

Bioconjugation

International Edition: DOI: 10.1002/anie.201906737

German Edition: DOI: 10.1002/ange.201906737

Squaramate-Modified Nucleotides and DNA for Specific Cross-Linking with Lysine-Containing Peptides and Proteins

Ivana Ivancová, Radek Pohl, Martin Hubálek, and Michal Hocek*

Abstract: Squaramate-linked 2'-deoxycytidine 5'-O-triphosphate was synthesized and found to be good substrate for KOD XL DNA polymerase in primer extension or PCR synthesis of modified DNA. The resulting squaramate-linked DNA reacts with primary amines to form a stable diamide linkage. This reaction was used for bioconjugations of DNA with Cy5 and Lys-containing peptides. Squaramate-linked DNA formed covalent cross-links with histone proteins. This reactive nucleotide has potential for other bioconjugations of nucleic acids with amines, peptides or proteins without need of any external reagent.

Protein–DNA interactions are of crucial importance in DNA packaging, replication, transcription, epigenetic modifications, and repair.^[1] Transcription factors (TFs) are particularly important DNA-binding proteins that regulate gene expression through sequence-specific binding to promoter sequences. Among the approximately 1600 known human TFs, the detailed biological role and binding motifs are fully understood only for a small fraction.^[2] Although, there is a number of methods for studying of protein–DNA interactions and for identification of DNA-binding proteins,^[3] there is still an urgent need of other alternative methods, in particular for weakly binding proteins. Covalent cross-linking is one of the most promising methods for identification of DNA-binding proteins, but covalent protein–DNA conjugates are also useful for other applications in chemical biology or biosensing.^[4]

There are some general non-specific cross-linking methods based on photochemical generation of radicals (from 5-halouracils)^[5] or carbenes (from diazirine-linked nucleobases)^[6] in DNA which bind randomly to neighboring amino

acids through C–H activations. More challenging but potentially very useful are reactions specific for one or several amino-acid side-chains, but so far a very limited number of them have been reported for DNA–protein cross-linking. Thiol-linked DNA can cross-link with Cys-containing proteins through disulfide formation.^[7] Vinylsulfonamide-linked DNA was reported to cross-link with Cys,^[8] whereas chloroacetamide cross-linked with proteins through Cys or His.^[9] In both cases, proximity effect was crucial for efficient formation of the covalent cross-link between modified DNA and protein. Most frequent were reports on cross-linking of aldehyde-linked DNA with Lys either through inefficient and reversible Schiff-base formation^[10] or (more often) through irreversible reductive amination,^[11,12] which, requires an additional stoichiometric reductant (for example, toxic NaBH₃CN, which complicates any in cellulose or in vivo usage). Lys–DNA interactions are very frequent and important, in particular in histones. So far, no reactive nucleobase-modification in DNA has been reported to form irreversible cross-links with Lys without an external reagent.

Mono-amides of squaric acid (squaramates) are often used for bioconjugations with Lys and other amines.^[13] Diamides (squaramides) have been used as a phosphate surrogate in nucleotide^[14] or oligonucleotide (ON) analogues.^[15] A chemically synthesized 2'-sugar-linked squaramate–RNA conjugate, prepared through reaction of 2'-amino-modified RNA with diethyl squarate, was reported to cross-link to aminoacyl-transferase FemX_W,^[16] as the only example of its use in nucleic-acid conjugation. Within the framework of our program aimed at base-functionalized nucleic acids for applications in chemical biology,^[17] we designed novel squaramate-linked cytosine 2'-deoxyribonucleoside triphosphate (dNTP) for the enzymatic synthesis of modified DNA and cross-linking with proteins.

