Epigenetics Hot Paper

DOI: 10.1002/anie.201308469

Synthesis of a DNA Promoter Segment Containing All Four Epigenetic Nucleosides: 5-Methyl-, 5-Hydroxymethyl-, 5-Formyl-, and 5-Carboxy-2′-Deoxycytidine**

*Arne S. Schröder, Jessica Steinbacher, Barbara Steigenberger, Felix A. Gnerlich, Stefan Schiesser, Toni Pfaffeneder, and Thomas Carell**

Abstract: *A 5-formyl-2′-deoxycytidine (fdC) phosphoramidite building block that enables the synthesis of fdC-containing DNA with excellent purity and yield has been developed. In combination with phosphoramidites for 5-methyl-dC, 5-hydroxymethyl-dC, and carboxy-dC, it was possible to prepare a segment of the OCT-4 promoter that contains all four epigenetic bases. Because of the enormous interest in these new epigenetic bases, the ability to insert all four of them into DNA should be of great value for the scientific community.*

Aside from the four canonical nucleosides, the genome contains several DNA modifications that are of epigenetic importance. The nucleoside 2′-deoxycytidine is often methylated at the C5 position to give 5-methyl-2′-deoxycytidine (**1**, mdC), which is involved in epigenetic gene silencing.[1] In recent years, the oxidation products of mdC were discovered.[2–5] Ten-eleven-translocation proteins (TET1–3) are now known to oxidize mdC to 5-hydroxymethyl- (**2**, hmdC), 5-formyl- (**3**, fdC), and 5-carboxy-2′-deoxycytidine (**4**, cadC).[6,7] These "new" modifications (Figure 1 A) are primarily found in the CpG context in DNA promoter sequences, where they seem to be involved in regulating gene expression.[8] Lately, the synthesis and analysis of a tRNA segment of human mitochondrial $tRNA^{Met}$ that carries 5-formylcytidine in the anticodon loop was reported.[9]

To study the function of these new nucleosides, it is of utmost importance to develop methods that enable the synthesis of DNA strands that contain hmdC, fdC, and cadC. Recently, we reported the enzymatic incorporation of these nucleosides into DNA by means of their triphosphates.[10] However, a sequence-specific incorporation of the nucleosides, particularly into CpG islands, which would be required for proteomics studies, is not possible with this method.

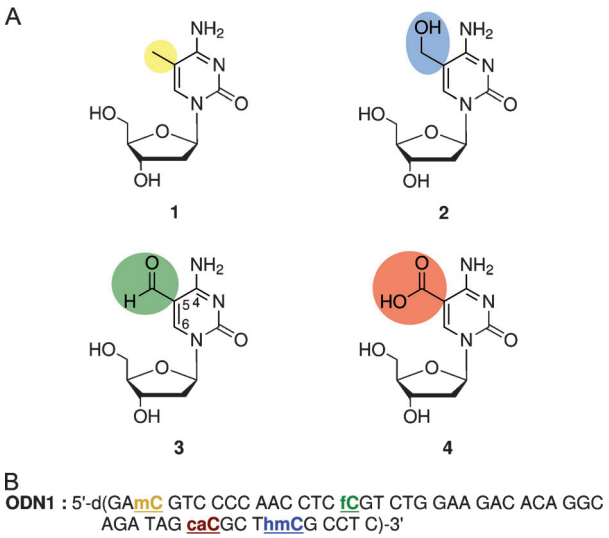


Figure 1. A) The recently found epigenetically relevant nucleosides 5-methyl- (**1**, mdC), 5-hydroxymethyl- (**2**, hmdC), 5-formyl- (**3**, fdC), and 5-carboxy-2′-deoxycytidine (**4**, cadC). B) Sequence of an OCT-4 promoter segment, in which cytidines in CpG contexts were substituted with the epigenetic nucleosides **1**–**4**.

