in transcription yield for T3 was found to be drastic (Fig. 3B). However, the labelling efficiencies remained largely constant for T3 RNAP, decreased only slightly for T7 and increased drastically for SP6 RNAP with increasing NTP concentrations (Fig. 3B). Based on these observations it can be concluded that if a transcription has to be performed at a relatively high NTP concentration, it is desirable to use SP6 RNAP since the transcription yield remains nearly constant but the labelling efficiency increases with increasing NTP concentrations. Similarly if a transcription needs to be performed at a relatively low NTP concentration, it is recommended to use T3 RNAP since this polymerase resulted in high transcription yield with a moderate labelling efficiency. However, under these low-NTP conditions

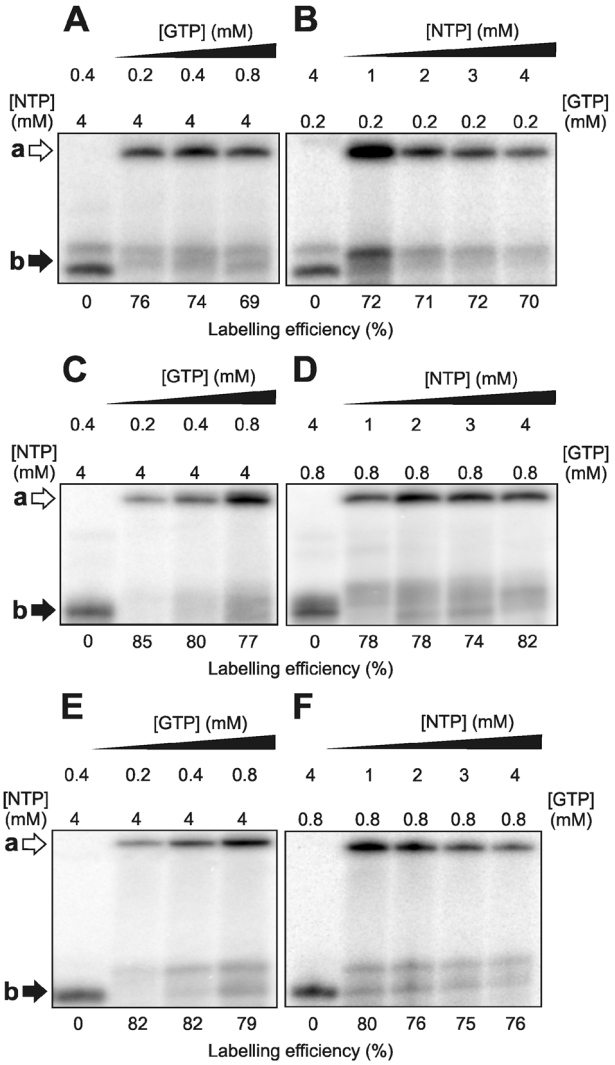

Fig. 2 Optimisation of transcriptional labelling by T3 (A and B), SP6 (C and D) and T7 (E and F) RNAPs. A, C and E denote labelling at varying GTP concentrations, while B, D, and F indicate the same at varying NTP concentrations. 'a' and 'b' denote streptavidin-shifted biotinylated RNA and unmodified RNA, respectively. The left lane in every gel denotes transcription initiation with GMP. The initiator nucleotide (GMP/OdUpG) was added to the transcription mixture at a concentration of 4 mM in every experiment. The purified transcript was subjected to CuAAC with biotin azide (compound 5), followed by incubation with streptavidin and gel analysis.

if the transcription yield is not the major concern but the labelling efficiency is, it is recommended to use T7 RNAP since the best labelling efficiency could be achieved with this polymerase while maintaining a moderate transcription yield.

Although the optimal transcription protocol is always a compromise between labelling efficiency and transcription yield, the best conditions in our opinion were found to be (1) 4 mM initiator, 0.2 mM GTP and 1 mM of the other NTPs for T3 RNAP with a labelling efficiency of 72%, (2) 4 mM initiator, 0.8 mM GTP and 4 mM of the other NTPs for SP6 RNAP with a labelling efficiency of 82% and (3) 4 mM initiator, 0.8 mM GTP and 1 mM of the other NTPs for T7 RNAP with a labelling efficiency of 80%. Initiated transcripts were further characterised by high resolution

