VIP Nucleic Acid Analogues

DOI: 10.1002/anie.201303288

Binary Genetic Cassettes for Selecting XNA-Templated DNA Synthesis In Vivo**

Valérie Pezo, Feng Wu Liu, Mikhail Abramov, Mathy Froeyen, Piet Herdewijn, and Philippe Marlière*

Information transfer between natural nucleic acids (DNA and RNA) and xenobiotic nucleic acids (XNA) is rapidly gaining momentum for extending the range of chemical constitutions and the format of molecular evolution accessible to living organisms.[1–3] Artificial coding by nucleic acid analogues previously focused on structural alterations of base pairs to expand the alphabet of genetic messages.[4–7] Studies were mostly conducted ex vivo and few experiments have succeeded in vivo thus far. Kool and collaborators demonstrated that size-expanded nucleobases can serve as template for DNA synthesis in *E. coli*.[8] Substitution of thymine for 5-chlorouracil in a whole genome could be performed through automated evolution of *E. coli*.[9]

Conveying genetic information to DNA from an XNA with a chemically deviant backbone is amenable to tight metabolic selection, as demonstrated for hexitol nucleic acid (HNA) using the thymidylate synthase screen in *E. coli*.[10] We have now shown that various combinations of only the two bases guanine and thymine can be used to encode the active site of thymidylate synthase. This finding was exploited to simplify the synthesis of XNA to be assayed as templates for DNA biosynthesis in vivo, by halving the alphabet needed for this purpose. It could thus be demonstrated that cyclohexenyl nucleic acid (CeNA) can serve in vivo as template, mobilizing a limited effort of chemical synthesis. Further simplification of the binary system to uracil and hypoxanthine enabled to reprogram *E. coli* with templates simultaneously bearing noncanonical bases and a noncanonical backbone, namely arabinofuranosyl nucleic acid (AraNA) and HNA.

A functional *thyA* gene encoding thymidylate synthase is absolutely required by *E. coli* cells to grow in nutrient medium devoid of thymine or thymidine (TLM, thymidine-less medium).[11] We took advantage of this selection scheme for constructing a plasmid carrying a defective *thyA* gene in which the six codons specifying the active site around the cysteine at position 146 have been deleted, leaving a gap when digested with the restriction enzymes NheI and NsiI.[10] Mosaic DNA oligonucleotides in which several of the six codons are carried by an XNA backbone can be tested for informational transfer simply by selecting for active *thyA* genes after transformation of the *thyA*-deficient strain G929 with heteroduplex ligation products (Figure 1). Up to six contiguous HNA nucleotides were found to serve as a short template for *E. coli* replication enzymes.[10]

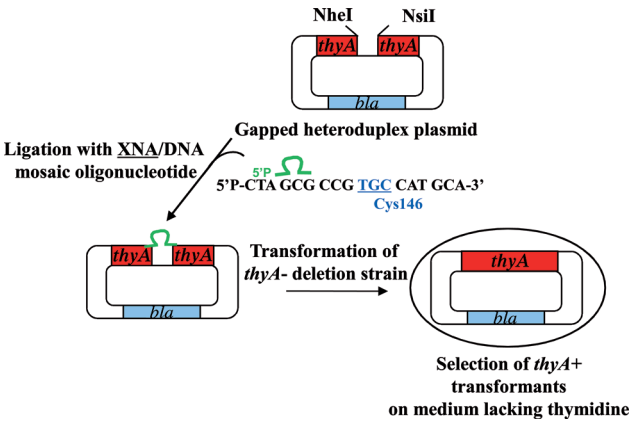


Figure 1. Selection screen to identify artificial XNA oligomers capable of templating DNA synthesis in vivo. *bla* = gene encoding β lactamase.

