Synthesis and multiple incorporations of 2≠-*O*-methyl-5-hydroxymethyl-, 5-hydroxymethyl- and 5-formylcytidine monomers into RNA oligonucleotides

Arun A. Tanpure and Shankar Balasubramanian*

Abstract: The synthesis of 2≠-*O*-methyl-5-hydroxymethylcytidine (hm^{5}Cm), 5-hydroxymethylcytidine (hm^{5}C) and 5-formylcytidine (f^{5}C) phosphoramidite monomers has been developed. Optimization of mild post-synthetic deprotection conditions enabled the synthesis of RNA containing all four naturally occurring cytosine-modifications (hm^{5}Cm, hm^{5}C, f^{5}C plus 5-methylcytosine). Given the considerable interest in RNA modifications and epitranscriptomics, the availability of synthetic monomers and RNAs containing these modifications will be valuable to elucidate their biological function(s).

Post-transcriptional chemical modifications in RNA are more diverse and complex than epigenetic modifications in DNA and in histones. So far more than 140 chemically distinct RNA modifications have been identified in various species. The majority of these modifications involve methylation, of which 2≠ ribose sugar methylation is the most abundant.[1] These modifications were initially considered as static and stable marks, however, recent studies have revealed their dynamic nature and involvement in important gene regulatory functions.[2] For instance, *N*^{6}-methyladenosine (m^{6}A) a predominant internal modification in eukaryotic messenger RNA (mRNA) can be oxidatively converted to adenosine by demethylase such as fat mass and obesity-associated protein (FTO) and alkB homologue 5 (ALKBH5).[3] This reversible adenosine methylation is proposed to be involved in RNA maturation, protein translation and gene expression.[4] 5-Methylcytidine (m^{5}C), is another important methylated ribonucleoside that exists in transfer RNA (tRNAs), ribosomal RNA (rRNAs), and in the untranslated regions of mRNAs.[2,5] We and others have demonstrated that m^{5}C can be oxidatively metabolised to produce 5-hydroxymethylcytidine (hm^{5}C) and 5-formylcytidine (f^{5}C) (Figure 1).[6,7] Additionally, it is proposed that m^{5}C is essential for mRNA export and post-transcriptional regulation.[8] Significant enrichment of hm^{5}C in polyadenylated RNA compared to total RNA further suggests that biosynthesis of hm^{5}C might be a part of a dynamic regulatory mechanism of RNA.[6b,9] The f^{5}C modification is prevalent at the wobble position of an anticodon loop of mitochondrial methionine tRNA in many species including human. It was observed that f^{5}C provides flexibility to the loop of tRNA and affords an ability to decode both AUG and AUA in translational initiation and elongation sites of mRNA.[10] Chemical labeling coupled with liquid chromatography-mass spectrometry (LC-MS) analysis revealed the existence of 5-carboxycytidine (ca^{5}C in mouse liver tissue, albeit with a very low concentration.[11] Recently, we discovered the existence of 2≠-*O*-methyl-5-hydroxymethylcytidine (hm^{5}Cm), a second methylated metabolite of m^{5}C, in RNA of higher organisms (Figure 1).[12] Formation of hm^{5}Cm from m^{5}C by stepwise oxidation via hm^{5}C as an intermediate supports the dynamic nature and complexity of these cytosine-modifications in RNA.[12-13] An efficient synthesis of oligonucleotides (ONs) containing these modifications is essential to elucidate the chemistry and function of RNA cytosine derivatives. For example, site-specific incorporation of these cytosine-modifications in RNA will enable the development of sequencing methods to decode the modification[14] and also help identify the reader proteins to understand the cellular functions of these modifications.[8]

