Noncovalent Interactions

Deutsche Ausgabe: DOI: 10.1002/ange.201603510
Internationale Ausgabe: DOI: 10.1002/anie.201603510

Glucose–Nucleobase Pseudo Base Pairs: Biomolecular Interactions within DNA

*Empar Vengut-Climent, Irene Gómez-Pinto, Ricardo Lucas, Pablo Peñalver, Anna Aviñó, Célia Fonseca Guerra, F. Matthias Bickelhaupt, Ramón Eritja, Carlos González, and Juan C. Morales**

Abstract: *Noncovalent forces rule the interactions between biomolecules. Inspired by a biomolecular interaction found in aminoglycoside–RNA recognition, glucose-nucleobase pairs have been examined. Deoxyoligonucleotides with a 6-deoxyglucose insertion are able to hybridize with their complementary strand, thus exhibiting a preference for purine nucleobases. Although the resulting double helices are less stable than natural ones, they present only minor local distortions. 6-Deoxyglucose stays fully integrated in the double helix and its OH groups form two hydrogen bonds with the opposing guanine. This 6-deoxyglucose-guanine pair closely resembles a purine-pyrimidine geometry. Quantum chemical calculations indicate that glucose-purine pairs are as stable as a natural T-A pair.*

Noncovalent forces govern interactions among biomolecules, drug–target molecular recognition and assembly processes. Hydrogen bonds, π–π stacking, van der Waals forces, electrostatic forces, and hydrophobic interactions are observed in RNA recognition by drugs such as macrolides, tetracyclines,[1] and new designed RNA binders.[2] Aminoglycoside antibiotics are a well-known family of RNA binders which target the 16S rRNA in the small ribosomal subunit. Apart from electrostatic forces and hydrogen bonds, solution and X-ray structures of aminoglycosides with rRNA show a singular biomolecular interaction, a monosaccharide–nucleobase stacking interaction [e.g., 2′-amino-2′-deoxyglucose (ring I) of paromomycin stacks over guanine 1491; Figure 1].[3]

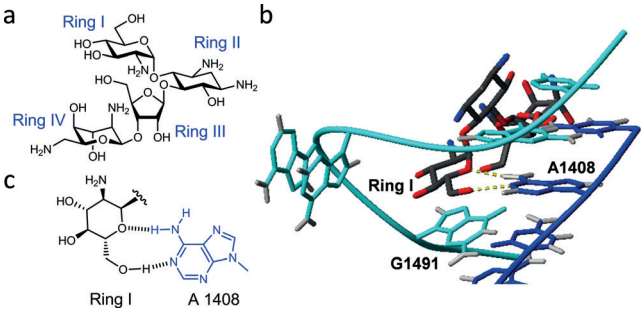


Figure 1. a) Structure of paromomycin. b) Detail of the solution structure of paromomycin binding a 16S RNA model sequence. c) Drawing of the glycoside-adenine 4108 pseudo base pair.

Recently, we have studied sugar–DNA stacking interactions using a self-complementary CGCGCG sequence with the carbohydrates directly linked to the 5′-end of DNA. We observed stacking of mono- and disaccharides on top of the terminal DNA base pairs, which stabilized the duplex between −0.5 to −1.8 kcal mol^{-1}.[4] Stacking of sugars on top of the guanine tetrad of a DNA G-quadruplex was also characterized.[4b]

A second singular biomolecular interaction found when paramomycin binds rRNA is a glycoside-adenine pseudo base pair (Figure 1c).[3] Two hydrogen bonds are formed between ring I of paromomycin and A1408. Inspired by this interaction we decided to study the possible formation of glucose-nucleobase pairs. Since monosaccharides can stack on DNA bases and possess OH groups capable of making hydrogen bonds at the edge of their coinlike structure, like natural bases (Figure 2a), we placed a potential glucose-nucleobase pair inside a DNA double helix to investigate its properties (Figure 2b). This model system also allows us to study possible sugar-sugar pairs. The only non-aromatic nucleobases reported previously are Leumann's phenyl cyclohexyl interstrand base[5] and Kashida's isopropylcyclohexane and methylcyclohexane DNA base insulators.[6]