The "Achilles' heel" of the synthesis of DNA strands containing all new epigenetic nucleosides is the incorporation of fdC because of the high reactivity of the formyl group.[11–13] The only currently available fdC building block is inserted as the 1,2-diol precursor, and periodate treatment is required to obtain the formyl moiety;[14] this reagent is, for instance, incompatible with hmdC.[15] Moreover, the periodate-based deprotection promotes side reactions that prevent the synthesis of DNA strands that contain multiple fdCs. Consequently, DNA strands that contain the four epigenetic nucleosides mdC, hmdC, fdC, and cadC at defined sites are synthetically not accessible.[16]

Herein, we report the development of a new fdC phosphoramidite building block that can be inserted multiple times, or even consecutively, into DNA strands by using standard phosphoramidite chemistry. High coupling yields during solid-phase synthesis as well as mild deprotection conditions even enabled the preparation of a OCT-4 promoter sequence that contains all four epigenetic nucleosides (Figure 1 B, **ODN1**).

The electron-poor heterocycle of **3** is a major chemical obstacle for fdC incorporation, as this moiety promotes glycosidic bond cleavage even under the mildly acidic

[*] A. S. Schröder, J. Steinbacher, B. Steigenberger, F. A. Gnerlich, S. Schiesser, T. Pfaffeneder, Prof. Dr. T. Carell
Center for Integrated Protein Science
Department für Chemie, Ludwig-Maximilians-Universität
Butenandtstrasse 5–13, 81377 München (Germany)
E-mail: Thomas.Carell@lmu.de
Homepage: http://www.carellgroup.de

[**] We thank the Deutsche Forschungsgemeinschaft (SFB749 and SFB646) for financial support. Additional support from the Fonds der Chemischen Industrie (predoctoral fellowships for A.S.S, S.S., T.P.) and two excellence clusters (CiPS[M], EXC114) is acknowledged.

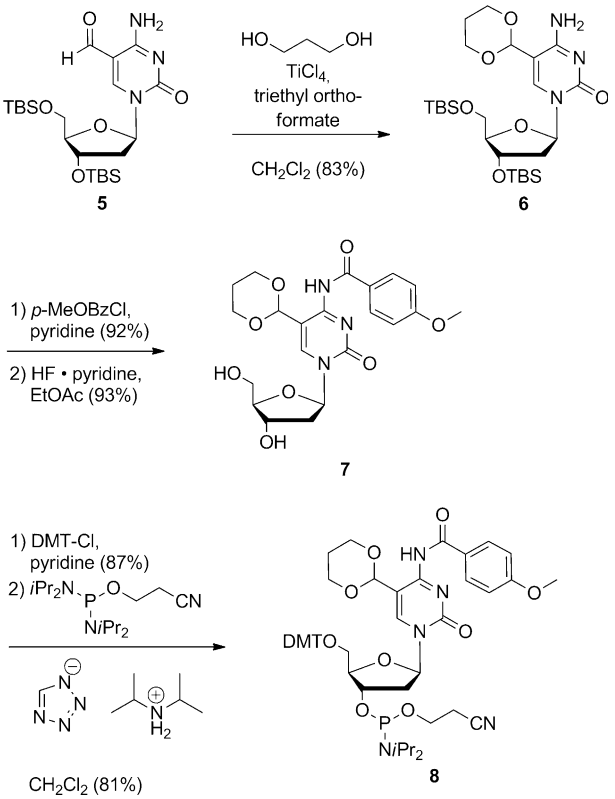
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201308469.