The first goal of the present study was to compact the coding set for specifying the active site of thymidylate synthase to two letters, namely T and G, or their pairing equivalents uracil (U) and hypoxanthine (I). In this way, we hoped to reduce the burden of organic synthesis for evaluating the coding potential of various XNA structures in vivo. Although the four A, G, C, and T monomers are available in the case of CeNA, HNA, araNA, and glycerol nucleic acid (GNA), quicker exploration of other XNA candidates should result from binary coding of amino acids at a selectable site. The choice of hypoxanthine (I) has the advantage over G that no base protection is required during oligonucleotide syn-

[*] Dr. F. W. Liu, Dr. M. Abramov, Prof. M. Froeyen, Prof. P. Herdewijn
Laboratory of Medicinal Chemistry, Rega Institute for Medical Research, KU Leuven
Minderbroedersstraat 10, 3000 Leuven, (Belgium)
E-mail: Piet.Herdewijn@rega.kuleuven.be
Dr. V. Pezo, Prof. P. Herdewijn, Dr. P. Marlière
ISSB, Génopole, Genavenir6, Equipe Xénome
5 rue Henri Desbruères, 91030 Evry Cedex (France)
Dr. V. Pezo
CEA, DSV, IG, Genoscope
2 rue Gaston Crémieux, 91057 Evry Cedex (France)

[**] We thank G. Schepers for the synthesis of the oligonucleotides, Dr. J. Rozenski for mass spectrometric analysis, and Lan N'Guyen and Mélanie Crouzard for technical assistance. We thank KU Leuven (GOA, IDO), FWO (Vlaanderen) and Genoscope (Genopole, Evry) for financial support (ATIGE, EA 4527, UPS 3509). The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement n° ERC-2012-ADG_20120216/320683. XNA = xenobiotic nucleic acids.

Supporting information (including synthesis, spectroscopic characterization of monomers and oligonucleotides, molecular modeling, and biological testing) for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201303288.

thesis, just as for T and U.[12] In addition, encoding an enzyme in a two-letter rather than four-letter code poses a functional challenge of "protein simplification", as investigated by Hilvert and collaborators.[13,14]

The catalytic residue Cys146 of ThyA (Figure 2), which is known not be replaceable by any other amino acid, can be encoded by the triplet TGT. The five other amino acids that

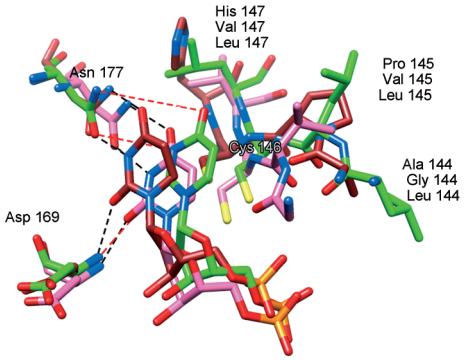


Figure 2. Superposition of thymidylate active sites from wild-type (WT) *E. coli* (in dark red), active mutant A144G-P145V-C146-H147V (in pink) and inactive mutant A144L-P145L-C146-H147L (in green). The WT structure corresponds to the B subunit of the homodimer in the X-ray structure. Representative active and inactive mutant structures were inferred from the 3 ns MD trajectory to highlight the hydrogen bond (HB) changes with the uracil moiety of the dUMP substrate shown at the bottom of the figure. The methylenetetrahydrofolate cosubstrate is not shown. The HB distances UMP903.O2/Asp169.N (3.0, 2.9, 4.0 Å for WT, active and inactive mutant), UMP903.O4/Asp177.ND2 (3.0, 2.9, 5.2 Å for WT, active mutant, and inactive mutant) and UMP903.N3/Asp177.OD1 (3.0, 2.9, 4.3 Å for WT, active mutant, and inactive mutant) are indicated by dashed lines. HBs in the inactive mutant are red. Structural comparison thus points to the loss of these HBs in the inactive mutant. Figure generated using the Chimera software.[26]

can be encoded with binary G/T triplets, Phe, Leu, Trp, Val, and Gly, belong to a small subset of aliphatic and aromatic residues.

We resorted to a combinatorial library of synthetic oligonucleotides for assessing whether a stretch of G's or T's could be extended upstream and downstream of the Cys codon without losing the function of the ThyA enzyme.

Table 1 shows the combinatorial cassettes (K7 and K10) that were used in the selection process. The cassette K7 consisted of seven nucleotides (K stands for T or G) covering three amino acid codons, corresponding to a library of 128 different sequences. The cassette K10 consisted of ten nucleotides covering four amino acid codons, corresponding to a library of 1024 sequences. Independent ligation of the two oligonucleotide libraries K7 and K10 with the gapped plasmid heteroduplex (Figure 1) and subsequent transformation of the *thyA*-deleted strain G929 and growth on TLM plates resulted in growing colonies in both cases.