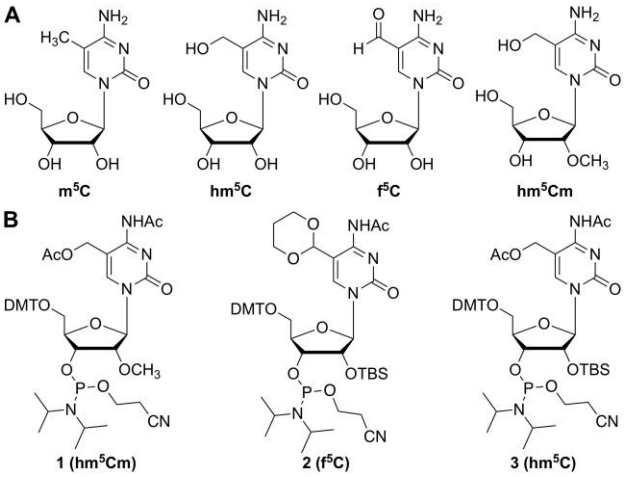


Figure 1. A) Chemical structure of major cytosine-modifications known to exist in eukaryotic RNA. **B)** Corresponding phosphoramidite monomers synthesized in this report.

A phosphoramidite monomer used for the synthesis of m^{5}C containing RNA ONs is commercially available, while syntheses of f^{5}C and hm^{5}C monomers have been reported in the literature.[10,15] We set out to develop a synthesis of a hm^{5}Cm monomer compatible with solid-phase RNA ONs synthesis, which has not yet been described. In addition, the reported monomer for the synthesis of f^{5}C containing RNA ONs requires a multi-step synthesis via 2′-bis(2-acetoxyethoxy)methyl (ACE) orthoester and 5′-*O*-benzhydroxy-bis (trimethylsiloxy)silyl (BZH) protection.[10] A drawback of this monomer is the presence of a free formyl group, which is susceptible to oxidation and nucleophilic attack encountered during ON assembly and post-synthetic resin cleavage respectively.[10,16] Consequently, multiple incorporations using the existing f^{5}C phosphoramidite is difficult, especially in the synthesis of longer oligomers and in the

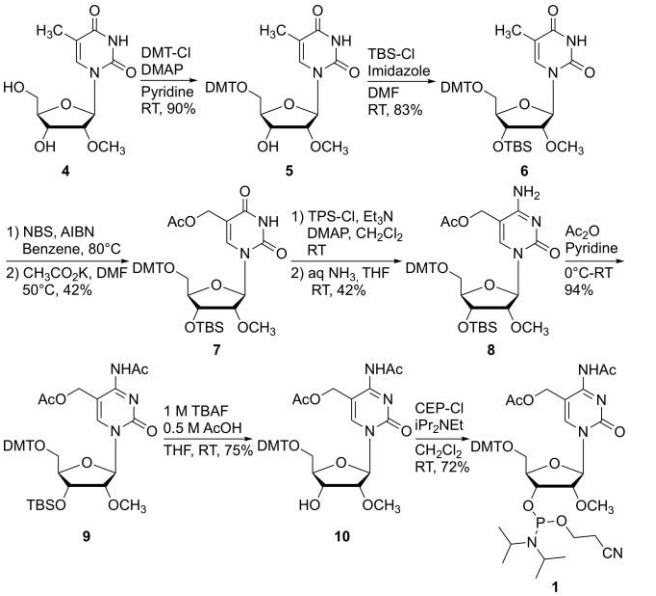
Dr. A. A. Tanpure, Prof. S. Balasubramanian
Department of Chemistry, University of Cambridge Lensfield Road, Cambridge, CB2 1EW (UK)
E-mail: sb10031@cam.ac.uk
Prof. S. Balasubramanian
Cancer Research, UK, Cambridge Institute, Li Ka Shing Centre University of Cambridge Robinson Way, Cambridge, CB2 0RE (UK) and
School of Clinical Medicine, University of Cambridge Cambridge, CB2 0SP (UK)

Supporting information for this article is given via a link at the end of the document.

