ORGANIC LETTERS

2013
Vol. 15, No. 2
366–369

Synthesis of 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxycytidine-triphosphates and Their Incorporation into Oligonucleotides by Polymerase Chain Reaction

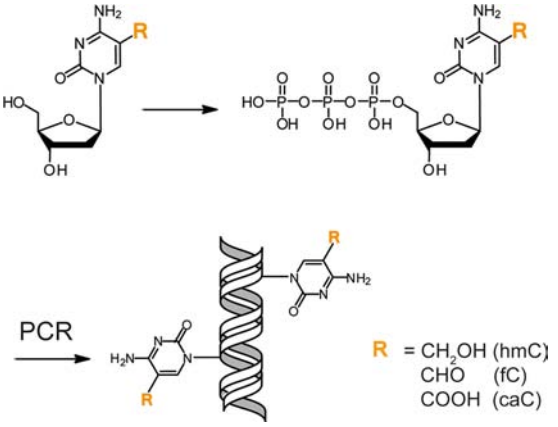
Barbara Steigenberger,[†] Stefan Schiesser,[†] Benjamin Hackner, Caterina Brandmayr, Silvia K. Laube, Jessica Steinbacher, Toni Pfaffeneder, and Thomas Carell*

Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians Universität München, Butenandtstr. 5-13, 81377, Munich, Germany

Thomas.Carell@lmu.de

Received December 4, 2012

ABSTRACT



The synthesis of the triphosphates of 5-hydroxymethyl-, 5-formyl-, and 5-carboxycytidine and the incorporation of these building blocks into long DNA fragments using the polymerase chain reaction (PCR) are reported. In this way DNA fragments containing multiple hmC, fC, and caC nucleobases are readily accessible.

Recently three new nucleobases were discovered in DNA isolated from mouse embryonic stem cells.[1−4] The three nucleobases are oxidation products of 5-methylcytosine (mC), which is a base that regulates transcriptional activity.[5] The oxidation reaction of 5-methylcytosine to 5-hydroxymethylcytosine (hmC), 5-formylcytosine (fC), and finally 5-carboxycytosine (caC) is now understood to be performed by 10-11-translocon proteins (TET1−3).[6] These enzymes are α-ketoglutarate dependent oxidases, which directly utilize molecular oxygen for the oxidation reaction.[7] The three new nucleobases are currently thought to be involved in epigenetic programming of cells, and they could be intermediates of a long searched for pathway of active demethylation.[8−10] Recently new sequencing methods that allow the genome wide localization of hmC[11,12] and fC[13] in genomic DNA were reported.

[†] These authors contributed equally.

(1) Kriaucionis, S.; Heintz, N. *Science* **2009**, *324*, 929.
(2) Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A. *Science* **2009**, *324*, 930.
(3) Pfaffeneder, T.; Hackner, B.; Truss, M.; Münzel, M.; Müller, M.; Deiml, C. A.; Hagemeier, C.; Carell, T. *Angew. Chem., Int. Ed.* **2011**, *50*, 7008.
(4) He, Y.-F.; Li, B.-Z.; Li, Z.; Liu, P.; Wang, Y.; Tang, Q.; Ding, J.; Jia, Y.; Chen, Z.; Li, L.; Sun, Y.; Li, X.; Dai, Q.; Song, C.-X.; Zhang, K.; He, C.; Xu, G.-L. *Science* **2011**, *333*, 1303.
(5) Law, J. A.; Jacobsen, S. E. *Nat. Rev. Genet.* **2010**, *11*, 204.
(6) Ito, S.; Shen, L.; Dai, Q.; Wu, S. C.; Collins, L. B.; Swenberg, J. A.; He, C.; Zhang, Y. *Science* **2011**, *333*, 1300.
(7) Loenarz, C.; Schofield, C. J. *Chem. Biol.* **2009**, *16*, 580.
(8) Jurkowski, T. P.; Jeltsch, A. *ChemBioChem* **2011**, *12*, 2543.
(9) Gu, T.-P.; Guo, F.; Yang, H.; Wu, H.-P.; Xu, G.-F.; Liu, W.; Xie, Z.-G.; Shi, L.; He, X.; Jin, S.-g.; Iqbal, K.; Shi, Y. G.; Deng, Z.; Szabo, P. E.; Pfeifer, G. P.; Li, J.; Xu, G.-L. *Nature* **2011**, *477*, 606.
(10) Ladwein, K. I.; Jung, M. *Angew. Chem., Int. Ed.* **2011**, *50*, 12143.