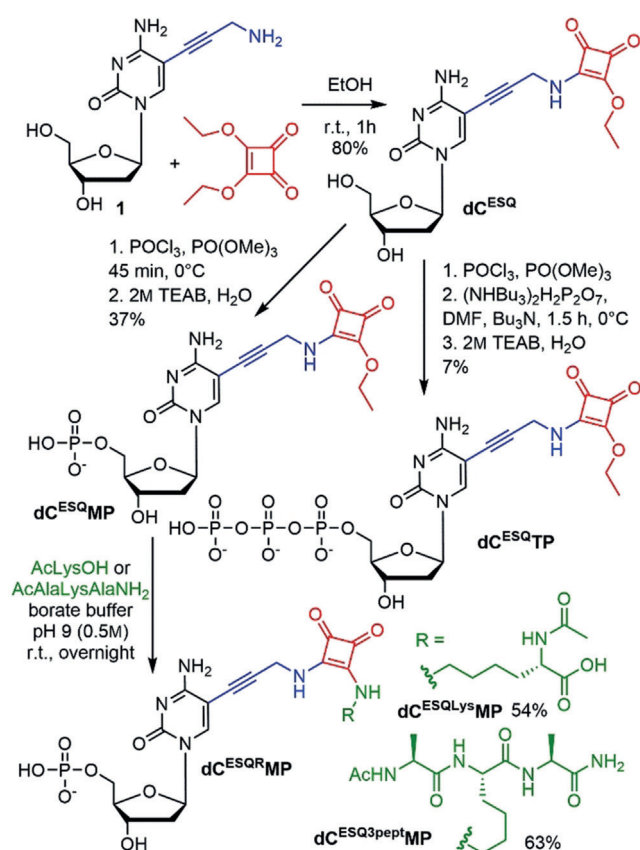
The synthesis of the desired modified nucleotides started with the preparation of 5-(3-aminopropynyl)-2'-deoxycytidine (**1**) by deacylation of known trifluoroacetamide,^[18] see Scheme S1 in Supporting Information. The reaction of amine **1** with 2 equiv of diethyl squarate gave the squaramate-linked nucleoside **dC^{ESQ}** in 80% yield (Scheme 1). Standard Yoshikawa phosphorylation^[19] with POCl₃ gave the monophosphate **dC^{ESQ}MP** in 37% yield, whereas the triphosphorylation^[20] with POCl₃ followed by pyrophosphate and triethylammonium bicarbonate (TEAB) gave the triphosphate **dC^{ESQ}TP** in 7% yield. The monophosphate **dC^{ESQ}MP** served as model compound for reactions with Lys and peptides (Scheme 1). The reaction of **dC^{ESQ}MP** with Ac-Lys or a Lys-containing tripeptide proceeded at room temperature in borate buffer (pH 9) overnight to give the desired conjugates **dC^{ESQ}LysMP** or **dC^{ESQ}3pepMP** in 54 and 63%,

[*] I. Ivancová, Dr. R. Pohl, Dr. M. Hubálek, Prof. Dr. M. Hocek
Institute of Organic Chemistry and Biochemistry,
Czech Academy of Sciences
Flemingovo nám. 2, 16610 Prague 6 (Czech Republic)
E-mail: hocek@uochb.cas.cz
Homepage: <http://www.uochb.cas.cz/hocekgroup>

I. Ivancová, Prof. Dr. M. Hocek
Department of Organic Chemistry, Faculty of Science,
Charles University in Prague
Hlavova 8, CZ-12843 Prague 2 (Czech Republic)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.201906737>.

© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Scheme 1. Synthesis of squamate-modified 2'-deoxycytidine and 2'-deoxycytidine mono- and triphosphate and dC^{ESQ}MP adducts with AcLys and lysine-containing tripeptide (AcAlaLysAlaNH₂).

respectively. These nucleotide-peptide conjugates were isolated by HPLC in pure form and fully characterized by NMR spectroscopy and mass spectrometry (MS) to confirm the expected formation of the amide bond with Lys.

Then, we tested the squamate-linked dNTP (dC^{ESQ}TP) as substrate for DNA-polymerase-catalyzed synthesis of modified DNA (Figure 1a). First, we performed the primer extension (PEX) in the presence of KOD XL DNA polymerase with either 19-, 20-, 31- or 98-mer template and a 13-, 15- or 25-mer primer (for the oligonucleotide sequences, see Tables S1 and S2 in the Supporting Information). In all cases (Figure 1b and Supporting Information, Figure S1), we observed the formation of full-length PEX products containing one, four or eighteen dC^{ESQ} modifications. PCR amplification with either 98-bp or 235-bp templates also proceeded well, giving a strong band corresponding to the modified amplicon (Supporting Information, Figure S2). This shows that the reactive dC^{ESQ}TP nucleotide does not react with DNA polymerase (neither during extension nor when modified DNA is used as template) and is a good substrate and building block for the enzymatic synthesis of reactive, modified DNA probes.

