COMMUNICATION

Vitamin-Guanosine Monophosphate conjugates for *in vitro* transcription priming

Received 00th January 20xx,
Accepted 00th January 20xx

Nikolaos Papastavrou,^a Omprakash Bande,^a Philippe Marlière,^{b*} Elisabetta Groaz,^{a*} and Piet Herdewijn^{a*}

DOI: 10.1039/x0xx00000x

Expanding the catalytic repertoire of ribozymes to include vitamin synthesis requires efficient labelling of RNA with the substrate of interest, prior to *in vitro* selection. For this purpose, we rationally designed and synthesized six GMP-conjugates carrying a synthetic pre-thiamine or biotin precursor and investigated their transcription incorporation properties by T7 RNA polymerase.

Ribonucleic acid (RNA) has come a long way since being first described as an information-carrying molecule in the “central dogma”.¹ At present, it is well known that RNA can catalyse reactions (e.g., the ubiquitous peptide bond formation), regulate gene expression (small regulatory RNAs or riboswitches), and even act as a defence mechanism in bacteria (RNA interference).^{2, 3} Efforts aimed at further elucidating the structure and biological function of RNA as well as expanding its catalytic repertoire and therapeutic potential rely on its ability to be chemically modified. While chemical derivatisation of ribonucleotides can occur naturally to some extent,⁴ the development of general methods to introduce more diverse and functional modifications (alternative sugars, affinity tags, fluorescent dyes) remains an important goal. Generally, RNA labelling can be achieved either synthetically, through solid-phase synthesis, or biochemically, with the aid of enzymes. In some cases, a post-synthetic approach can be adopted, where a nucleotide bearing a reactive group, such as a thiol or activated ester, is incorporated into an RNA strand by using one of the two aforementioned methods and later conjugated with more complex molecules through the appropriate chemistry.⁵

Solid-phase synthesis offers a direct way of incorporating modified nucleotides in a sequence-specific manner. Furthermore, it allows to accommodate a variety of 2'-modifications and changes to the sugar moiety, thus finding applications in antisense technology.⁶ Nonetheless, this method is limited by the availability of the corresponding phosphoramidite building blocks as well as the size of the resulting RNA molecules, which are typically up to 100 nucleotides long. On the other hand, the enzymatic synthesis of RNA does not pose such size restriction, but entails little sequence specificity. Despite the recognised ability of natural enzymes such as T4 RNA ligase⁷ or, as recently demonstrated, an *in vitro* selected ribozyme⁸ to carry out end modifications, internal labelling and the majority of end-modified RNA is achieved by using T7 RNA polymerase (T7 RNAP). This viral polymerase has been shown to incorporate an adequate number of sugar- and base- modifications, thus allowing for fluorescent labelling, cross-linking, and pull-down studies as well as post-transcriptional functionalisation.⁹ This method has proven especially useful in the preparation of 5'-modified RNA libraries for use in ribozyme selection and evolution through SELEX (Systemic Evolution of Ligands by Exponential enrichment). Typically, a ribozyme substrate is chemically conjugated to a purine, most often guanosine monophosphate (GMP), and used for transcription priming by T7 RNAP at the beginning of each round.¹⁰⁻¹² Alternatively, the substrate can be post-transcriptionally attached to an appropriately 5'-modified RNA pool.¹³⁻¹⁵ This technique has resulted in a handful of successful ribozyme selections with interesting chemistries, such as the carbon-carbon forming Diels-Alder¹⁶ and aldol reaction.¹⁷

In vitro selection has also allowed for the identification of ribozymes that mimic the function of enzymes involved in metabolic reactions. For instance, RNA sequences that either catalyse the synthesis of common coenzymes from their precursors or utilise essential cofactors for complex reactions have been reported in the literature.^{18, 19} [ENREF 17](#) With the

^a KU Leuven, Rega Institute for Medical Research, Medicinal Chemistry, Herestraat 49-box 1041, 3000 Leuven, Belgium