10.1021/ol3033219
Published on Web 01/03/2013

In the past couple of years phosphoramidite building blocks of hmC, fC, and caC, which allow the solid phase synthesis of oligonucleotides containing the new bases at defined sites, were developed.[14−18] However, to search for proteins that interact with these new epigenetic bases, and to decipher the biological/biochemical questions associated with the new nucleobases, longer oligonucleotides containing multiple hmC, fC, and caC bases are required. In this direction, we thought that using the corresponding triphosphates (Figure 1) in combination with the polymerase chain reaction (PCR) would solve this chemical problem.[19−21]

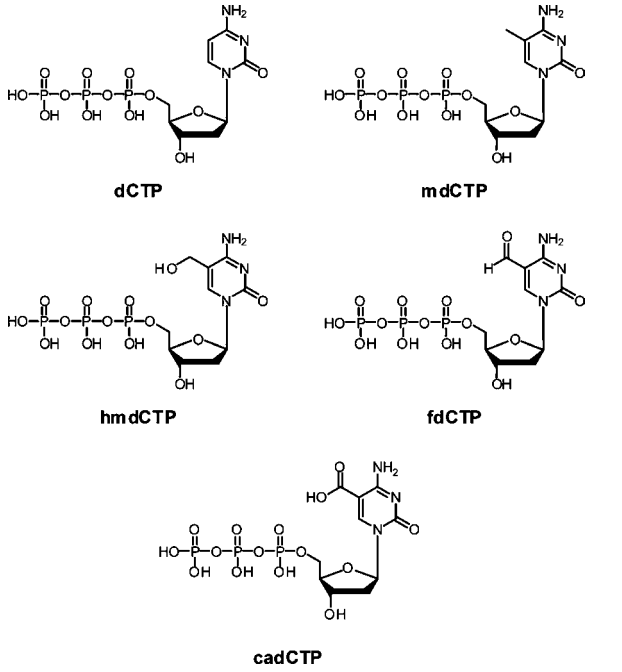
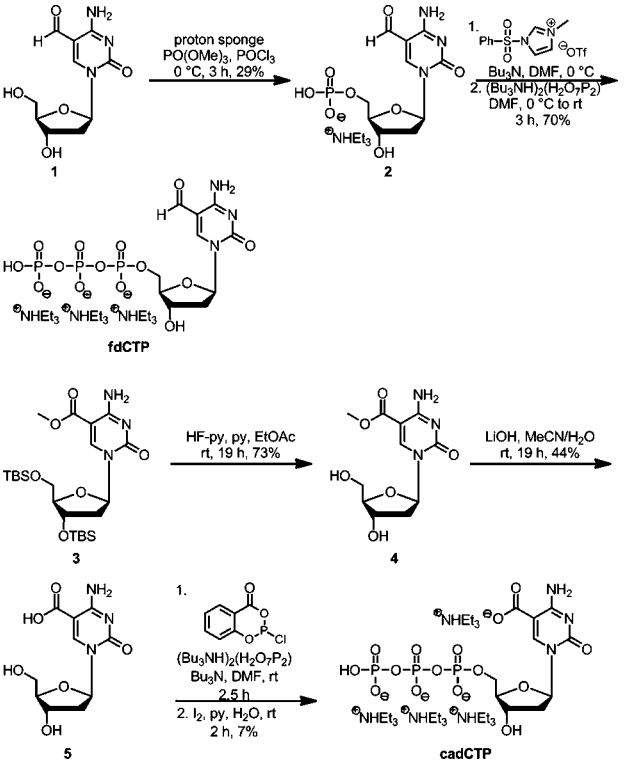


Figure 1. Depiction of the canonical DNA base dC and of mC as well as of the new epigenetic bases hmC, fC, and caC as triphosphates.