conditions that are commonly used for oligonucleotide synthesis. Furthermore, the formyl group of **3** is sensitive to oxidation, and, most importantly, 5-formylcytosine is readily attacked by nucleophiles, particularly at the C6 position (Figure 1 A).[16] To mask the reactivity of the aldehyde functionality, we examined various protecting groups (for details, see the Supporting Information). Finally, the best results were obtained when the aldehyde is converted into a 1,3-dioxane unit with propane-1,3-diol. This acetal turned out to be sensitive enough for later cleavage under mild conditions (80% aqueous acetic acid, 20°C); nevertheless, it is sufficiently stable to remain intact during solid-phase DNA synthesis. However, the temperature at which the deprotection is carried out at the end of the synthesis is a crucial parameter. Cleavage at 15°C is too slow, whereas higher temperatures ($\geq$25°C) lead to a significant amount of glycosidic bond cleavage, mostly at fdC (for details, see the Supporting Information). Therefore, deprotection of the fdC acetal has to occur at approximately 20°C. The amino group at the C4 position was protected with 4-methoxybenzoyl chloride, which is compatible with the acetal moiety. For fdC, we discovered that the commonly used benzoyl group is already efficiently cleaved during the solid-phase DNA synthesis, which gives rise to branched DNA products.

The synthesis of the fdC building block was achieved starting from *tert*-butyldimethylsilyl (TBS)-protected fdC **5**, which was synthesized from 5-iodo-dC according to literature procedures (Scheme 1).[17] Notably, standard procedures for acetal protection did not provide the desired product **6**. Activation of the formyl group is strictly required, and $TiCl_4$ turned out to be the only Lewis acid that gave acceptable yields. Subsequent protection of the amino group at the C4 position with 4-methoxybenzoyl chloride and silyl deprotection furnished the fdC derivative **7**, which was converted into the fdC phosphoramidite building block **8** by standard procedures.[18]

To examine the quality of the new fdC building block **8**, we prepared oligonucleotide **ODN2**, which contains five fdCs (Figure 2 B). The solid-phase synthesis was performed using standard phosphoramidite conditions.[19,20] For fdC, the coupling times were increased from 30 to 60 seconds to ensure maximum coupling yields. For deprotection of the standard nucleosides and cleavage of the oligonucleotide from the solid support, we first treated the solid-phase material with saturated aqueous NH_4OH solution (24 h, 25°C). The HPL chromatogram depicted in Figure 2 B shows the result for the crude product of the synthesis of the dimethoxytrityl (DMT)-protected oligonucleotide. In a second deprotection step, the oligonucleotide was treated with acetic acid for six hours at 20°C, which cleaved the acetal and DMT protecting groups. The HPL chromatogram that was obtained for the unpurified DNA strand **ODN2**, which contains five fdCs, is shown in Figure 2 C. As only one peak was observed, the synthesis had occurred in high yield.

To verify that five fdC nucleosides are indeed present in the DNA strand, we measured a MALDI-TOF spectrum (Figure 2 D), which showed the expected exact mass. Further proof for correct incorporation was obtained by digestion of the synthesized, but still crude oligonucleotide with nucle-



Scheme 1. Synthesis of acetal-protected fdC phosphoramidite **8**.