^b Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 2 Rue Gaston Crémieux 91057 Evry, France

* piet.herdewijn@kuleuven.be, elisabetta.groaz@kuleuven.be, phmarliere@gmail.com

Electronic Supplementary Information (ESI) available: Experimental section, synthetic schemes, gel pictures, HPLC profiles, ¹H, ¹³C, and ³¹P NMR spectra. See DOI: 10.1039/x0xx00000x

ultimate goal of establishing an autonomous RNA-catalysed metabolic

Fig. 1 Proposed catalytic reactions mediated by RNAzymes.

network like the one thought to exist in a protocellular environment,²⁰ we are currently pursuing the *in vitro* selection of two RNAzymes, i.e., one that could catalyse the hydrolysis of a phosphorothiolate analogue of an intermediate in thiamine (vitamin B₁) biosynthesis and another that could convert a ring-opened precursor of biotin (vitamin H or B₇) to the active cofactor (Fig. 1). To this end, we describe here the chemical synthesis and transcription incorporation of a total of six vitamin-GMP conjugates, i.e., four bearing a pre-thiamine precursor and two carrying a pre-biotin analogue (Fig. 2), which is a prerequisite for the envisaged selection.

The pre-thiamine initiators were designed to incorporate a phosphorothiolate bond for two main reasons. Firstly, the hydrolysis of a P-S bond falls within the scope of RNA catalysis, since an *in vitro* selected ribozyme was previously demonstrated to catalyse this type of reaction.²¹ Secondly, it was envisaged that the successful cleavage of this bond could provide a selectable functionality, since the resulting active RNAs would carry a terminal thiol group. A flexible hydrophilic linker was introduced between the substrate and the GMP residue (**T1-T4**). This is a common practice,²² which allows for bulky GMP-conjugates to be incorporated by **T7** RNAP, whilst providing a degree of freedom for the substrate to reach distant catalytic sites on the folded RNA. In this study, a tetra(ethylene glycol) was chosen (**T1, T3-T4**) and compared with a longer hexa(ethylene glycol) chain (**T2**) to assess the influence of the distance between the substrate and GMP on the transcription incorporation efficiency. The chemical environment next to the reaction site was further varied through the insertion of either an aromatic sulfonamide (**T3**) or heterocyclic triazole (**T4**). The purpose of these extra chemical moieties was to increase the binding of the substrate to RNA through stacking interactions or hydrogen bonds, thus fine-tuning the **T7** RNAP's substrate tolerance. On the other hand, the proposed pre-biotin initiators shared a very similar structure, which varied only at one position (R). Conjugate **B1** is a carbamate analogue, whereas conjugate **B2** is a urea derivative. This difference, however, is expected to have a big impact on the reaction rate, since the urea moiety is more stable to hydrolysis compared to the

carbamate functionality. Here, no additional chemical groups were added near the reaction site and a tetra(ethylene glycol)

Fig. 2 General structures (framed) of the four thiamine (top) and two biotin (bottom) initiators synthesised in this study. The different substituents are shown in red.

linker was used in both cases.

The synthetic routes and experimental procedures utilized for the preparation of all six initiator molecules are described in detail in the Supplementary information (Supplementary Schemes S1-S2). Briefly, the stepwise assembly of three common synthons (substrate, linker, GMP) was the basis for all syntheses, with the only difference being the order of conjugation. For the thiamine initiators, the glycol linker was first appropriately modified and attached to a suitably protected guanosine through standard phosphoramidite chemistry. The resulting GMP-linker was then conjugated to a thiophosphate containing substrate by either direct coupling or “click” chemistry. Subsequent removal of all protecting groups, resulted in compounds **T1-T4**. In contrast, for the pre-biotin initiators, the reaction site was built on the biotin residue, which was then coupled to the linker through the formation of an amide bond. The substrate-linker was then coupled to a guanosine phosphoramidite, followed by complete deprotection to afford compounds **B1** and **B2**.