The DNA and protein sequences corresponding to such colonies with functional *thyA* genes are given in Table 2 A, B. As expected, only cysteine (encoded by TGT) was found at the catalytically active site 146 (cassette K10; Table 1 and Table 2).

Despite the high conservation throughout evolution of the Ala-Pro-Cys-His sequence in the active site of thymidylate synthase,[11] Val (GTT or GTG), Cys (TGT), and Gly (GGT or GGG) were found to be acceptable replacements at positions

Table 2: Plasmid sequences from thymidine prototrophic transformants of *E. coli* originating from G/T DNA binary cassettes. A) DNA sequences; B) protein sequences.

A)	WT		Leu^{143}Ala^{144}Pro^{145}Cys^{146}His^{147}Ala^{148}		No. of *thyA*$^+$ colonies
	K7	5′P-	CTA GKK KKK TGT KKT GCA	-3′	
			CTA GGT GTT TGT GTT GCA		2
			CTA GGT GTT TGT TGT GCA		1
			CTA GGG GTT TGT TGT GCA		1
			CTA GGG GTG TGT TGT GCA		2
			CTA GTT GTT TGT GGT GCA		3
			CTA GTT TGT TGT GGT GCA		2
			CTA GTT TGT TGT TGT GCA		1
			CTA GTG TGT TGT GGT GCA		3
			CTA GTG TGT TGT GTT GCA		2
			CTA GTG GTG TGT GGT GCA		1
			CTA GTG GTT TGT GGT GCA		1
			CTA GTG GTT TGT TGT GCA		1
			CTA GTG GGT TGT GTT GCA		1
	K10	5′P-	CTA GKK KKK KKK KKT GCA	-3′	
			CTA GTG TGT TGT GTT GCA		2
			CTA GGT GTG TGT TGT GCA		1

B)	WT	Ala^{144}Pro^{145}Cys^{146}His^{147}	No. of clones
	K7	Gly Val Cys Val	4
		Gly Val Cys Cys	2
		Val Val Cys Gly	5
		Val Cys Cys Gly	5
		Val Cys Cys Cys	1
		Val Cys Cys Val	2
		Val Val Cys Cys	1
		Val Gly Cys Val	1
	K10	Val Cys Cys Val	1
		Gly Val Cys Cys	1

Table 1: Statistical composition of G/T binary DNA cassettes encoding the active site of thymidylate synthase in the *E. coli thyA* gene.

	Oligonucleotides[a] Leu^{143}Ala^{144}Pro^{145}Cys^{146}His^{147}Ala^{148}	No. of oligonucleotides in the library	No. of *bla*$^+$ colonies $\times 10^4$	No. of *thyA*$^+$ colonies	Ratio *thyA*$^+$/*bla*$^+$
WT	5′P-CTA GCG CCG TGC CAT GCA-3′		63	245 000	0.38
K7	5′P-CTA GKK KKK TGT KKT GCA-3′	128	97	24	0.000027
K10	5′P-CTA GKK KKK KKK KKT GCA-3′	1024	16	3	0.000018

[a] K = T or G.

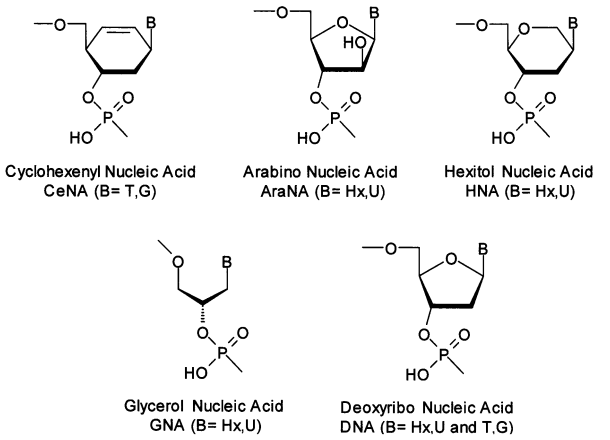
upstream and downstream of Cys 146 (Table 2). The larger G/T-encoded amino acids Phe, Leu, and Trp were found neither downstream nor upstream. Active sites further simplified were constructed separately by ligating appropriate oligonucleotides to the gapped plasmid heteroduplex. It was thus found that the sequences Cys-Cys-Cys-Cys and Val-Val-Cys-Val led to an active site, whereas the sequences Gly-Gly-Cys-Gly and Leu-Leu-Cys-Leu led to inactive enzymes.