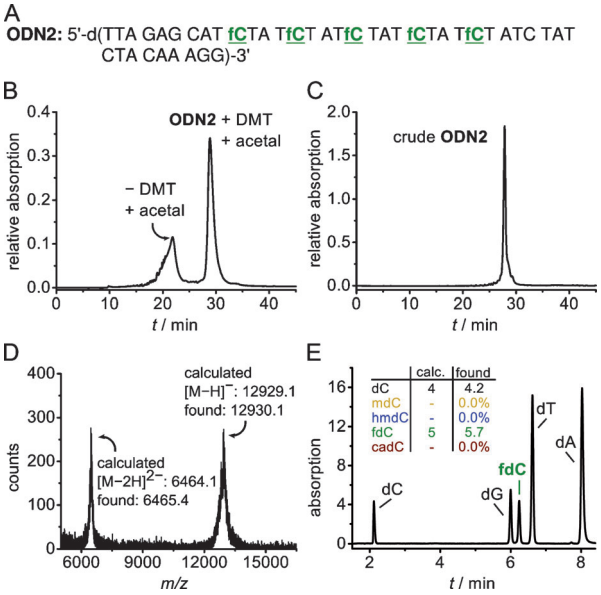


Figure 2. A) Sequence of the synthesized **ODN2** with fivefold incorporation of fdC using **8**. B) Reversed-phase HPL chromatogram of crude **ODN2** directly after basic cleavage from the resin (NH_4OH, 23 h, 25°C, 0–80% buffer B in 45 min). C) Reversed-phase HPL chromatogram of crude **ODN2** directly after acidic cleavage of the DMT and acetal groups (80% aq. acetic acid, 6 h, 20°C, 0–40% buffer B in 45 min). D) MALDI-TOF spectrum of fully deprotected, but crude **ODN2**. E) UHPL chromatogram and quantification data of unpurified digested **ODN2**.

ase S1 and snake-venom phosphodiesterase I to give the 5′-monophosphates. These were further hydrolyzed to the nucleosides by antarctic phosphatase. The obtained nucleoside mixture was analyzed by UHPLC-MS/MS (QQQ; Figure 2E). The obtained UV chromatogram was clean and showed signals only for the expected nucleosides. The fdC nucleoside was identified based on the retention time and its characteristic MS/MS fragmentation ($m/z = 256.1 \rightarrow 140.1$). Next, we quantified the amount of dC, fdC, and dT (Figure 2E, inset) by using synthetic isotopologues of the nucleosides as internal standards, which is a method that was recently described.[16] We obtained the expected dC/fdC ratio of approximately 4:5, which is in full agreement with the sequence. The fact that these values are obtained with the crude synthesis product underpins the superior properties of the new fdC phosphoramidite. Most importantly, we did not detect any α-fdC nucleosides, which indicates that β- to α-anomerization was fully suppressed with the new building block. We also investigated whether hmdC and cadC are present during DNA synthesis; these are the reduction and oxidation products of fdC, which might be formed during the iodine oxidation step. As expected, none of these compounds were detected (Figure 2E, inset). In summary, the fdC phosphoramidite **8** enabled the synthesis of fdC-containing oligonucleotides. With this building block in hand, we have established the basis for the synthesis of oligonucleotides that contain all four epigenetic nucleosides.

Therefore, we then synthesized oligonucleotide **ODN1** with the sequence of a segment of the OCT-4 promoter (*mus musculus*, chromosome 17, 35505895-35505943; Figure 3B). All cytidines in the respective CpG contexts were substituted with the epigenetic nucleosides **1**–**4**. The best results were obtained with the cyclic-carbamate-protected hmdC phosphoramidite (for detailed conditions of the oligonucleotide synthesis, see the Supporting Information).[17] This carbamate building block and the ester-protected cadC unit forced us to use NaOH (0.4 M in methanol/water (4:1)) instead of NH_4OH in the first deprotection step to avoid the formation of amide moieties (Figure 3A).[12,17] After treatment of the synthesized **ODN1** with NaOH, the oligonucleotide was precipitated with ethanol. We next performed the second deprotection step with acetic acid. Figure 3C shows the HPL chromatogram of the crude material after the second deprotection step. It is clearly visible that the process yielded **ODN1** as the major product, despite the fact that **ODN1** is a 49 mer oligonucleotide that contains four modified nucleosides (for detailed data of crude **ODN1**, see the Supporting Information). Purification by HPLC after the first deprotection step and subsequent acidic treatment as described above gave pure **ODN1** in 31% (Figure 3C, inset; see also the Supporting Information). The oligonucleotide was fully digested, and the product of hydrolysis was analyzed by UHPLC-MS/MS, which corroborated the presence of all incorporated nucleosides. Isotope-based quantification confirmed the amounts of the epigenetic nucleosides relative to dC to be 1.2:1.2:0.9:1.0:15.0, which is in full agreement with the sequence (Figure 3B,D).