Next, we tested the cross-linking reactions of dC^{ESQ}-linked DNA with amines and peptides (Figure 1a,c). The reactions of 20-bp DNA-C^{ESQ} were performed at room temperature in borate buffer (pH 9). The reaction with Sulfo-

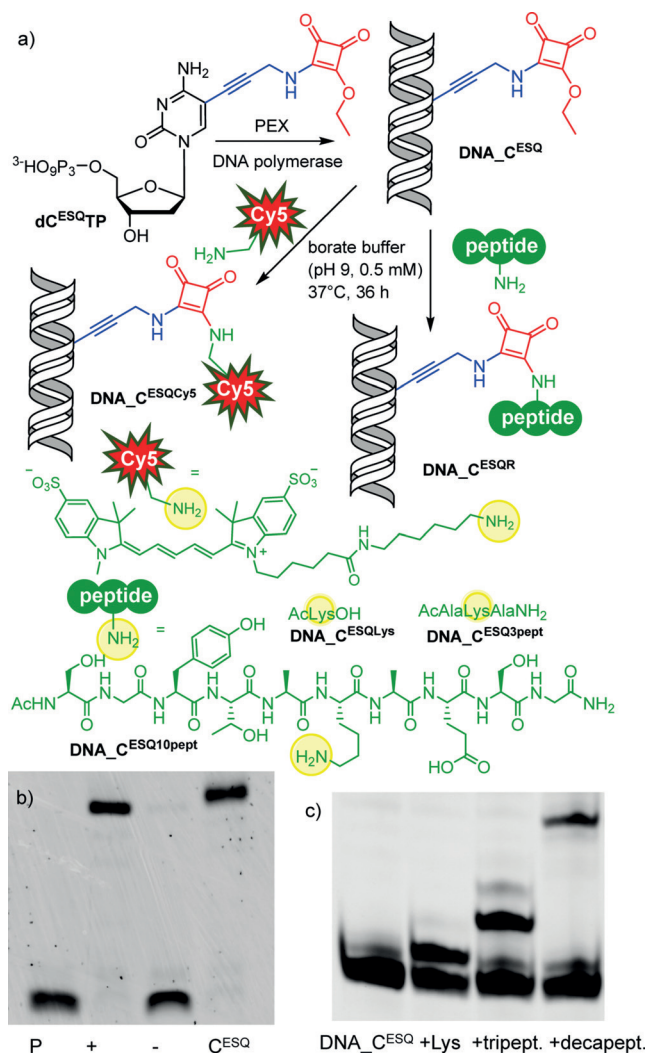


Figure 1. a) Synthesis of squamate-modified DNA (DNA-C^{ESQ}) and its cross-linking with Sulfo-Cy5-NH₂, lysine, tri- and decapeptide. b) SDS-PAGE analysis of the incorporation of dC^{ESQ}TP into DNA using KOD XL polymerase and temp²⁰ 1°C. P: primer; (+): natural dNTPs; (-): natural dNTPs without dC^{ESQ}TP, dGTP, dTTP, dATP. c) SDS-PAGE analysis of the conjugation of DNA-C^{ESQ} (4.3 μM) with Ac-Lys (11 mM), tripeptide (11 mM), and decapeptide (11 mM). Conditions: borate buffer (0.5 M, pH 9), 37°C, 36 h. For the oligonucleotide sequence see Table S1 in the Supporting Information.

Cy5-NH₂ (100 equiv) gave the desired fluorescently labeled DNA-C^{ESQ}Cy5 (Supporting Information, Figure S3). Analogous reactions of DNA-C^{ESQ} with Ac-Lys and Lys-containing tripeptide or decapeptide were conducted with large excess (approximately 2500 equiv) of the peptides, since no proximity effect was expected in these non-DNA-binding peptides. Under these conditions, the reactions proceeded with moderate efficiency and gave the desired cross-linked conjugates DNA-C^{ESQ}Lys, DNA-C^{ESQ}3pept or DNA-C^{ESQ}10pept with 50, 43, and 20% conversions, respectively (Figure 1c). All conjugates were characterized and confirmed by MALDI MS (Supporting Information, Table S4).

Finally, the reactive DNA probe DNA-C^{ESQ} was tested in reactions with proteins. We used bovine serum albumin

(BSA) as negative control of a protein containing 60 lysines that does not interact with DNA, the core-domain of p53^[21] as a DNA-binding protein containing lysines but not in the binding site, and a set of recombinant H2A, H2B, H3.1, and H4 histones, as examples of Lys-rich proteins that strongly bind to DNA. The cross-linking reactions were performed with only 2 equiv of the corresponding proteins. Since the histones are known to form dimers and oligomers, we assume that this ratio is probably effectively close to equimolar. To be closer to physiological conditions, we used phosphate (or TRIS or HEPES, see Figure S8 in the Supporting Information) buffers (pH 7.4).