With this set of molecules in hand, we then sought to test their ability to act as substrates for **T7** RNAP and individually determine the optimal conditions of RNA incorporation. To this end, a short (34-mer) DNA template bearing the consensus **T7**

Fig. 3 5'-RNA labelling with GMP-conjugates **T1-T4** and **B1-B2**. **A**. Schematic representation of the transcription incorporation experiments. A short (34-nt) double stranded DNA template bearing the **T7** promoter sequence (diagonal lines) and two 2'-O-methyl modified nucleotides (yellow stars) was transcribed *in vitro* by **T7** RNA polymerase (**T7** RNAP) in the presence of the different initiators (X-GMP), and the transcription products were analysed by denaturing PAGE. **B**. Bar graphs summarising the incorporation efficiency of each individual initiator.

promoter sequence and two methylated residues at the 5'-end of its non-coding strand was *in vitro* transcribed in the presence of different GTP and initiator concentrations. The transcription products were then analysed by denaturing polyacrylamide gel electrophoresis (PAGE) (Fig. 3a). The choice for a short DNA template with two terminally modified residues on the antisense strand was motivated by the inherent difficulty in resolving the 5'-modified from the native transcripts. The absence of any functional groups on the initiator molecules allows only for size separation and shorter sequences are easier to separate even when they differ by a single nucleotide. Furthermore, it is well-established that run-off transcription by **T7** RNAP results in the addition of non-encoded nucleotides at

the 3'-end of the nascent RNA ("n+1 band")¹¹. The modification of the last two nucleotides at the 5'-end of the DNA template with a 2'-O-methyl group, however, has shown to greatly attenuate the appearance of the n+1 band,²³ although not leading to its complete elimination. Moreover, transcription products were purified and analysed by mass spectrometry to verify the chemical stability of the GMP-conjugates post-transcriptionally (Supplementary information). The results of the RNA labelling experiments are summarized in Fig. 3b (gel pictures can be found in the Supplementary information, Supplementary Figures S3-S7) and are expressed as percent of RNA yield compared to a standard *in vitro* transcription (no initiator, 2 mM GTP). Three different GTP-to-

initiator ratios (1:1 to 1:4) for three different GTP concentrations (0.5–2 mM) were tested in order to have a comprehensive view of the different ability of each GMP-conjugate to initiate transcription.

All six compounds were found to be incorporated during *in vitro* transcription, however the two classes of initiators exhibited a very different behaviour. With regard to the thiamine-modified nucleotides, increasing concentrations of the initiator led to higher incorporation efficiencies at the expense of RNA yields, which is in line with previous findings.²⁴ This effect was observed to be more pronounced for compounds **T2** and **T3**, which have a longer spacer between the substrate and GMP moiety. These initiators showed good incorporation ratios and RNA yields at lower GTP (and initiator) concentrations (0.5–1 mM), however a nearly total inhibition of **T7** RNAP occurred at higher concentrations. In contrast, GMP-conjugates **T1** and **T4** exhibited a more consistent transcription profile with generally lower labelling efficiencies. Notably, phenyl sulphonamide **T3** emerged as the preferred substrate for transcription initiation when added in a 4-fold excess in the presence of 1 mM GTP, thus leading to the highest percentage of 5'-modified RNA among the members of its class. Nonetheless, its laborious synthesis makes initiator **T2** an attractive alternative for *in vitro* selection, where high quantities of compound are typically needed. Conversely, the biotin-modified nucleotides displayed a more favourable labelling profile than the thiamine based molecules as they were readily accepted by **T7** RNAP as substrates, despite their bulky size. For both initiator nucleotides, an increase in GTP concentration and initiator-to-GTP ratio was translated into a higher transcription yield and RNA labelling, respectively. Here, a structure lacking a negatively charged phosphorothiolate group seems to be better tolerated by the polymerase. A predominantly 5'-modified RNA population was generated with either initiator, rendering the choice solely based on the preferred stability.

