

“Glucose-nucleobase pseudo base pairs as a new biomolecular interaction in a DNA context”
Vengut-Climent, E., Gómez-Pinto, I., Lucas, R., Peñalver, P., Aviñó, A., Fonseca-Guerra, C., Bickelhaupt, F.M., Eritja, R., González, C., Morales, J.C. *Angew. Chem. Int. Ed.*, 55(30), 8643-8647, (2016). doi: 10.1002/anie201603510

Glucose–Nucleobase Pseudo Base Pairs: Biomolecular Interactions within DNA

Empar Vengut-Climent,^a Irene Gómez-Pinto,^b Ricardo Lucas,^{a,c} Pablo Peñalver,^{a,c} Anna Aviñó,^d Célia Fonseca Guerra,^e F. Matthias Bickelhaupt,^{e,f} Ramón Eritja,^d Carlos González,^b and Juan C. Morales*^{a,c}

^aDepartment of Bioorganic Chemistry, Instituto de Investigaciones Químicas CSIC—Universidad de Sevilla, Américo Vespucio 49, 41092 Sevilla (Spain)
E-mail: jcmorales@ipb.csic.es

^bInstituto de Química Física “Rocasolano”, CSIC 28006 Madrid (Spain)

^cDepartment of Biochemistry and Molecular Pharmacology, Instituto de Parasitología y Biomedicina, CSIC, Parque Tecnológico, Ciencias de la Salud 18016 Armilla, Granada (Spain)

^dInstituto de Química Avanzada de Cataluña (IQAC), CSIC, CIBER—BBN Networking Centre on Bioengineering, Biomaterials and Nanomedicine, 08034 Barcelona (Spain),

^eDepartment of Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Vrije Universiteit Amsterdam 1081 HV Amsterdam (The Netherlands),

^fInstitute of Molecules and Materials (IMM), Radboud University, 6525 AJ Nijmegen (The Netherlands).

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]

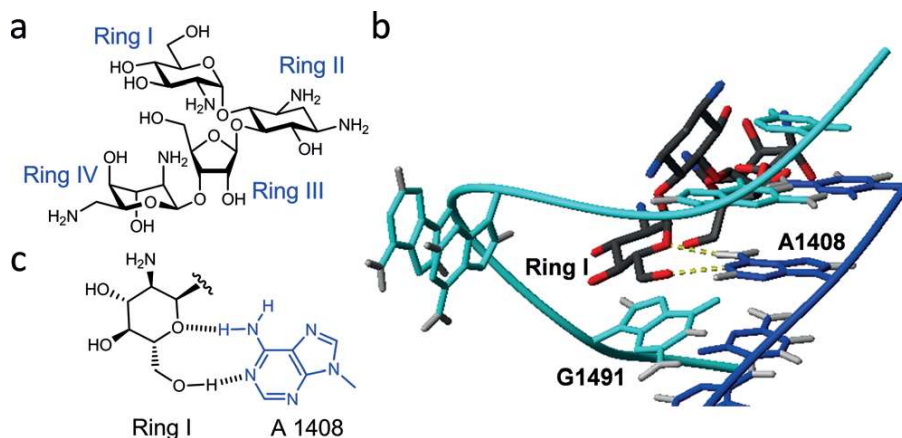


Figure 1. a) Structure of paromomycin. b) Detail of the solution structure of paromomycin binding a 16S RNA model sequence. c) Drawing of the glycoside-adenine 4108 pseudo base pair.