In such a PCR, one would exchange the dCTP either completely or partially by the corresponding xdCTP (x = hm, f or ca) so that these DNA fragments contain the new bases hmC, fC, and caC at the corresponding dC positions.

While the triphosphate of hmC (hmdCTP) is already commercially available and its incorporation via PCR is established, the corresponding fdCTP and cadCTP compounds were unknown at the beginning of this study. The chemical synthesis of both building blocks was achieved as outlined in Scheme 1. For fdCTP we started the synthesis with the fC nucleoside **1**, which was prepared as described recently by us.[22] This compound was converted into the 5′-monophosphate **2**, which was obtained after HPLC purification in 29% yield. We subsequently employed the new triphosphate method recently described by S. D. Taylor et al. using sulfonyl imidazolium triflate as the activating reagent and pyrophosphate.[23] The procedure allowed us to access the triphosphate from the monophosphate in 70% yield. The triphosphate was best isolated by ion exchange chromatography at 4 °C using a DEAE-cellulose column[24] with a gradient from 100% water to 0.5 M TEAB (pH 7.5). The crude triphosphate product was further purified by FPLC (0.1 M TEAB, 1 M TEAB; 0−100% over 30 min) using a MonoQ 5/50 GL anion exchange column (GE). This two-step procedure allowed us to generate the reactive aldehyde-containing fC triphosphate in sufficient yield for all further studies.

Scheme 1. Synthesis of fdCTP and cadCTP



(11) Booth, M. J.; Branco, M. R.; Ficz, G.; Oxley, D.; Krueger, F.; Reik, W.; Balasubramanian, S. *Science* **2012**, *336*, 934.
(12) Yu, M.; Hon, G. C.; Szulwach, K. E.; Song, C.-X.; Zhang, L.; Kim, A.; Li, X.; Dai, Q.; Shen, Y.; Park, B.; Min, J.-H.; Jin, P.; Ren, B.; He, C. *Cell* **2012**, *149*, 1368.
(13) Raiber, E.-A.; Beraldi, D.; Ficz, G.; Burgess, H. E.; Branco, M. R.; Murat, P.; Oxley, D.; Booth, M. J.; Reik, W.; Balasubramanian, S. *Genome Biol.* **2012**, *13*, R69.
(14) Tardy-Planechaud, S.; Fujimoto, J.; Lin, S. S.; Sowers, L. C. *Nucleic Acids Res.* **1997**, *25*, 553.
(15) Münzel, M.; Globisch, D.; Trindler, C.; Carell, T. *Org. Lett.* **2010**, *12*, 5671.
(16) Dai, Q.; He, C. *Org. Lett.* **2011**, *13*, 3446.
(17) Münzel, M.; Lischke, U.; Stathis, D.; Pfaffeneder, T.; Gnerlich, F. A.; Deiml, C. A.; Koch, S. C.; Karaghiosoff, K.; Carell, T. *Chem.—Eur. J.* **2011**, *17*, 13782.
(18) Dai, Q.; He, C. *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: 2001.
(19) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. *J. Am. Chem. Soc.* **2005**, *127*, 15071.
(20) Kuwahara, M.; Nagahima, J.; Hasegawa, M.; Tamura, T.; Kitagata, R.; Hanawa, K.; Hososhima, S.; Kasamatsu, T.; Ozaki, H.; Sawai, H. *Nucleic Acids Res.* **2006**, *34*, 5383.
(21) Shoji, A.; Hasegawa, M.; Hososhima, S.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 776.
(22) Globisch, D.; Münzel, M.; Müller, M.; Michalakis, S.; Wagner, M.; Koch, S.; Brückl, T.; Biel, M.; Carell, T. *PLoS ONE* **2010**, *5*, e15367.
(23) Mohamady, S.; Desoky, A.; Taylor, S. D. *Org. Lett.* **2011**, *14*, 402.