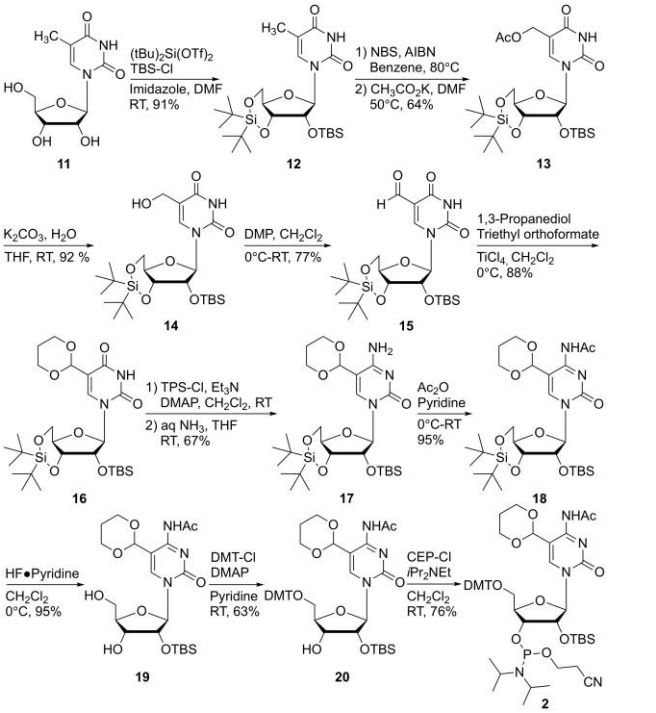
combination with other cytosine-modifications such as hm^{5}C and hm^{5}Cm containing a nucleophilic moiety. To overcome these practical limitations we designed an alternative f^{5}C monomer in which the reactive formyl group is appropriately masked. Riml et al. reported the synthesis and incorporation of hm^{5}C monomer into RNA ONs, their route provides this monomer in a 3% overall yield in eight steps from 5-hydroxymethyluridine (hm^{5}U),[15] which itself is obtained in a three step synthesis from uridine and totalling the number of steps to 11.[17] We therefore, considered the development of a faster and more efficient route to this monomer at the same time. The presence of the 2≠ hydroxyl group, makes the synthesis and incorporation of functionally-modified ribonucleoside phosphoramidites particularly challenging, compared to the analogous DNA phosphoramidite.[10,15,18] Moreover, an orthogonal protection strategy developed for DNA ON synthesis does not necessarily work for RNA due to its inherently more labile nature.[18a] Herein, we describe the development of phosphoramidite building blocks of hm^{5}Cm, f^{5}C and hm^{5}C (Figure 1) and demonstrate the incorporation of all these cytosine-modifications into RNA ONs on multiple positions and combinations by solid-phase RNA synthesis.

To synthesize hm^{5}Cm phosphoramidite **1** we decided to protect the 5-hydroxymethyl with an acetyl group, which is compatible with the incorporation of hm^{5}C into RNA ONs.[15] Starting with commercially available 2'-*O*-methyl-5-methyluridine **4** we protected the 5≠-hydroxyl with a 4,4'-dimethoxytrityl (DMT) and the 3≠-hydroxyl with a *tert*-butyldimethylsilyl (TBS) group (scheme 1).[19] The DMT and TBS protected uridine **6** was then subjected to azobisisobutyronitrile (AIBN) catalyzed bromination of 5-methyl group using *N*-bromosuccinimide (NBS).[20] The crude bromo-derivative was treated with potassium acetate to yield the fully protected uridine analogue **7** in moderate yield.[19] The reaction of **7** with 2,4,6-triisopropylbenzenesulfonyl chloride (TPS-Cl) resulted into regioselective O^4-trisylation, which was readily converted into cytidine analogue **8** upon ammonolysis. Acetylation of the exocyclic amino group was achieved using acetic anhydride in pyridine to provide nucleoside **9**. Cleavage of the 3′-*O*TBS group with TBAF and acetic acid in THF gave precursor **10**. Finally, phosphitylation of the 3′-hydroxyl with 2-cyanoethyl-*N*,*N*-diisopropylchloro-phosphoramidite (CEP-Cl) in the presence of *N*,*N*-diisopropylethylamine afforded phosphoramidite **1** in a 6.7% overall yield in seven steps from **4** (Scheme 1).[19]