[*] Dr. E. Vengut-Climent, Dr. R. Lucas, Dr. P. Peñalver, Dr. J. C. Morales
Department of Bioorganic Chemistry
Instituto de Investigaciones Químicas
CSIC—Universidad de Sevilla
Américo Vespucio 49, 41092 Sevilla (Spain)
E-mail: jcmorales@ipb.csic.es
Dr. I. Gómez-Pinto, Prof. C. González
Instituto de Química Física "Rocasolano", CSIC
28006 Madrid (Spain)
Dr. R. Lucas, Dr. P. Peñalver, Dr. J. C. Morales
Department of Biochemistry and Molecular Pharmacology
Instituto de Parasitología y Biomedicina, CSIC, Parque Tecnológico Ciencias de la Salud
18016 Armilla, Granada (Spain)
Dr. A. Aviñó, Prof. R. Eritja
Instituto de Química Avanzada de Cataluña (IQAC), CSIC
CIBER—BBN Networking Centre on Bioengineering, Biomaterials and Nanomedicine, 08034 Barcelona (Spain)
Dr. C. Fonseca Guerra, Prof. Dr. F. M. Bickelhaupt
Department of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Vrije Universiteit Amsterdam
1081 HV Amsterdam (The Netherlands)
Prof. Dr. F. M. Bickelhaupt
Institute of Molecules and Materials (IMM), Radboud University
6525 AJ Nijmegen (The Netherlands)
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201603510.

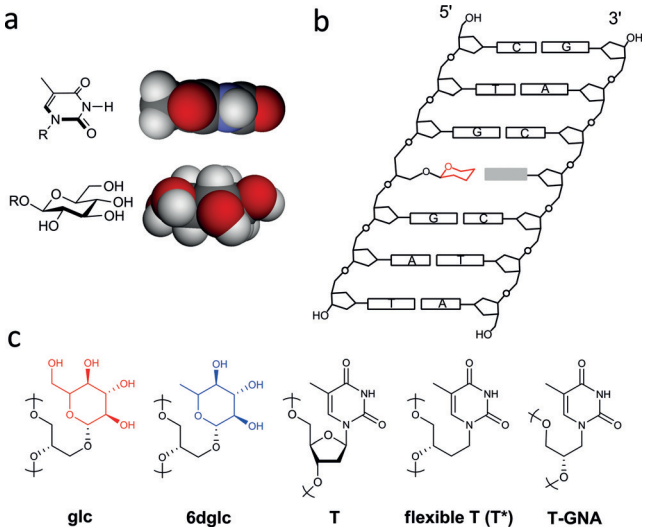


Figure 2. Description of the carbohydrate derivatives under study. a) CPK models of thymine and glucose. b) Schematic drawing of a DNA double helix containing a glycoside-nucleobase pair. c) Structures of the two sugar units prepared, (*S*)-2,3-dihydroxypropyl glucose (glc) and (*S*)-2,3-dihydroxypropyl 6-deoxyglucose (6dglc), thymidine (T), (*S*)-3,4-dihydroxybutyl thymine or flexible T (T*), and glycol T (T-GNA).

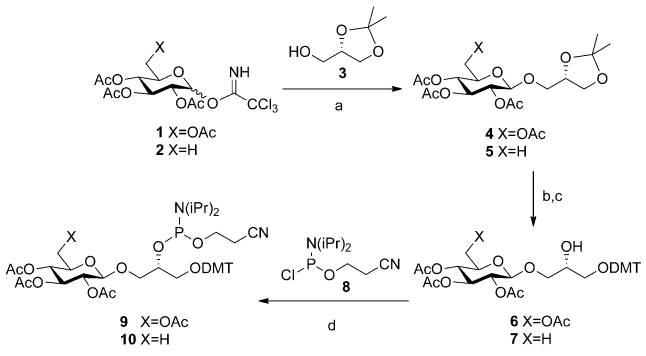
Table 1: Melting temperature (T_m) for DNA duplexes containing T*, glc, and 6dglc, and GNA duplexes containing 6dglc.