In conclusion, efficient synthetic procedures were established for the preparation of six GMP-conjugates bearing a thiamine precursor or biotin analogue. These modified nucleotides were designed for the ultimate *in vitro* selection of metabolically relevant ribozymes. Herein, their ability to initiate transcription was assayed and optimal GTP and initiator concentrations were determined for each compound, leading to the identification of the preferred initiator for each structural class. Work is now underway to establish a suitable selection methodology for each type of reaction and identify active RNA sequences that can utilise these initiators as substrates.

N.P. was supported by an FWO Aspirant grant (11V4218N). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007–2013)/ERC (grant agreement no. ERC2012-ADG 20120216/320683). The authors would like to extend their gratitude to Luc Baudempretz for recording the NMR spectra, Jef Rozenski for the mass spectrometry analysis and Chantal Biernaux for editorial help.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1. F. Crick, *Nature*, 1970, **227**, 561–563.
2. M. J. Fedor and J. R. Williamson, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 399–412.
3. S. R. Eddy, *Nat. Rev. Genet.*, 2001, **2**, 919–929.
4. M. Helm and Juan D. Alfonzo, *Chem. Biol.*, 2014, **21**, 174–185.
5. E. Defrancq, Y. Singh and N. Spinelli, *Curr. Org. Chem.*, 2008, **12**, 263–290.
6. T. P. Prakash, *Chem. Biodivers.*, 2011, **8**, 1616–1641.
7. K. Cole, V. Truong, D. Barone and G. McGall, *Nucleic Acids Res.*, 2004, **32**, e86–e86.
8. B. Samanta, D. P. Horning and G. F. Joyce, *Nucleic Acids Res.*, 2018, **46**, e103–e103.
9. M. Flamme, L. K. McKenzie, I. Sarac and M. Hollenstein, *Methods*, 2019, **161**, 64–82.
10. J. Wolf, V. Dombos, B. Appel and S. Muller, *Org. Biomol. Chem.*, 2008, **6**, 899–907.
11. R. Fiammengio, K. Musilek and A. Jaschke, *J. Am. Chem. Soc.*, 2005, **127**, 9271–9276.
12. F. Huang, *Nucleic Acids Res.*, 2003, **31**, e8.
13. M. Wecker, D. Smith and L. Gold, *RNA*, 1996, **2**, 982–994.
14. J. C. Schlatterer and A. Jaschke, *Biochem. Biophys. Res. Commun.*, 2006, **344**, 887–892.
15. S. Tsukiji, S. B. Pattnaik and H. Suga, *Nat. Struct. Biol.*, 2003, **10**, 713–717.
16. B. Seelig and A. Jaschke, *Chem Biol*, 1999, **6**, 167–176.
17. S. Fusz, A. Eisenfuhr, S. G. Srivatsan, A. Heckel and M. Famulok, *Chem Biol*, 2005, **12**, 941–950.
18. P. Cernak and D. Sen, *Nat. Chem.*, 2013, **5**, 971–977.
19. F. Huang, C. W. Bugg and M. Yarus, *Biochemistry*, 2000, **39**, 15548–15555.
20. G. F. Joyce and J. W. Szostak, *Cold Spring Harb. Perspect. Biol.*, 2018, **10**.
21. T.-P. Wang, Y.-C. Su, Y. Chen, Y.-M. Liou, K.-L. Lin, E.-C. Wang, L.-C. Hwang, Y.-M. Wang and Y.-H. Chen, *Biochemistry*, 2012, **51**, 496–510.
22. B. Seelig and A. Jäschke, *Bioconjug. Chem.*, 1999, **10**, 371–378.
23. C. Kao, M. Zheng and S. Rüdisser, *RNA*, 1999, **5**, 1268–1272.
24. A. Samanta, A. Krause and A. Jaschke, *Chem. Commun. (Camb.)*, 2014, **50**, 1313–1316.