In DNA synthesis, acetal chemistry has been successfully utilized to protect the formyl group.[21] To explore the analogous protection of f^{5}C for RNA, we synthesized phosphoramidite building block **2** as depicted in scheme 2.[19] Starting with 5-methyluridine **11**, this time we protected the 3'- and 5'-hydroxyl with the cyclic di-*tert*-butylsilylene and subsequently 2'-hydroxyl with TBS group.[22] Nucleoside **12** was selectively brominated at the C5-methyl group followed by acylation yielded uridine-derivative **13** in good yield. Hydrolysis of **13** provided 5-hydroxymethyl derivative **14,** which was then converted into aldehyde **15** by Dess–Martin periodinane (DMP) oxidation.[23] Next, the formyl group was protected as a 1,3-dioxane using propane-1,3-diol in presence of $TiCl_4$, following the procedure described for synthesizing an analogous 2'-deoxy phosphoramidite.[21] Fully protected uridine analogue **16** was then treated with TPS-Cl, followed by ammonolysis to give cytidine analogue **17**. Acetylation of the exocyclic amino group was carried out using acetic anhydride in pyridine to provide **18.**



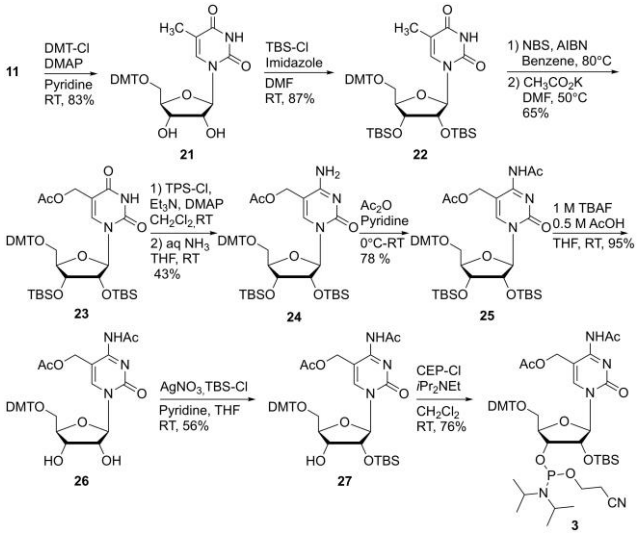
Scheme 1. Synthesis of hm^{5}Cm phosphoramidite **1**. (DMT: 4,4′-dimethoxytrityl; TBS-Cl: *tert*-butyldimethylsilyl chloride; NBS: *N*-bromosuccinimide; AIBN: azobisisobutyronitrile; TPS-Cl: 2,4,6-triisopropylbenzenesulfonyl chloride; DMAP: 4-dimethylaminopyridine; CEP-Cl: 2-cyanoethyl-*N*,*N*-diisopropylchloro-phosphoramidite).[19]



Scheme 2. Synthesis of f^{5}C phosphoramidite **2**.[19]

Selective deprotection of 3', 5'-*O*-di-*tert*-butylsilylene group was achieved by treating **18** with hydrogen fluoride in pyridine. Finally, 5'-OH DMT protection and subsequent phosphitylation yielded target monomer **2**. Starting from 5-methyluridine and following this ten step route we obtained monomer **2** in a good 10.5 % overall yield (Scheme 2).[19] Compared to the previously reported synthesis for an f^5C phosphoramidite, we have significantly improved the overall yield and more importantly, we have now deployed a formyl protective group.[10] We anticipated that this fully protected monomer would allow us to incorporate f^5C at multiple positions in longer ONs and also in the combination with the other cytosine-modifications m^5C, hm^5C and hm^5Cm.