Molecular modeling was conducted to interpret these surprising genetic findings by structural analysis. Nine functional mutants (Table I in the Supporting Information) and one defective mutant (Leu-Leu-Cys-Leu) of thymidylate synthase were investigated using molecular dynamics over a simulation time of 3 ns[15] (for experimental data, see the Supporting Information). No difference was observed between the distance of the catalytic Cys146 to the C6 atom of dUMP between the active and inactive mutants (attack by the sulfur atom of Cys146 to this C6 atom is the first step in the catalytic process leading to dTMP). However, a clear difference was seen between the hydrogen bonding networks of dUMP in the active sites of the functional and defective mutants, thus suggesting that the substrate critically loses bonding strength in the latter case (Figure 2).

The second goal of the present study was to assay the templating properties of chemically simplified XNA oligonucleotides. We call transliteration the process catalyzed by DNA polymerases and other enzymes of the replication machinery through which a stretch of XNA acts as template in the condensation of deoxynucleoside triphosphates (dNTPs) to form a DNA sequence complementary to the XNA. Once formed that DNA sequence undergoes canonical replication in subsequent cell generation.

The structure of HNA and the other chemically modified nucleic acids that were used in this study are depicted in Scheme 1. Detailed description of the synthesis and analysis of the nucleoside[16–19] and of the oligonucleotide[20–22] analogues can be found in the Supporting Information.

As the dodecamer GTT-GTT-TGT-GGT was the most abundantly selected sequence among the combinatorial libraries, we used this binary sequence for testing the coding potential of CeNA, that is, a polynucleotide analogue whose backbone bears little resemblance to the oxygenated five-membered ring of ribose and deoxyribose. Yet, CeNA is still able to hybridize in a sequence-selective manner with DNA as



Scheme 1. Structure of sugar- and base-modified XNA tested as binary genetic cassettes.

well as RNA, and the triphosphates of CeNA nucleosides can serve as substrate of several DNA polymerases.[23] It is shown in Table 3 that substituting one (TGT), two (GTT-TGT or TGT-GGT), three (GTT-TGT-GGT), or four (GTT-GTT-TGT-GGT) codons in the active site of ThyA by the binary CeNA-corresponding G/T sequences led to the recovery of *thyA*$^+$ colonies with a decreasing efficiency. This experiment provided a proof of principle validating the use of binary code for systematically assessing informational transfer by xenobiotic heredity supports.

We then tested whether the binary G/T cassettes can be replaced by binary I/U XNA cassettes. Indeed, when synthesizing sugar analogues of nucleic acids, the nucleoside with the G base is invariably the most difficult to obtain. Furthermore, the presence of multiple G residues in DNA and XNA oligonucleotides frequently hinders purification, thus rendering the protection of the G base (as well as the C and A bases) mandatory for oligonucleotide synthesis. Therefore, we investigated a further chemical simplification of the G/T binary code to a I/U binary code, with hypoxanthine serving as guanine surrogate for pairing with cytosine, and uracil serving as thymine surrogate for pairing with adenine. It should be noted that both these simplified purine and pyrimidine are actively repaired when present in cellular DNA by dedicated enzymes hydrolyzing hypoxanthine or uracil deoxynucleotides.[15,24]

Table 3: Templating by G/T binary CeNA for restoring an active thymidylate synthase gene in *E. coli*.