DNA duplexes		5′-d(GATGAC**X**GCTAG)[a,b] 3′-d(CTACTG**Y**CGATC)			
X-Y	T_m [°C]	X-Y	T_m [°C]	X-Y	T_m [°C]
T*-A	37.7	glc-A	30.6	6dglc-A	33.2
T*-T	32.8	glc-T	28.5	6dglc-T	29.9
T*-C	32.0	glc-C	28.7	6dglc-C	29.9
T*-G	35.6	glc-G	31.3	6dglc-G	32.7
T*-T*	31.1	glc-glc	29.6	6dglc-6dglc	32.0
GNA duplexes		*3′-TAAAATTTA**X**ATTATTAA*[b,c] *2′-ATTTTAAAT**Y**TAATAATT*			
X-Y	T_m [°C]			X-Y	T_m [°C]
T-A	57.3			A-6dglc	45.4
T-T	48.0			T-6dglc	37.4
				6dglc-T	38.7
				6dglc-6dglc	43.0

[a] The natural DNA duplex containing X-Y = T-A results in a T_m value of 47.9 °C. [b] Conditions for DNA duplexes: 10 mM NaH_2PO_4, 150 mM NaCl, pH 7.0. For GNA duplexes: 10 mM NaH_2PO_4, 200 mM NaCl, pH 7.0. Estimated errors are: ±0.4 °C (in DNA, except for 6dglc-6dglc: ±1.0 °C) and ±1.0 °C (GNA). Average value of three experiments measured at 1.2 μM conc (DNA) and 0.7 μM conc (GNA). [c] GNA monomers in italics.

Glucose (glc) was linked to the phosphodiester backbone of DNA through a flexible glycerol linker through its anomeric position. Thus, we prepared the DMT-glucose phosphoramidite derivative **9** (Scheme 1) for introduction at the required position of the oligonucleotide, similar to a standard natural DNA base phosphoramidite. To modulate the high polarity of glucose, we also prepared the 6dglc amidite **10**. A flexible T (T*) was synthesized for comparison of glc with a pyrimidine base in the same environment (Figure 2c; see Figure S1 in the Supporting Information).

We then measured thermal denaturation of DNA duplexes containing our modifications. Firstly, the introduction of a flexible linker, such as in T*, with respect to a natural



Scheme 1. Synthesis of the carbohydrate DMT phosphoramidites. Glycosyl trichloroacetimidate donors were used as starting material. Reagents and conditions: a) TMSOTf, CH_2Cl_2 for glc; $BF_3{\cdot}OEt_2$, CH_2Cl_2 for 6dglc; b) AcOH/H_2O (4:1), 80 °C; c) DMTCl, DMAP, CH_2Cl_2; d) DIPEA, CH_2Cl_2. DIPEA = diisopropylethylamine, DMAP = 4-(*N*,*N*-dimethylamino)pyridine, DMT = 4,4′-dimethoxytrityl, Tf = trifluoromethanesulfonyl, TMS = trimethylsilyl.