Next, we synthesized the hm^5C phosphoramidite **3** (Scheme 3). During the synthesis of the previous two monomers, we noted that the methyl group of appropriately protected 5-methyluridine could be functionalized efficiently. Hence, we started our synthesis with 5-methyluridine **11** followed by protecting 5≠-hydroxyl with DMT (**21**) and the protection of the 2≠- and 3≠-hydroxyls with TBS. Conversion of the 5-methyl group of **22** into the 5-acetyloxymethyl group was the key step in this synthesis, which we achieved by bromination and subsequent acylation (Scheme 3).[19] From here onwards we followed the route reported by Riml et al. and obtained the monomer **3** in 8 steps with 6.4 % overall yield (Scheme 3).[15,19] During the preparation of this manuscript the Micura group improved their initial synthesis and reported an alternative method, starting from cytidine to obtained this monomer in 8 steps with 9.2% overall yield.[24]



Scheme 3. Synthesis of hm^5C phosphoramidite **3**.[19]

To evaluate the utility of our building blocks to obtain ONs containing various modifications we synthesised a small series of RNA ONs (**1–4**) (Figure 2A). To further confirm the robustness of these monomers we synthesised the RNA ONs **1–3** containing each single modification at three different positions, and RNA ON **4** contains all of four cytosine-modifications. To ascertain the compatibility of these monomers in the presence of a biochemical tag we also performed the RNA ON synthesis on CPG solid support tethered to TEG-Biotin. Modified phosphoramidites were incorporated into RNA ONs following a standard solid-phase RNA ON synthesis protocol. Incorporations of modified phosphoramidite substrates **1□3** were performed with a coupling time of 10 min, with 80□90% coupling efficiency, based on a trityl release assay. Oligonucleotides were cleaved from the solid support with 20 % Ethanol in NH_4OH solution. 2'-*O*-TBS deprotection was performed by 1:1 mixture of anhydrous DMSO and triethylamine trihydrofluoride. RNA ONs **1** and **3** were purified by HPLC at this stage, while RNA ONs **2** and **4** were each subjected to deprotection of the acetal moiety to reveal the formyl group.[19] Conditions previously used for the acetal deprotection of fdC in DNA led to only partial deprotection of acetal group for RNA ONs **2** and **4**, along with precipitation of RNA. On screening several deprotection conditions we found that treatment of RNA ON with 20% aqueous acetic acid results in complete and clean scission of acetal group without cleavage of the RNA (Figure S1). The purity and integrity of all modified RNA ONs was confirmed by LC-MS analysis (Figure 2 and Figure S1, S2).[19]

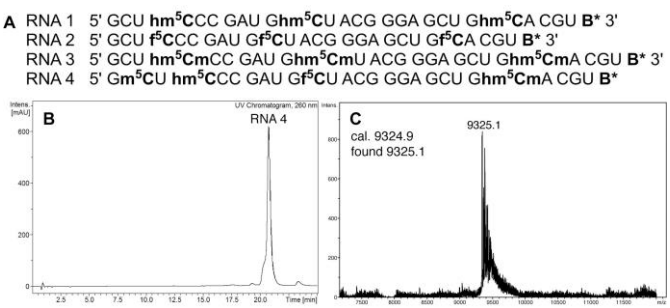


Figure 2. A) Sequences of modified RNA ONs synthesised in this study. RNA ONs **1**-**3** contains hm^5C, f^5C, hm^5Cm respectively at three different positions whereas, RNA ON **4** contains all four cytosine-modifications. The 3≠ end of each RNA ON is tagged with TEG-Biotin (B*). **B)** Representative LC trace of RNA ON **4**. **C**) Corresponding ESI-MS spectrum of RNA ON **4.**[19]