At first, we performed a simple kinetic study of the cross-linking reaction of **DNA_C^{ESQ}** with histone H3.1 to show that the reaction reaches the maximum conversion in 16–24 h (Supporting Information, Figure S5). Therefore, we used 36 h reactions in other cases to ensure sufficient conversions. Figure 2 shows the results of the cross-linking experiments with proteins. To our delight, the reactions of **DNA_C^{ESQ}** with all four recombinant histones gave the covalent cross-linked conjugates with lower mobility on a denaturing SDS-PAGE gel (Figure 2b). The conversions of these reactions calculated from the SDS-PAGE were 31–34% (Supporting Information, Table S5). The identity of the covalent DNA–protein conjugates with H2B, H3.1, and H4 histones was also confirmed by SDS-PAGE with protein staining (Coomassie Blue, Supporting Information, Figure S7) and by HPLC-MS analysis using electrospray ionization (Supporting Information, Figures S16–S18). Also, a longer 98-bp PEX product containing 18 squaramate groups reacted with histone H3.1, though

mixture of cross-linked products was obtained (Supporting Information, Figure S10). On the other hand, **DNA_C^{ESQ}** did not cross-link with BSA or p53 (Supporting Information, Figure S9) or with DNA polymerase during the PEX or PCR. These results show that the proximity effect (presence of lysine(s) close to the DNA-binding site of protein) is crucial for efficient cross-linking in the absence of large excess of the peptide or protein.

In conclusion, we designed and synthesized a novel squaramate-linked dNTP (**dC^{ESQ}TP**) and demonstrated that it was a very good substrate for KOD XL DNA polymerase in PEX or PCR synthesis of reactive DNA probes. The squaramate group reacts with amines to form a stable covalent diamide (squaramide) linkage. We have shown that the **dC^{ESQ}**-modified DNA probes reacted with amino-linked Cy5 to form fluorescently labeled DNA. Its reactions with Lys-containing peptides proceeded only when a large excess of the peptide was present. On the other hand, in reactions with Lys-containing DNA-binding proteins, where the proximity effect helps, the reactions proceed with good conversions even in almost equimolar ratio. Compared to previously reported DNA–Lys conjugations based on reductive amination,^[11,12] the squaramate modification and its transformation to a stable amide proceeds under physiological conditions (at pH 7.4–9) and does not require any external reagent (i.e. toxic NaBH₃CN used in reductive aminations^[11,12]). Therefore, this reactive modification and the presented methodology has good potential in the post-synthetic labeling of DNA,^[22] bioconjugations of DNA with peptides, proteins or other biomolecules,^[4] as well as in cross-linking experiments to identify and study DNA-binding proteins. Further research along these lines is under way in our lab.

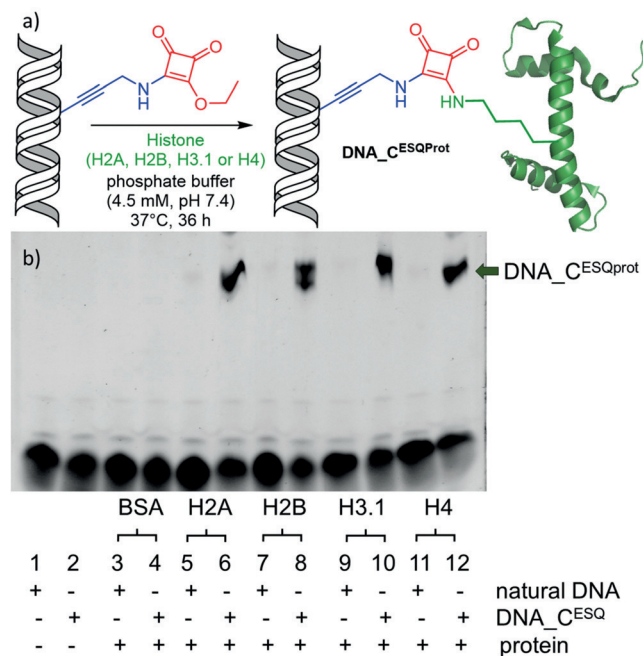


Figure 2. a) Cross-linking of squaramate-modified DNA (**DNA_C^{ESQ}**) with histone recombinant proteins. b) SDS-PAGE analysis of cross-linking experiments with natural or modified DNA and various recombinant proteins (2 equiv of protein to DNA): 17.5% SDS-PAGE gel. Conditions: phosphate buffer (4.5 mM, pH 7.4), 37°C, 36 h.