In summary, the new fdC building block **8** enabled the synthesis of fdC-containing DNA strands in unprecedented yield and quality, which will facilitate biological studies of fdC. In combination with phosphoramidite building blocks for mdC, hmdC, and cadC, it is now possible to synthesize oligonucleotides that contain all four epigenetically important cytidine nucleosides **1**–**4**. This will pave the way for a more detailed analysis of how these nucleosides influence biological processes and stem-cell development.[8]

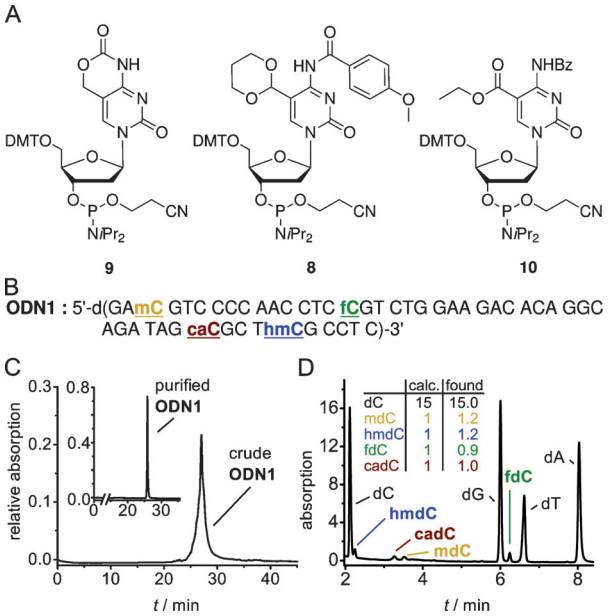


Figure 3. A) Used phophoramidites. B) Sequence of the synthesized OCT-4 promoter segment, in which cytidines in CpG contexts were substituted with the epigenetic nucleosides **1**–**4**. C) Reversed-phase HPL chromatogram of crude and purified (inset) **ODN1** after complete deprotection (0–40% buffer B in 45 min). D) UHPL chromatogram and quantification data of digested **ODN1**. Bz = benzoyl.

Received: September 28, 2013
Published online: November 26, 2013

Keywords: 5-formylcytosine · DNA methylation · epigenetics · oligonucleotides · solid-phase synthesis

[1] J. A. Law, S. E. Jacobsen, *Nat. Rev. Genet.* **2010**, *11*, 204–220.
[2] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
[3] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
[4] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem.* **2011**, *123*, 7146–7150; *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
[5] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
[6] C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583.
[7] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
[8] C. X. Song, C. He, *Trends Biochem. Sci.* **2013**, *38*, 480–484.
[9] 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.

[10] B. Steigenberger, S. Schiesser, B. Hackner, C. Brandmayr, S. K. Laube, J. Steinbacher, T. Pfaffeneder, T. Carell, *Org. Lett.* **2013**, *15*, 366 – 369.
[11] S. Tardy-Planechaud, J. Fujimoto, S. S. Lin, L. C. Sowers, *Nucleic Acids Res.* **1997**, *25*, 553 – 559.
[12] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, *17*, 13782 – 13788.
[13] Q. Dai, C. He, *Org. Lett.* **2011**, *13*, 3446 – 3449.
[14] N. Karino, Y. Ueno, A. Matsuda, *Nucleic Acids Res.* **2001**, *29*, 2456 – 2463.
[15] T. M. A. Shaikh, L. Emmanuvel, A. Sudalai, *J. Org. Chem.* **2006**, *71*, 5043 – 5046.
[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] M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671 – 5673.
[18] W. Bannwarth, A. Trzeciak, *Helv. Chim. Acta* **1987**, *70*, 175 – 186.
[19] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859 – 1862.
[20] N. D. Sinha, J. Biernat, J. McManus, H. Köster, *Nucleic Acids Res.* **1984**, *12*, 4539 – 4557.