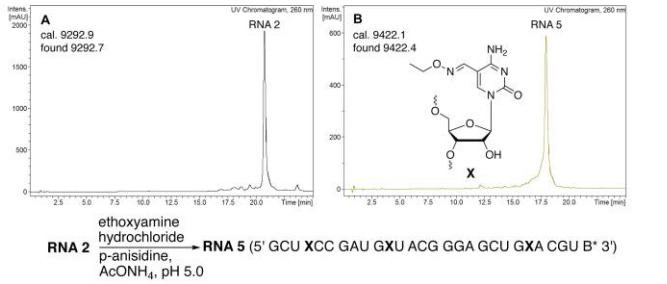


Figure 3. LC-MS chromatogram of RNA ON **2** before **(A)** and after **(B)** post-synthetic chemical reaction with ethoxyamine hydrochloride at pH 5.0.[24] Sequence of the RNA ON **5** is given at the bottom and structure of the modified nucleoside **(X)** formed upon selective condensation reaction between f^5C and ethoxyamine is shown in chromatogram **(B).**[19]

To further demonstrate the intactness of the modified ribonucleosides upon RNA synthesis and deprotection beyond mass spectroscopy, we perform a post-synthetic chemical functionalization of the RNA. Firstly, we subjected the RNA ON **2** containing a reactive formyl group for nucleophile addition with ethoxyamine hydrochloride at pH 5.0 in the presence of anisidine.[(24)] All three f^5C nucleotides present in RNA **2** each

undergo facile addition of ethoxyamine, followed by elimination of water to form a stable imine derivative, which we confirmed by LC-MS analysis (Figure 3).[19,24] Next, when we performed the same reaction with RNA ONs **1** and **3** containing three hm^5C and hm^5Cm we observed that both the RNA ONs remain unaltered by this reaction. This experiment supports the functional integrity of these modifications in RNA.

In summary, we have demonstrated the efficient syntheses of hm^5Cm, hm^5C, and f^5C building blocks and the synthesis of RNA ONs containing these modifications at multiple positions in excellent yield and purity. The availability of these monomers and the capacity to prepare RNA oligomers with any combination, in desired positions, will be vital for studying the function(s) of cytosine modifications in biology.

Acknowledgements

We acknowledge support from University of Cambridge and Cancer Research UK program. The Balasubramanian laboratory is supported by core funding from Cancer Research UK (C14303/A17197). S.B. is a Senior Investigator of the Wellcome Trust (grant no. 099232/z/12/z).

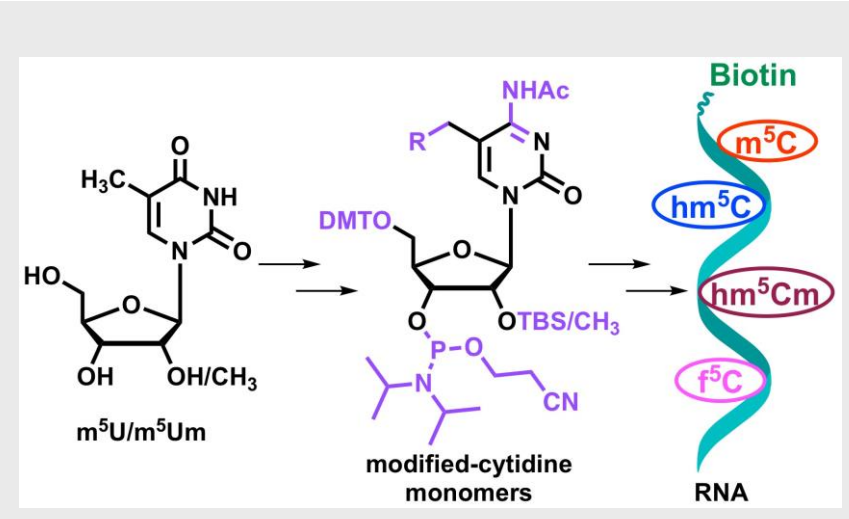
Keywords: RNA epigenetic • 5-methylcytosine • 5-hydroxymethylcytosine • 5-formylcytosine • 2≠-O-methyl-5-hydroxymethylcytidine