	Oligonucleotides Leu^{143}Ala^{144}Pro^{145}Cys^{146}His^{147}Ala^{148}	No. of CeNA residues	No. of *bla*$^+$ colonies $\times 10^4$	No. of *thyA*$^+$ colonies	Ratio *thyA*$^+$/*bla*$^+$
WT	CTA GCG CCG TGC CAT GCA	0	634	2450000	0.38
Cys 146 deleted	CTA GCG CCG ΔΔΔ CAT GCA	0	285	0	0
DNA/CeNA mosaic template	CTA GTT GTT TGT GGT GCA	3	1273	2920000	0.23
	CTA GTT GTT TGT GGT GCA	6	805	490000	0.06
	CTA GTT GTT TGT GGT GCA	6	1207	700000	0.058
	CTA GTT GTT TGT GGT GCA	9	1000	3600	0.00036
	CTA GTT GTT TGT GGT GCA	12	611	900	0.00014
	CTA GTG GTG TGT GGT GCA	12	556	370	0.000066

Table 4: Templating by I/U binary XNAs for restoring an active thymidylate synthase gene in *E. coli*. A) I/U arabino nucleic acid; B) I/U hexitol nucleic acid.

		Oligonucleotides Leu[143]Ala[144]Pro[145]Cys[146]His[147]Ala[148]	No. of XNA residues	No. of *bla*$^+$ colonies $\times 10^4$	No. of *thyA*$^+$ colonies	Ratio *thyA*$^+$/*bla*$^+$
A)	DNA/**arabino I/U** mosaic template	CTA GTT GTT TGT GGT GCA	0	150.5	124000	0.08
		CTA GTT GTT UGU GGT GCA	2	294	180000	0.06
		CTA GTT GTT UIU GGT GCA	3	215	112000	0.05
		CTA GTT IUU UIU GGT GCA	6	260	6300	0.002
		CTA GTT GTT UIU IIU GCA	6	0.5	20	0.004
		CTA GTG IUU UIU IIU GCA	9	11.5	220	0.002
		CTA IUU IUU UIU GGT GCA	9	12.5	140	0.001
B)	DNA/**hexitol I/U** mosaic template	CTA GTT GTT TGT GGT GCA	0	108	130000	0.12
		CTA GTT GTT TIT GGT GCA	1	142	88800	0.06
		CTA GTT GTT UGU GGT GCA	2	57.5	45000	0.08
		CTA GTT GTT UIU GGT GCA	3	393	151200	0.04
		CTA GTT IUU UIU GGT GCA	6	42	19500	0.05
		CTA GTT GTT TGT GGT GCA	0	1739	480000	0.02
		CTA GTT GTT UIU IIU GCA	6	124	19300	0.01
		CTA GTG IUU UIU IIU GCA	9	11.5	140	0.0012
		CTA IUU IUU UIU GGT GCA	9	141	170	0.0001
		CTA IUU IUU UIU IIU GCA	12	97	100	0.0001

The same experiments as performed with binary G/T CeNA cassettes were repeated using binary I/U sequences borne on the DNA backbone and on the XNA backbones HNA[25], AraNA,[19] and GNA.[18] As expected, the presence of I (deoxyinosine) and U (deoxyuridine) in the DNA template resulted in the complete absence of *thyA*$^+$ colonies, presumably owing to degradation through the action of the N-glycosylases Ung and Fpg.

The GNA/DNA mosaic templates did not yield any *thyA*$^+$ colony, even when only one GNA codon (UIU coding for Cys146) was present. Apparently the glycerol backbone structure was too alien for being used as template by the DNA replication machinery of *E. coli*. Implementation of I/U binary XNA sequences in the two other XNA templates AraNA and HNA succeeded in restoring a functional *thyA* gene (Table 4). The yields of prototrophs were found to be similar for AraNA and HNA as for G/T binary CeNA sequences. The yield of thymidine prototrophic colonies decreased steeply with the length of the XNA stretch in the mosaic template. For each chemically modified template assayed, 16 *thyA*$^+$ colonies growing on TLM were subjected to DNA sequencing of the Cys 146 region of ThyA. This analysis revealed that the genetic message conveyed under the form of binary G/T sequences in CeNA and of binary I/U sequences in AraNA and HNA was correctly copied into DNA in each of the scrutinized cases. Faithful transmission of hereditary information can thus be accomplished from templates lacking a canonical backbone and canonical bases at the same time. This augurs well for the propagation of a third type of nucleic acid aside from DNA and RNA in bacterial cells.