T leads to a 10.2–15.9 °C decrease in stability (Table 1). The decrease in stability of DNA containing single or multiple acyclic nucleosides has been previously reported.[7] Secondly, the DNA duplexes with glc and 6dglc opposite to the four natural nucleosides were destabilized in comparison to the duplexes with T-A and T*-A. Considering the penalty of the flexible linker, destabilization of glc pairs compared to T*-A ranges from 6.4 to 9.2 °C. Interestingly, the presence of the more apolar 6dglc improved the DNA stability, when compared to glc, by 1.2–3.9 °C. Still, the more stable pair, 6dglc-A, led to a DNA duplex which is 4.5 °C less stable than that with T*-A, and 15.7 °C less stable than that with T-A. The larger volume of the pyranose ring, in comparison with an aromatic ring, may distort the nearby DNA base pairs and provoke this decrease in DNA stability. Thirdly, we observed certain selectivity for glc-nucleobase and 6dglc-nucleobase pairs. Sugar-purine pairs were more stable than sugar-pyrimidine pairs (from 1.9 to 3.3 °C). This difference may be a consequence of direct hydrogen bonding between the OH groups of glc and 6dglc and the donor and acceptor groups in A and G. Alternatively, purines may be preferred in front of glucose in the sequence just for the better stacking as observed in abasic sites,[8] but the NMR spectroscopy and theoretical calculations shown below indicate that hydrogen bonding between glc and 6dglc with A and G is possible. Fourthly, the stability of DNA duplexes containing sugar-sugar pairs ranges from 29.6 to 32.3 °C. Unexpectedly, pairs containing two flexible spacers and two bulky pyranose rings do not show further DNA destabilization in comparison to sugar-natural nucleobase pairs. For example, a 6dglc-6dglc pair is as stable as a 6dglc-purine pair.

We also incorporated glc and 6dglc into a glycerol nucleic acid (GNA) double helix. GNA strands have an acyclic backbone of propylene glycol nucleosides which incorporate

the natural nucleobases in the strand and are connected by phosphodiester bonds (Figure 2c).[9] In this context, both natural bases and our sugar modifications are linked through flexible linkers to the skeleton and no energetic penalty is expected. The T_m values measured showed that the GNA duplex with an A-6dglc pair was 11.9 °C less stable than with T-A (Table 1) and only 2.6 °C less stable than with a mismatched T-T. It is quite notable that the selectivity between T-A and T-T pairs (9.3 °C) is similar to that between 6dglc-A and 6dglc-T pairs (6.7–8.0 °C). Lastly, a 6dglc-6dglc pair results in a GNA duplex stability of 43 °C, thus pointing to the formation of hydrogen bonds between the OH groups of each 6dglc unit.

We next determined the three-dimensional high-resolution structure of DNA duplexes containing a 6dglc-G and a 6dglc-T pair using restrained molecular dynamics methods based on experimental NMR distance constraints (Figure 3). The exchangeable proton region of the NMR spectra exhibited 11 imino proton signals between $\delta = 12.5$ and 14.5 ppm, thus indicating the formation of a duplex structure with Watson–Crick base pairs (see Figure S4). Additional imino proton signals are observed in the $\delta = 10$–11 ppm region corresponding to the nucleotide located opposite to the carbohydrate. Full proton assignment of the DNA and sugar units was performed with only a few exceptions (see Tables S1 and S2). DNA chemical-shift differences between the conjugates and the natural DNA control duplexes are limited to the surroundings of the carbohydrate moieties, thus indicating the overall duplex structure is not distorted (see Figure S5).