References

1. a) Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611−631; b) M. A. Machnicka, K. Milanowska, O. Osman Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262−D267; c) W-J. Sun, J-H. Li, S. Liu, J. Wu, H. Zhou, L-H. Qu, J-H. Yang, *Nucleic Acids Res.* **2016**, *44*, D259−D265.
2. a) M. Frye, S. R. Jaffrey, T. Pan, G. Rechavi, T. Suzuki, *Nat. Rev. Genet.* **2016**, *17*, 365−372; b) B. S. Zhao, I. A. Roundtree, C. He, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 31–42.
3. a) G. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindhal, T. Pan, Y. G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885−887; b) G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C. M. Huang, C. J. Li, C. B. Vågbø, Y. Shi, W. L. Wang, S. H. Song, Z. Lu, R. P. G. Bosmans, Q. Dai, Y. J. Hao, X. Yang, W-M. Zhao, W-M. Tong, X. J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y. G. Yang, C. He, *Mol. Cell.* **2013**, *49*, 18−29.
4. a) I. A. Roundtree, C. He, *Curr. Opin. Chem. Biol.* **2016**, *30*, 46–51; b) C. J. Lewis, T. Pan, A. Kalsotra, *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 202–210.
5. J. E. Squires, H. R. Patel, M. Nousch, T. Sibbritt, D. T. Humphreys, B. J. Parker, C. M. Suter, T. Preiss, *Nucleic Acids Res.* **2012**, *40*, 5023–5033.
6. a) L. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S. G. Jin, L. J. Niedernhofer, G. P. Pfeifer, G. L. Xu, Y. Wang, *J. Am. Chem. Soc.* **2014**, *136*, 11582−11585; b) S. M. Huber, P. van Delft, L. Mendil, M. Bachman, K. Smollett, F. Werner, E. A. Miska, S. Balasubramanian, *ChemBioChem* **2015**, *16*, 752−755.
7. a) S. Nakano, T. Suzuki, L. Kawarada, H. Iwata, K. Asano, T. Suzuki, *Nat. Chem. Biol.* **2016**, *12*, 546−551; b) L. Van Haute, S. Dietmann, L. Kremer, S. Hussain, S. F. Pearce, C. A. Powell, J. Rorbach, R. Lantaff, S. Blanco, S. Sauer, U. Kotzaeridou, G. F. Hoffmann, Y. Memari, A. Kolb-Kokocinski, R. Durbin, J. A. Mayr, M. Frye, H. Prokisch, M. Minczuk, *Nat Commun.* **2016**, *7*, 12039, DOI: 10.1038/ncomms12039; c) S. Haag, K. E. Sloan, N. Ranjan, A. S. Warda, J. Kretschmer, C. Blessing, B. Hübner, J. Seikowski, S. Dennerlein, P. Rehling, M. V. Rodnina, C. Höbartner, M. T. Bohnsack, *EMBO J.* **2016**, *35*, 2104−2119.
8. X. Yang, Y. Yang, B-F. Sun, Y-S. Chen, J-W. Xu, W-Y. Lai, A. Li, X. Wang, D. P. Bhattarai, W. Xiao, H-Y. Sun, Q. Zhu, H-L. Ma, S. Adhikari, M. Sun, Y-J. Hao, B. Zhang, C-M. Huang, N. Huang, G-B. Jiang, Y-L. Zhao, H-L. Wang, Y-P. Sun, Y-G. Yang, *Cell Res.* **2017**, *27*, 606–625.
9. B. Delatte, F. Wang, L. V. Ngoc, E. Collignon, E. Bonvin, R. Deplus, E. Calonne, B. Hassabi, P. Putmans, S. Awe, C. Wetzel, J. Kreher, R. Soin, C. Creppe, P. A. Limbach, C. Gueydan, V. Kruys, A. Brehm, S. Minakhina, M. Defrance, R. Steward, F. Fuks, *Science* **2016**, *351*, 282−285.