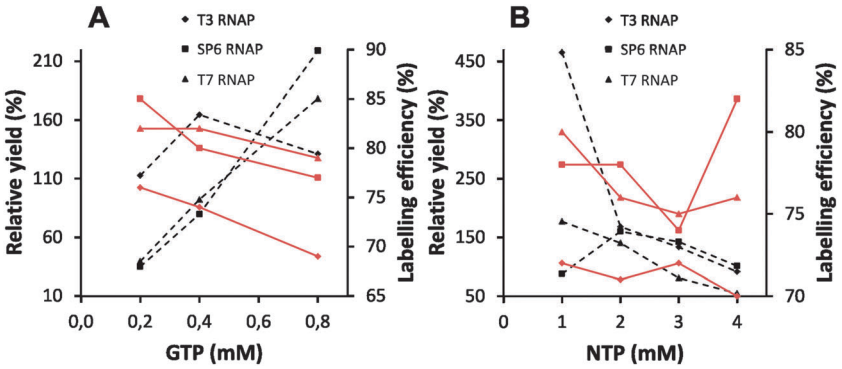

Fig. 3 Overview of the relative transcription yields and labelling efficiencies at varying GTP (A) and NTP (B) concentrations. Broken and solid lines denote relative yields and labelling efficiencies, respectively. Relative yields were calculated with respect to the corresponding GMP-primed reactions in Fig. 2 (see ESI† for calculation details).

mass spectrometry (ESI,† Table S1). To the best of our knowledge, this is the only study reporting comparative data on individual labelling efficiencies and transcription yields of T3, SP6 and T7 RNAPs with an initiator nucleotide.

To demonstrate the robustness of this click-type labelling approach, transcription priming was performed with T7 RNAP using the aforesaid optimised conditions from a template carrying two 2′-OMe substitutions (ESI,† Table S2).[5] The purified primed-transcript was then reacted by CuAAC with a variety of commercial and self-synthesised compounds (ESI,† Fig. S2B). Attachment of these various tags resulted in substantial gel-shifts for the products compared to the starting materials and thus enabling a straightforward assessment of the click-conjugation efficiency (ESI,† Fig. S2A). These data indicate near-quantitative conversion in all cases, irrespective of the nature of the coupling partner. This represents a major advancement in comparison to all other previously described initiators, as one central intermediate is sufficient to synthesise a wide variety of different conjugates without the need to synthesise dozens of initiator nucleotides and to optimise their enzymatic incorporation.

Furthermore, a glycine riboswitch (length 209 nucleotides) from *Bacillus subtilis* and a random RNA pool (length 233 nucleotides) (ESI†) were subjected to transcriptional labelling with OdUpG by T7 RNAP under the aforesaid optimised conditions. The resulting transcripts were conjugated with a Cy5-azide by CuAAC (ESI,† Fig. S3A and B, respectively). Both of these experiments clearly demonstrate the applicability and sequence-independence of this RNA-labelling approach for labelling long, structured RNAs that are beyond the limits of standard solid-phase chemistry.

The presence of a free 5′-OH in the OdUpG primed transcript renders the transcript to be amenable to splinted ligation with another RNA strand. Thus the 5′-terminal modification will be transformed to an internal one as part of a ligation fragment. This would however require the non-natural 5′-nucleoside to be accepted as a substrate by both a polynucleotide kinase and a ligase. To test this hypothesis, an OdUpG-primed transcript (25 nucleotides) was subjected to enzymatic 5′-end phosphorylation using T4 polynucleotide kinase (PNK), followed by ligation to an acceptor RNA (40 nucleotides) in the presence of a complementary DNA splint (65 nucleotides) (ESI,† Table S2). Nearly 7-fold excess of the acceptor RNA over the donor strand along with a mixture of T4 RNA ligase 2, T4 RNA ligase 1 and T4 DNA ligase was found to be necessary to achieve efficient ligation (ESI,† Fig. S4). The requirement of RNA ligase 1, a single strand specific ligase, is plausible due to the fact that the currently designed splint does not incorporate the added alkyne-bearing nucleoside in a duplex environment, thus creating a single nucleotide bulge at the ligation joint. Current efforts include the optimisation of the ligation protocol for improved ligation yield at stoichiometric ratios of the donor to acceptor strand with a perfectly double stranded ligation site. However, it can be concluded, by comparing this result with the ligation results obtained from an unmodified UpG-primed transcript, that the OdUpG initiator is indeed tolerated by three different classes of enzymes widely used in molecular biology, namely RNA polymerases, polynucleotide kinases and ligases. The resulting ligated RNA can further be derivatised by CuAAC, thereby enabling site-specific, internal modification of RNA, which considerably broadens the scope of this approach.