A tight genetic selection screen based on the restoration of the active site of thymidylate synthase was elaborated for assaying the use of short XNA stretches carrying only the two bases T and G, as templates for DNA biosynthesis in the model bacterium *E. coli*. This screen allowed us to demonstrate that hereditary information can be conveyed to DNA under the form of CeNA, AraNA, and HNA messages but not GNA. Moreover, simultaneously altering the structure of the nucleobase and the backbone sugar did not abolish templating of DNA biosynthesis, as exemplified by faithful reading of dodecameric hexitol stretches bearing hypoxanthine and uracil. This study incidentally demonstrated that various amino acid sequences are tolerated in the active site region of thymidylate synthase. These versatile G/T or I/U binary cassettes selectable in the *thyA* gene of *E. coli* will now be applied to improve XNA-dependent DNA polymerase activity in vivo.

Received: April 18, 2013
Published online: June 26, 2013

Keywords: gene expression · hypoxanthine · nucleic acids · thymidylate synthetase · uracil

[1] P. Herdewijn, P. Marliere, *Chem. Biodiversity* **2009**, *6*, 791 – 808.
[2] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S. Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341 – 344.
[3] V. B. Pinheiro, P. Holliger, *Curr. Opin. Chem. Biol.* **2012**, *16*, 245 – 252.
[4] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33 – 37.
[5] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950 – 954.
[6] D. L. McMinn, A. K. Ogawa, Y. Q. Wu, J. Q. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585 – 11586.
[7] T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4922 – 4925.
[8] A. T. Krueger, L. W. Peterson, J. Chelliserry, D. J. Kleinbaum, E. T. Kool, *J. Am. Chem. Soc.* **2011**, *133*, 18447 – 18451.

[9] P. Marlière, J. Patrouix, V. Döring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, R. Mutzel, *Angew. Chem.* **2011**, *123*, 7247 – 7252; *Angew. Chem. Int. Ed.* **2011**, *50*, 7109 – 7114.
[10] S. Pochet, P. A. Kaminski, A. Van Aerschot, P. Herdewijn, P. Marliere, *C. R. Biol.* **2003**, *326*, 1175 – 1184.
[11] B. Lemeignan, P. Sonigo, P. Marliere, *J. Mol. Biol.* **1993**, *231*, 161 – 166.
[12] F. Seela, X. Ming, *Helv. Chim. Acta* **2008**, *91*, 1181 – 1200.
[13] K. U. Walter, K. Vamvaca, D. Hilvert, *J. Biol. Chem.* **2005**, *280*, 37742 – 37746.
[14] M. M. Müller, J. R. Allison, N. Hongdilokkul, L. Gaillon, P. Kast, W. F. van Gunsteren, P. Marliere, D. Hilvert, *PLoS Genet.* **2013**, *9*, e1003187.
[15] B. K. Tye, I. R. Lehman, *J. Mol. Biol.* **1977**, *117*, 293 – 306.
[16] B. De Bouvere, L. Kerremans, J. Rozenski, G. Janssen, A. van Aerschot, P. Claes, R. Busson, P. Herdewijn, *Liebigs Ann.* **1997**, 1453 – 1461.
[17] J. Wang, J. Morral, C. Hendrix, P. Herdewijn, *J. Org. Chem.* **2001**, *66*, 8478 – 8482.
[18] M. K. Schlegel, E. Meggers, *J. Org. Chem.* **2009**, *74*, 4615 – 4618.
[19] A. M. Noronha, C. J. Wilds, C.-N. Lok, K. Viazovkina, D. Arion, M. A. Parniak, M. J. Damha, *Biochemistry* **2000**, *39*, 7050 – 7062.
[20] C. Hendrix, H. Rosemeyer, I. Verheggen, F. Seela, A. Van Aerschot, P. Herdewijn, *Chem. Eur. J.* **1997**, *3*, 110 – 120.
[21] J. Wang, B. Verbeure, I. Luyten, E. Lescrinier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *J. Am. Chem. Soc.* **2000**, *122*, 8595 – 8602.
[22] L. Zhang, A. E. Peritz, P. J. Carroll, E. Meggers, *Synthesis* **2006**, 645 – 653.
[23] V. Kempeneers, M. Renders, M. Froeyen, P. Herdewijn, *Nucleic Acids Res.* **2005**, *33*, 3828 – 3836.
[24] B. Weiss, *Mutat. Res.* **2001**, *461*, 301 – 309.
[25] P. Herdewijn, *Chem. Biodiversity* **2010**, *7*, 1 – 59.
[26] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605 – 1612.