The cadCTP compound was prepared from the TBS protected caC methyl ester **3** which was also prepared as described recently by us.[17] We first cleaved the TBS groups to obtain compound **4**. The methyl ester was subsequently saponified which provided the unprotected caC nucleoside **5**. **5** was next introduced into the one-step triphosphate synthesis reported originally by Eckstein et al.[25] Here the yield could be improved to 7% if the conditions developed by Huang et al. were employed.[26]

We recently reported the synthesis of hmC, fC, and caC phosphoramidites and the incorporation of these building blocks into DNA strands. Primer extension studies showed that none of the new bases are mutagenic.[17] Here we report the development of PCR conditions for the incorporation of fdCTP and cadCTP into long oligonucleotides. We chose the oct4 promotor sequence (see Supporting Information (SI)) as the DNA template. The primers for the PCR were designed to yield a 150 bp product containing 77 modified dCs (4 dC are present in the primer; these are not exchanged). For this purpose the forward and reverse primers were annealed to the template at 55 °C. The elongation of the primers was best performed at 75 °C (for fdCTP) and 72 °C (for cadCTP). Different polymerases were screened. We discovered that the DNA polymerase Vent (exo$^-$) (for fdCTP) and KOD XL polymerase for (cadCTP) provided the best results (see SI). To ensure complete extension of the primer the elongation time was lengthened compared to the time used for incorporation of dCTP. The experimental results of the PCRs are described in Figure 2.

The PCR products obtained with cadCTP can only be visualized when a 1 × TBE buffer system is used for the analysis. When other buffers such as TAE were used we noted that the obtained oligonucleotide products did not give a distinct band in the gel electrophoresis potentially because of the additional carboxylic acid groups present on caC. Rather a broad smear is detected due to the lower buffer capacity. The TBE buffer system in contrast provides sharp bands for the caC containing DNA products. As depicted in lanes 5 and 6, both triphosphates fdCTP (lane 5) and cadCTP (lane 6) yielded PCR products with the correct length (negative control lanes 1 and 2) if our developed PCR methods are employed. In the shown experiments we replaced the dCTP completely by the corresponding xdCTP. Hence full length PCR product can only be formed when the triphosphate is accepted, as further shown by the negative control in lane 2. Further proof for the correct incorporation of fC and caC using PCR was obtained by LC-MS experiments. The PCR products were to this end fully digested. For this purpose the sugar phosphate backbone was first cleaved with nuclease S1 and snake venom phosphodiesterase, giving the 5′-monophosphates. These were further hydrolyzed to the nucleoside level by Antarctic phosphatase.[3] Using this

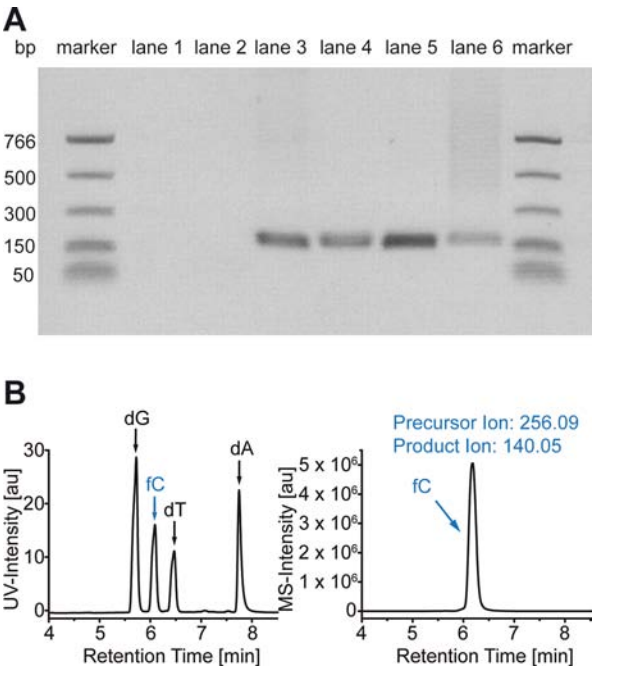


Figure 2. (A) Depiction of the results of the PCR analyzed by gel electrophoresis. Lane 1: without template. Lane 2: without any dC derivative. Lane 3: dCTP. Lane 4: protected D_2,$^{15}N_2$-hmdCTP. Lane 5: fdCTP. Lane 6: cadCTP. (B) UV trace (left) and mass trace of fC (right) of a fully digested PCR mixture with fdCTP.