Acknowledgements

This work was supported by the Czech Academy of Sciences (Praemium Academiae award to M. Ho.) and by the Czech Science Foundation (18-03305S to I. I. and M. Ho.). The authors thank J. Konč (University of Cambridge) for advice with MS characterization of the conjugates.

Conflict of interest

The authors declare no conflict of interest.

Keywords: bioconjugation · cross-linking reactions · DNA · DNA polymerase · proteins

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 13345–13348
Angew. Chem. **2019**, *131*, 13479–13482

- [1] R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig, R. S. Mann, *Annu. Rev. Biochem.* **2010**, *79*, 233–269.
- [2] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, M. T. Weirauch, *Cell* **2018**, *172*, 650–665.
- [3] T. S. Furey, *Nat. Rev. Genet.* **2012**, *13*, 840–852.

- [4] a) J. B. Trads, T. Tørring, K. V. Gothelf, *Acc. Chem. Res.* **2017**, *50*, 1367–1374; b) N. Y. Tretyakova, A. Groehler, S. Ji, *Acc. Chem. Res.* **2015**, *48*, 1631–1644; c) C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2010**, *49*, 1200–1216; *Angew. Chem.* **2010**, *122*, 1220–1238; d) J. L. Vinkenborg, G. Mayer, M. Famulok, *Angew. Chem. Int. Ed.* **2012**, *51*, 9176–9180; *Angew. Chem.* **2012**, *124*, 9311–9315; e) C. B. Rosen, A. L. B. Kodal, J. S. Nielsen, D. H. Schaffert, C. Scavenius, A. H. Okholm, N. V. Voigt, J. J. Enghild, J. Kjems, T. Tørring, K. V. Gothelf, *Nat. Chem.* **2014**, *6*, 804–809; f) T. B. Nielsen, R. P. Thomsen, M. R. Mortensen, J. Kjems, P. F. Nielsen, T. E. Nielsen, A. L. B. Kodal, E. Clo, K. V. Gothelf, *Angew. Chem. Int. Ed.* **2019**, <https://doi.org/10.1002/anie.201903134>; *Angew. Chem.* **2019**, <https://doi.org/10.1002/ange.201903134>.
- [5] C. L. Norris, P. L. Meisenheimer, T. H. Koch, *J. Am. Chem. Soc.* **1996**, *118*, 5796–5803.
- [6] a) M. Winnacker, S. Breeger, R. Strasser, T. Carell, *ChemBioChem* **2009**, *10*, 109–118; b) L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.* **2013**, *52*, 10553–10558; *Angew. Chem.* **2013**, *125*, 10747–10752.
- [7] a) H. Huang, R. Chopra, G. L. Verdine, S. C. Harrison, *Science* **1998**, *282*, 1669–1675; b) Y. Mishina, C. He, *J. Am. Chem. Soc.* **2003**, *125*, 8730–8731.
- [8] J. Dadová, P. Orság, R. Pohl, M. Brázdová, M. Fojta, M. Hocek, *Angew. Chem. Int. Ed.* **2013**, *52*, 10515–10518; *Angew. Chem.* **2013**, *125*, 10709–10712.
- [9] A. Olszewska, R. Pohl, M. Brázdová, M. Fojta, M. Hocek, *Bioconjugate Chem.* **2016**, *27*, 2089–2094.
- [10] a) E.-A. Raiber, G. Portella, S. Martínez Cuesta, R. Hardisty, P. Murat, Z. Li, M. Iurlaro, W. Dean, J. Spindel, D. Beraldi, Z. Liu, M. A. Dawson, Reik, W. S. Balasubramanian, *Nat. Chem.* **2018**, *10*, 1258–1266; S. Balasubramanian, *Nat. Chem.* **2018**, *10*, 1258–1266; b) F. Li, Y. Zhang, J. Bai, M. M. Greenberg, Z. Xi, C. Zhou, *J. Am. Chem. Soc.* **2017**, *139*, 10617–10620; c) S. Ji, H. Shao, Q. Han, C. L. Seiler, N. Y. Tretyakova, *Angew. Chem. Int. Ed.* **2017**, *56*, 14130–14134; *Angew. Chem.* **2017**, *129*, 14318–14322.