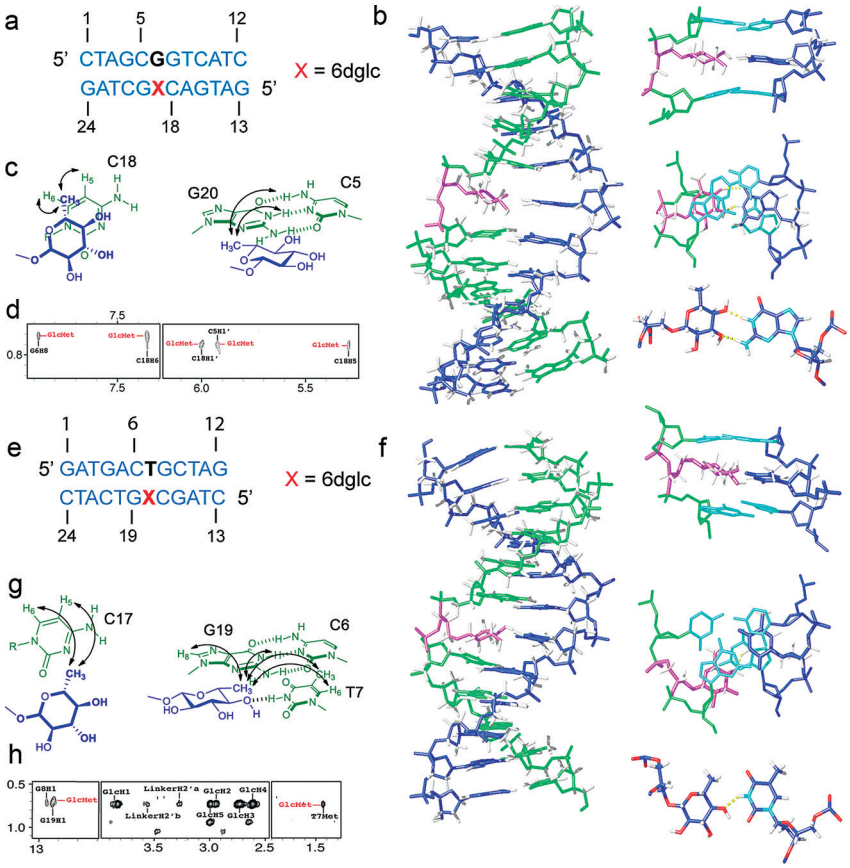


Figure 3. Solution structure of the helix 6dglc-G and helix 6dglc-T. a) Sequences of the helix 6dglc-G. b) Refined solution structure of the helix 6dglc-G and zoom views of the 6dglc-G pair. c) NOE interactions between 6dglc and its 5′-neighboring base C18 and its 3′-neighboring base pair G20-C5. d) NOESY trace showing key NOEs with the methyl group of 6dglc. e) Sequences of helix 6dglc-T. f) Refined solution structure of helix 6dglc-T and zoom views of the 6dglc-T pair. g) NOE interactions between 6dglc and its 5′-neighbouring base C17, and with its 3′-neighboring base pair G19-C6 and opposite base T7. h) NOESY trace showing key NOEs with the methyl group of 6dglc. Color code: sugar and spacer (purple), nucleobases (green and blue) and surrounding natural base pairs (light blue).

The 3D structures obtained are B-form helices without global distortions (Figure 3; see Figures S7 and S8). The carbohydrates and the nucleobases (G or T) located in the opposite position remain intercalated between the surrounding base-pairs maintaining extensive contacts at both sides. As a consequence, a large number of NOE crosspeaks (see Table S3) between the linker and the carbohydrate protons with the DNA are observed. In all cases the double helices are slightly unwound and the rise between flanking residues is increased, as usually occurs in intercalation complexes.

The structures of helix 6dglc-G and helix 6dglc-T are deposited in the PDB (2N9F and 2N9H, respectively). 6dglc is well located opposite to both guanine (Figure 3b) and thymine (Figure 3f), thus orienting its alpha face towards its 3'-neighboring nucleobase and its methyl group towards the major groove. In the helix 6dglc-G the sugar-nucleobase pair forms two hydrogen bonds. In six of the ten resulting structures, these hydrogen bonds are H1 G–O4 6dglc and HN2 G–O3 6dglc (Figure 3b). The other structures show different conformers, including the pattern obtained in the quantum calculations (see below). All orientations are experimentally supported by a number of intra- and inter-strand NOEs, like those between the methyl group of 6dglc with H5 and H6 C18, with amino protons of C5, and with H1 G20 (Figure 3c). In the helix 6dglc-T the monosaccharide forms only one hydrogen bond with the opposing T. In most structures (7 out of 10), this hydrogen bond is H3 T–O4 6dglc (Figure 3f), and in the other cases the hydrogen bond is HO4 6dglc–O2 T. The orientation is supported by NOEs, for example, between the methyl group of 6dglc with methyl and H6 T7; with H1 and H8 G19; and with H5 and H6 C17 (Figure 3g). The formation of an extra hydrogen bond, when comparing 6dglc-G and 6dglc-T, may contribute to the enhanced thermal stability and it may explain the observed selectivity for purines.