The combination of this approach with a previously reported tailing-based click modification method for RNA 3′-ends and internal positions would allow considerable freedom in the choice of the labelling positions, and also allows for sequential multiple labelling.[6]

Although in this study the initiator dinucleotide was self-synthesised, it can also be purchased from various oligonucleotide synthesis companies owing to the commercial availability of the monomers. Hence this labelling strategy is completely open to people lacking any expertise or equipment in organic synthesis. Moreover, since this labelling-protocol is compatible with all commonly used RNAPs in molecular biology, this approach can directly be implemented on one's RNA of interest without any further cloning or PCR manipulation of the available plasmids. In combination with the well-established protocols for enzymatic *in vitro* RNA synthesis, we expect that the initiator dinucleotide described in this study will help to overcome the present shortcomings of chemical RNA synthesis, thereby allowing for the introduction of variable modifications at the 5′-terminus or at internal positions of RNA by click chemistry.

This work was supported by the DFG Grant #Ja 794/3. The authors acknowledge S. Ameta, S. Suhm and A. Hertl.

Notes and references

1 H. Lönnberg, *Bioconjugate Chem.*, 2009, **20**, 1065–1094; E. Paredes, M. Evans and S. R. Das, *Methods*, 2011, **54**, 251–259.
2 S. Fusz, S. G. Srivatsan, D. Ackermann and M. Famulok, *J. Org. Chem.*, 2008, **73**, 5069–5077; F. Huang, J. He, Y. Zhang and Y. Guo, *Nat. Protocols*, 2008, **3**, 1848–1861; F. Huang and Y. Shi, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 6254–6257; N. Li, C. Yu and F. Huang, *Nucleic Acids Res.*, 2005, **33**, e37; E. Paredes and S. R. Das, *ChemBioChem*, 2011, **12**, 125–131; C. Pitulle, R. G. Kleineidam, B. Sproat and G. Krupp, *Gene*, 1992, **112**, 101–105; J. Schoch, S. Ameta and A. Jäschke, *Chem. Commun.*, 2011, **47**, 12536–12537; R. Fiammengo, K. Musilek and A. Jäschke, *J. Am. Chem. Soc.*, 2005, **127**, 9271–9276; B. Seelig and A. Jäschke, *Bioconjugate Chem.*, 1999, **10**, 371–378; J. Wolf, V. Dombos, B. Appel and S. Müller, *Org. Biomol. Chem.*, 2008, **6**, 899–907.
3 I. D. Pokrovskaya and V. V. Gurevich, *Anal. Biochem.*, 1994, **220**, 420–423.
4 A. H. El-Sagheer and T. Brown, *Chem. Soc. Rev.*, 2010, **39**, 1388–1405; A. H. El-Sagheer and T. Brown, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 15329–15334; J. Gierlich, G. A. Burley, P. M. Gramlich, D. M. Hammond and T. Carell, *Org. Lett.*, 2006, **8**, 3639–3642; J. Gierlich, K. Gutsmiedl, P. M. Gramlich, A. Schmidt, G. A. Burley and T. Carell, *Chemistry*, 2007, **13**, 9486–9494; V. Hong, N. F. Steinmetz, M. Manchester and M. G. Finn, *Bioconjugate Chem.*, 2010, **21**, 1912–1916; T. Ishizuka, M. Kimoto, A. Sato and I. Hirao, *Chem. Commun.*, 2012, **48**, 10835–10837; C. Y. Jao and A. Salic, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15779–15784; Y. Motorin, J. Burhenne, R. Teimer, K. Koynov, S. Willnow, E. Weinhold and M. Helm, *Nucleic Acids Res.*, 2011, **39**, 1943–1952; J. Qiu, A. H. El-Sagheer and T. Brown, *Chem. Commun.*, 2013, **49**, 6959–6961; H. Rao, A. A. Sawant, A. A. Tanpure and S. G. Srivatsan, *Chem. Commun.*, 2012, **48**, 498–500.
5 C. Kao, S. Rudisser and M. Zheng, *Methods*, 2001, **23**, 201–205.
6 M. L. Winz, A. Samanta, D. Benzinger and A. Jäschke, *Nucleic Acids Res.*, 2012, **40**, e78.