- [11] a) S. Ji, I. Fu, S. Naldiga, H. Shao, A. K. Basu, S. Broyde, N. Y. Tretyakova, *Nucleic Acids Res.* **2018**, *46*, 6455–6469; b) P. Pande, S. Ji, S. Mukherjee, O. D. Schärer, N. Y. Tretyakova, A. K. Basu, *Chem. Res. Toxicol.* **2017**, *30*, 669–677; c) S. Wickramaratne, S. Mukherjee, P. W. Villalta, O. D. Schärer, N. Y. Tretyakova, *Bioconjugate Chem.* **2013**, *24*, 1496–1506.
- [12] a) V. Raindlová, R. Pohl, M. Hocek, *Chem. Eur. J.* **2012**, *18*, 4080–4087; b) M. Krömer, K. Bártová, V. Raindlová, M. Hocek, *Chem. Eur. J.* **2018**, *24*, 11890–11894.
- [13] a) R. I. Storer, C. Aciro, L. H. Jones, *Chem. Soc. Rev.* **2011**, *40*, 2330–2346; b) F. R. Wurm, H.-A. Klok, *Chem. Soc. Rev.* **2013**, *42*, 8220–8236; c) L. A. Marchetti, L. K. Kumawat, N. Mao, J. C. Stephens, R. B. P. Elmes, *Chem* **2019**, *5*, 1398–1485.
- [14] a) S. Niewiadomski, Z. Beebeejaun, H. Denton, T. K. Smith, R. J. Morris, G. K. Wagner, *Org. Biomol. Chem.* **2010**, *8*, 3488–3499; b) F. Soukarieh, M. W. Nowicki, A. Bastide, T. Pöyry, C. Jones, K. Dudek, G. Patwardhan, F. Meullenet, N. J. Oldham, M. D. Walkinshaw, A. E. Willis, P. M. Fischer, *Eur. J. Med. Chem.* **2016**, *124*, 200–217; c) Y. Zhang, M. Jumppanen, M. M. Maksimainen, S. Auno, Z. Awol, L. Ghemtio, H. Venkannagari, L. Lehtiö, J. Yli-Kauhaluoma, H. Xhaard, G. Boije Af Gennäs, *Bioorg. Med. Chem.* **2018**, *26*, 1588–1597.
- [15] a) K. Sato, K. Seio, M. Sekine, *J. Am. Chem. Soc.* **2002**, *124*, 12715–12724; b) E.-M. Dürr, W. Doherty, S. Y. Lee, A. H. El-Sagheer, A. Shivalingam, P. J. McHugh, T. Brown, J. F. McGouran, *ChemistrySelect* **2018**, *3*, 12824–12829.
- [16] M. Fonvielle, N. Sakkas, L. Iannazzo, C. Le Fournis, D. Patin, D. Mengin-Lecreux, A. El-Sagheer, E. Braud, S. Cardon, T. Brown, M. Arthur, M. Etheve-Quellejeu, *Angew. Chem. Int. Ed.* **2016**, *55*, 13553–13557; *Angew. Chem.* **2016**, *128*, 13751–13755.
- [17] a) M. Hocek, *J. Org. Chem.* **2014**, *79*, 9914–9921; b) M. Hocek, *Acc. Chem. Res.* **2019**, *52*, 1730–1737.
- [18] S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.* **2005**, *127*, 15071–15082.
- [19] M. Yoshikawa, T. Kato, T. Takenishi, *Tetrahedron Lett.* **1967**, *8*, 5065–5068.
- [20] J. Ludwig, *Acta Biochim. Biophys. Hung.* **1981**, *16*, 131–134.
- [21] a) M. Brazdova, J. Palecek, D. I. Cherny, S. Billova, M. Fojta, P. Pecinka, B. Vojtesek, T. M. Jovin, E. Palecek, *Nucleic Acids Res.* **2002**, *30*, 4966–4974; b) M. Fojta, H. Pivonkova, M. Brazdova, K. Nemcova, J. Palecek, B. Vojtesek, *Eur. J. Biochem.* **2004**, *271*, 3865–3876.
- [22] Reviews: a) Y. Singh, P. Murat, E. Defrancq, *Chem. Soc. Rev.* **2010**, *39*, 2054–2070; b) M. Merkel, K. Peewasan, S. Arndt, D. Ploschik, H.-A. Wagenknecht, *ChemBioChem* **2015**, *16*, 1541–1553.

Manuscript received: May 30, 2019

Revised manuscript received: July 12, 2019

Accepted manuscript online: July 22, 2019

Version of record online: August 13, 2019