The pairing properties of glc and 6dglc in the gas phase and water were investigated quantum chemically by means of

dispersion-corrected density functional theory (DFT) at the BLYP-D3(BJ)/TZ2P level of theory.[10] The binding energies (ΔE) of the Watson–Crick base pairs A-T and G-C agreed

Table 2: Hydrogen-bond energies (in kcal mol^{-1}) of sugar-base pairs in the gas-phase (ΔE_{gas}) and in aqueous solution (ΔE_{water}).[a]

X-Y	ΔE_{gas}	ΔE_{water}	X-Y	ΔE_{gas}	ΔE_{water}
A-T	−18.5	−9.4	G-C	−34.0	−13.5
6dglc-G	−23.8	−10.5	glc-G	−23.3	−12.2
6dglc-T	−10.5	−6.1	glc-T	−15.4	−9.5
6dglc-A	−16.7	−10.5	glc-A	−16.7	−10.7
6dglc-C	−12.9	−6.7	glc-C	−17.7	−5.1
6dglc-6dglc	−8.3	−4.1	glc-glc	−12.7	−8.7
6dglc-glc	−11.4	−7.6			

[a] Calculated at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution.

with those reported in literature (Table 2).[11] Sugar-purine pairs show greater hydrogen-bonding energy than any other combination for both for glc and 6dglc. In fact, the G-6dglc pair shows greater energy of bonding than a natural A-T base pair, with two hydrogen-bonds between the OH3 and OH4 of 6dglc with the NH1 and C=O, respectively, of G (Figure 4).

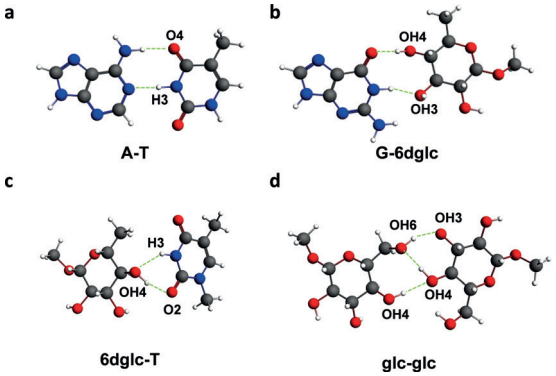


Figure 4. A-T, G-6dglc, 6dglc-T, and glc-glc pairs calculated at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution. Hydrogen bonds are shown in green.

Two hydrogen bonds are also observed for A-6dglc, G-glc, and A-glc pairs (see Figure S9). Differences in the hydrogen-bonding pattern in comparison to the NMR structures (such as in 6dglc-T) may be ascribed to factors such as base–base stacking or distance restrictions resulting from the DNA skeleton, and these are factors which are not included in these calculations.

Small differences between glc and 6dglc were observed when pairing with purines. In fact, the hydrogen bond formed between the OH4 of glc and the C=O of G is reinforced through a third hydrogen bond which is formed between the OH6 and OH4 (see Figure S9), thus yielding a better energy of bonding for glc-G than for 6dglc-G. Note that glc and 6dglc form significantly more stable pairs with pyrimidines than with purines. The only exception is the glc-T pair which has a bond energy of −9.55 kcal mol^{-1}.

Interestingly, sugar-sugar base pairs show similar bond energies to that of sugar-pyrimidine base pairs. One, two, and three hydrogen bonds are formed in 6dglc-6dglc, 6dglc-glc, and glc-glc, respectively, where the OH groups involved can act as either hydrogen-bond donors or acceptors (Figure 4d). In contrast to the experimental data found for DNA, the 6dglc-6dglc pair appears to be less stable than the glc-glc base pair. These differences may be ascribed to interactions with the surrounding bases,[11] which are not considered in the present calculations.

In conclusion, we have shown glycoside-nucleobase pairs form within a DNA double helix. They cause some destabilization of either a DNA or GNA duplex but they display selective pairing with purines versus pyrimidines. This selectivity can be explained by the formation of hydrogen bonds between either glc or 6dglc with the natural bases as shown by quantum chemical calculations and NMR studies. Moreover, 6dglc stacks within the interior of a duplex when paired with either G or T and does not stick out of the helix seeking better hydration. These 6dglc-G and 6dglc-T pairs infer only minor distortion in the double-helix structure.