procedure even the highly modified DNA prepared here was fully digested. The resulting nucleoside mixture was subsequently analyzed by LC-HRMS or LC-MS/MS. The data obtained for incorporated fdCTP are shown in Figure 2B (for cadCTP, see SI). Clearly evident is the presence of dA, dT, and dG in addition to the fC-nucleoside. Our results show that both fC and caC can be inserted as triphosphates into long DNA fragments using PCR. The observation that formyl group containing nucleosides can be incorporated into PCR products despite their high reactivity is in line with a recent report by Hocek et al. This group reported the PCR based synthesis of aldehyde containing DNA products.[27,28] It is interesting that the polymerase tolerates also the negative charge associated with the carboxylic acid present in caC. For future quantification of fC and caC in natural material we also prepared isotope labeled fC and caC triphosphates and incorporated them into DNA strands using basically the same PCR conditions showing the broad applicability of the here reported technology (see SI).

We next turned our attention to the hmC base. Here, the corresponding triphosphate is generated from the monophosphate, which is directly isolated from natural sources. A chemical synthesis of the hmC triphosphate was not

(24) Guan, L.; van der Heijden, G. W.; Bortvin, A.; Greenberg, M. M. *ChemBioChem* **2012**, *12*, 2184.
(25) Ludwig, J.; Eckstein, F. *J. Org. Chem.* **1989**, *54*, 631.
(26) Caton-Williams, J.; Smith, M.; Carrasco, N.; Huang, Z. *Org. Lett.* **2011**, *13*, 4156.
(27) Raindlová, V.; Pohl, R.; Šanda, M.; Hocek, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1064.
(28) Raindlová, V.; Pohl, R.; Hocek, M. *Chem.—Eur. J.* **2012**, *18*, 4080.