10. H. Lusic, E. M. Gustilo, F. A. P. Vendeix, R. Kaiser, M. O. Delaney, W. D. Graham, V. A. Moye, W. A. Cantara, P. F. Agris, A. Deiters, *Nucleic Acids Res.* **2008**, *36*, 6548–6557.
11. W. Huang, M-D. Lan, C-B. Qi, S-J. Zheng, S-Z. Wei, B-F. Yuan, Y-Q. Feng, *Chem. Sci.* **2016**, *7*, 5495–5502.
12. S. M. Huber, P. van Delft, A. Tanpure, E. A. Miska, S. Balasubramanian, *J. Am. Chem. Soc.* **2017**, *139*, 1766−1769.
13. L. Kawarada, T. Suzuki, T. Ohira, S. Hirata, K. Miyauchi, T. Suzuki, *Nucleic Acids Res.* **2017**, *45*, 7401−7415.
14. a) M. Schaefer, T. Pollex, K. Hanna, F. Lyko, *Nucleic Acids Res.* **2009**, *37*, e12; b) M. J. Booth, M. R. Branco G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934−937; c) M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435−440.
15. C. Riml, R. Micura, *Synthesis* **2016,** *48*, A–I.
16. S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593−14599.
17. D. Gavriliu, C. Fossey, G. Fontaine, S. Benzaria, A. Ciurea, Z. Delbederi, B. Lelong, D. Laduree, A. M. Aubertin, A. Kirn, *Nucleosides, Nucleotides & Nucleic Acids* **2000**, *19*, 1017–1031.
18. a) F. Wachowius, C. Höbartner, *Chembiochem*, **2010**, *11*, 469–480; b) B. Samanta, J. Seikowski, C. Höbartner, *Angew. Chem. Int. Ed.* **2016**, *55*, 1912–1916.
19. See supporting information for details.
20. a) R. K. Grover, S. J. K. Pond, Q.-Z. Cui, P. Subramaniam, D. A. Case, D. P. Millar, P. Jr. Wentworth, *Angew. Chem. Int. Ed.* **2007**, *46*, 2839−2843; b) Q. Sun, J. Sun, S-S. Gong, C-J. Wang, S-Z. Pua, F-D. Feng, *RSC Adv.* **2014**, *4*, 36036–36039.
21. A. S. Schrcöder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder, T. Carell, *Angew. Chem. Int. Ed.* **2014**, *53*, 315–318.
22. V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983–1985.
23. D. B. Dess, J. C. Martin, *J. Org. Chem.* **1983**, *48*, 4155–4156.
24. C. Riml, A. Lusser, E. Ennifar, R. Micura, *J. Org. Chem.* **2017**, *DOI: 10.1021/acs.joc.7b01171*
25. E.-A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.

Entry for the Table of Contents (Please choose one layout)

Layout 2:



*Arun A. Tanpure and Shankar Balasubramanian**

Page No. – Page No.

Synthesis and multiple incorporations of 2≠-*O*-methyl-5-hydroxymethyl-, 5-hydroxymethyl- and 5-formylcytidine monomers into RNA oligonucleotides

Synthesis and multiple incorporations of 2≠-*O*-methyl-5-hydroxymethylcytidine (hm^5Cm), 5-hydroxymethylcytidine (hm^5C) and 5-formylcytidine (f^5C) phosphoramidite monomers into RNA ONs have been developed. Optimization of mild post-synthetic deprotection conditions enabled the preparation of RNA oligomers containing all four of the cytosine-modifications at multiple positions and in combination.