Acknowledgments

We thank the Ministerio de Economía y Competitividad (CTQ2011-15203-E, CTQ2012-35360, CTQ2014-52588-R, BFU2014-52864-R), the Netherlands Organization for Scientific Research (NWO-CW and NWO-EW), and the National Research School Combination—Catalysis (NRSC-C) for financial support. E.V.C. thanks Ministerio de Educación, Cultura y Deporte for a FPU fellowship and Cost Action CM1005 for a STSM grant.

Keywords: DNA · hydrogen bonds · NMR spectroscopy · noncovalent interactions · nucleobases

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 8643–8647
Angew. Chem. **2016**, *128*, 8785–8789

[1] A. Blond, E. Ennifar, C. Tisne, L. Micouin, *ChemMedChem* **2014**, *9*, 1982–1996.
[2] a) R. Moumné, M. Catala, V. Larue, L. Micouin, C. Tisné, *Biochimie* **2012**, *94*, 1607–1619; b) S. P. Velagapudi, B. R. Vummidi, M. D. Disney, *Curr. Opin. Chem. Biol.* **2015**, *24*, 97–103.
[3] Q. Vicens, E. Westhof, *Biopolymers* **2003**, *70*, 42–57.
[4] a) R. Lucas, I. Gómez-Pinto, A. Aviñó, J. J. Reina, R. Eritja, C. González, J. C. Morales, *J. Am. Chem. Soc.* **2011**, *133*, 1909–1916; b) I. Gómez-Pinto, E. Vengut-Climent, R. Lucas, A. Aviñó, R. Eritja, C. Gonzalez, J. C. Morales, *Chem. Eur. J.* **2013**, *19*, 1920–1927; c) R. Lucas, P. Penalver, I. Gomez-Pinto, E. Vengut-Climent, L. Mtashobya, J. Cousin, O. S. Maldonado, V. Perez, V. Reynes, A. Aviñó, R. Eritja, C. Gonzalez, B. Linclau, J. C. Morales, *J. Org. Chem.* **2014**, *79*, 2419–2429.
[5] M. Kaufmann, M. Gisler, C. J. Leumann, *Angew. Chem. Int. Ed.* **2009**, *48*, 3810–3813; *Angew. Chem.* **2009**, *121*, 3868–3871.
[6] H. Kashida, K. Sekiguchi, H. Asanuma, *Chem. Eur. J.* **2010**, *16*, 11554–11557.
[7] a) K. C. Schneider, S. A. Benner, *J. Am. Chem. Soc.* **1990**, *112*, 453–455; b) D. Zhou, I. M. Lagoja, J. Rozenski, R. Busson, A.

Van Aerschot, P. Herdewijn, *ChemBioChem* **2005**, *6*, 2298 – 2304.

[8] a) P. Cuniasse, L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, J. A. Cognet, M. LeBret, W. Guschlbauer, G. V. Fazakerley, *Nucleic Acids Res.* **1987**, *15*, 8003 – 8022; b) P. Cuniasse, L. C. Sowers, R. Eritja, B. Kaplan, M. F. Goodman, J. A. Cognet, M. Le Bret, W. Guschlbauer, G. V. Fazakerley, *Biochemistry* **1989**, *28*, 2018 – 2026.

[9] a) L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174 – 4175; b) M. K. Schlegel, A. E. Peritz, K. Kittigowittana, L. Zhang, E. Meggers, *ChemBioChem* **2007**, *8*, 927 – 932.

[10] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931 – 967.

[11] a) J. Poater, M. Swart, C. Fonseca Guerra, F. M. Bickelhaupt, *Chem. Commun.* **2011**, *47*, 7326 – 7328; b) J. Poater, M. Swart, F. M. Bickelhaupt, C. Fonseca Guerra, *Org. Biomol. Chem.* **2014**, *12*, 4691 – 4700.

Received: April 11, 2016
Revised: May 19, 2016
Published online: June 22, 2016