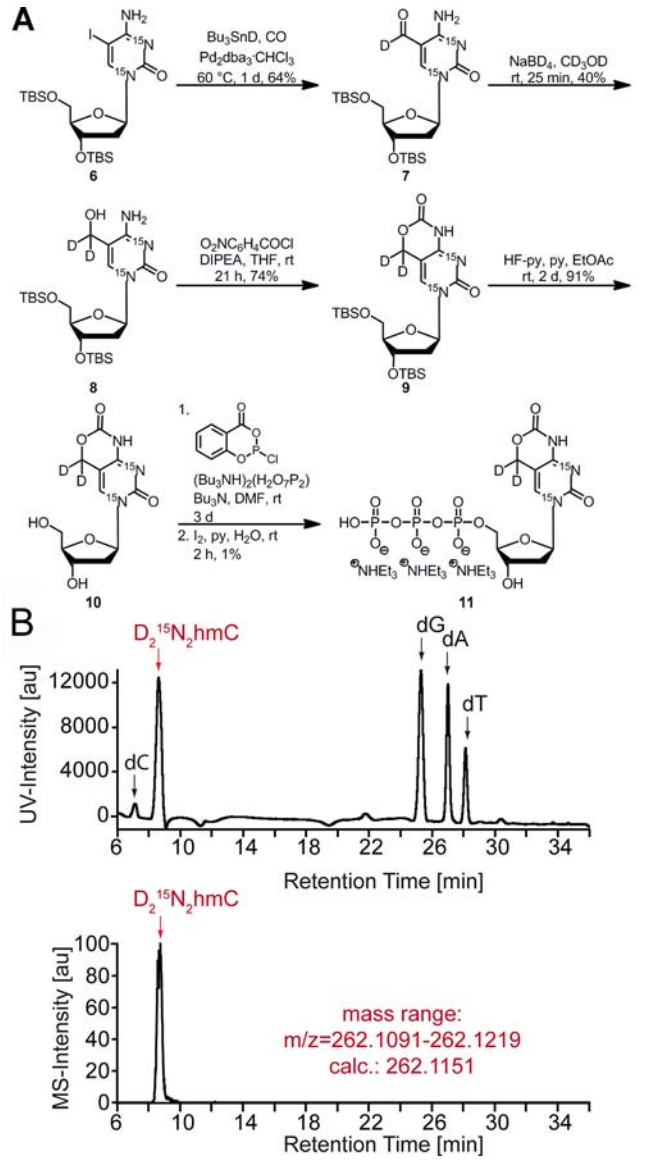


Figure 3. (A) Synthesis of the protected D_2,$^{15}N_2$-hmdCTP. (B) UV trace (top) and D_2,$^{15}N_2$-hmC mass trace (bottom) of a fully digested PCR mixture.

performed so far, which limits our ability to create DNA fragments with modified hmC, e.g. with isotopically labeled compounds as needed for mass spectrometry based quantification and proteomics studies. The synthetic challenge associated with the synthesis of hmdCTP is the benzylic hydroxyl group of the hmC heterocycle which is more reactive than the primary 5′-OH group. This makes it difficult to access the 5′-monophosphate directly from hmC. To circumvent this problem and to enable the PCR incorporation of modified hmC building blocks, we investigated the ability to insert a protected hmC derivative by PCR followed by deprotection. The synthesis of the protected hmdCTP building block and the results of the PCR study are shown in Figure 3.

To exemplify the possibilities associated with the chemical method, we prepared for this study the unnatural 2 × [^{15}N] and 2 × D modified hmC (SI). The starting point for the synthesis is the TBS protected iodouracil **6**,[29] which was carbonylated to **7** and reduced to **8**. Compound **8** was protected as the carbamate **9**, and the TBS groups were cleaved to obtain **10**. We utilized the optimized one-pot triphosphate synthesis described above to obtain the labeled hmdCTP **11**. Subsequent PCR based incorporation studies showed that the reaction is best performed with the KOD XL polymerase with again slightly prolonged elongation times (30 s instead of 15 s). The PCR product was subsequently deprotected with 0.1 M NaOH in water/methanol 1:4 for 1 h at rt. The DNA was finally purified using a silica membrane (see SI). The agarose gel of the deprotected and purified DNA fragment is depicted in Figure 2A, lane 4. Again a clean PCR product is observed. The results of the total digest performed under the optimized conditions reported above are shown in Figure 3B. Again next to dA, dG, and dT as well as small amounts of residual dC from the primers, an additional signal is observed with the correct retention time and exact molecular weight for D_2,$^{15}N_2$-hmC. Most importantly we do not observe a signal for un-deprotected hmC showing that full deprotection of the carbamate protecting group present on our hmdCTP building block was achieved. In summary we report here the first chemical synthesis of the three triphosphates hmdCTP, fdCTP, and cadCTP of the new epigenetic bases hmC, fC, and caC and describe PCR conditions which enable the incorporation of these building blocks into long DNA fragments. The chemical synthesis allows even the synthesis and incorporation of isotopologues of the new bases, which should facilitate mass spectrometry based quantification methods.

We believe that the reported synthetic methodologies disclosed here will strongly advance our ability to study the biology and biochemistry of the new epigenetic bases hmC, fC and caC.

Acknowledgment. We thank the excellence cluster EXC114 and the SFB749 as well as the Volkswagen foundation and the DFG for Grant CA275/8-4 for financial support. S.S. and T.P. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for predoctoral fellowship.

Supporting Information Available. Experimental procedures and spectroscopic data of all new compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

(29) Schiesser, S.; Hackner, B.; Pfaffeneder, T.; Müller, M.; Hagemeier, C.; Truss, M.; Carell, T. *Angew. Chem., Int. Ed.* **2012**, *51*, 6516.

The authors declare no competing financial interest.