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 kcalmol⁻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 paromomycin 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 glucosenucleobase 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]

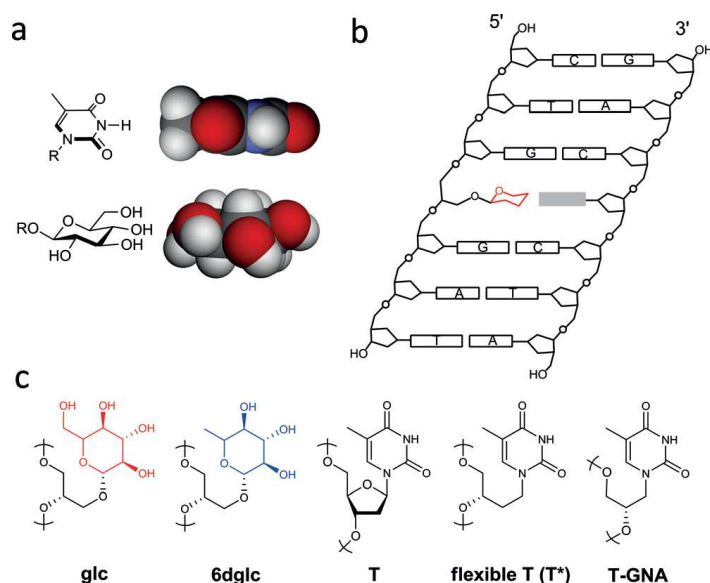
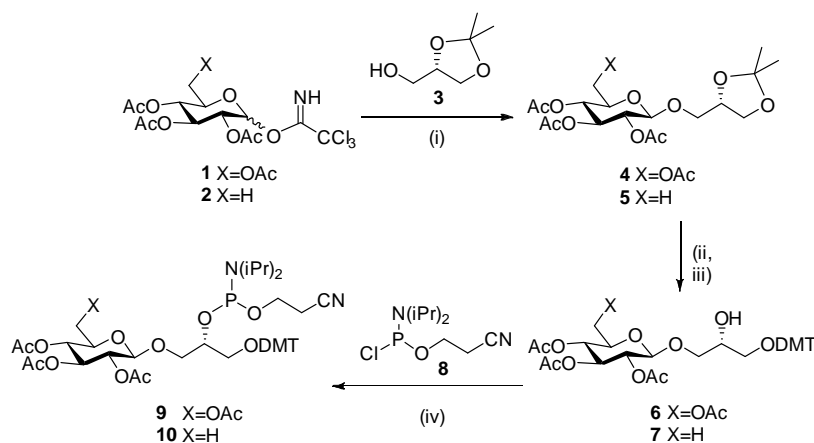


Figure 2. Description of the carbohydrate derivatives under study. a) CPK models of thymine and glucose. b) Schematic drawing of a DNA double helix containing a glycoside-nucleobase pair. c) Structures of the two sugar units prepared, (*S*)-2,3-dihydroxypropyl glucose (glc) and (*S*)-2,3-dihydroxypropyl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

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).



Scheme 1. Synthesis of the carbohydrate DMT phosphoramidites. Glycosyl trichloroacetimidate donors were used as starting material. Reagents and conditions: (i) TMSOTf, CH₂Cl₂ for **glc**; BF₃·OEt₂, CH₂Cl₂ for **6dglc**; (ii) AcOH-H₂O, 80 °C; (iii) DMTCl, DMAP,

CH₂Cl₂; (iv) DIPEA, CH₂Cl₂. DIPEA=diisopropylethylamine, DMAP=4-(N,Ndimethylamino)pyridine, DMT=4,4'-dimethoxytrityl, Tf=trifluoromethanesulfonyl, TMS=trimethylsilyl.

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

DNA duplexes		5'-d(GATGACXGCTAG) ^[a,b]			
		3'-d(CTACTGYCGATC)			
X-Y	T_m (°C)	X-Y	T_m (°C)	X-Y	T_m (°C)
T*-A	37.7	alc-A	30.6	6dalc-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	alc-	29.6	6dalc-6dalc	32.0
GNA duplexes		3'-TAAAATTTAXATTATTAA ^[b,c]			
		2'-ATTTTAAATYTAATAATT			
X-Y	T_m	X-Y	T_m (°C)		
T-A	57.3	A-6dalc	45.4		
T-T	48.0	T-6dalc	37.4		
		6dalc-T	38.7		
		6dalc-6dalc	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 NaH₂PO₄, 150 mM NaCl, pH 7.0. For GNA duplexes: 10 mM NaH₂PO₄, 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.

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 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

sugarpyrimidine 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 δ =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 δ =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).

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.

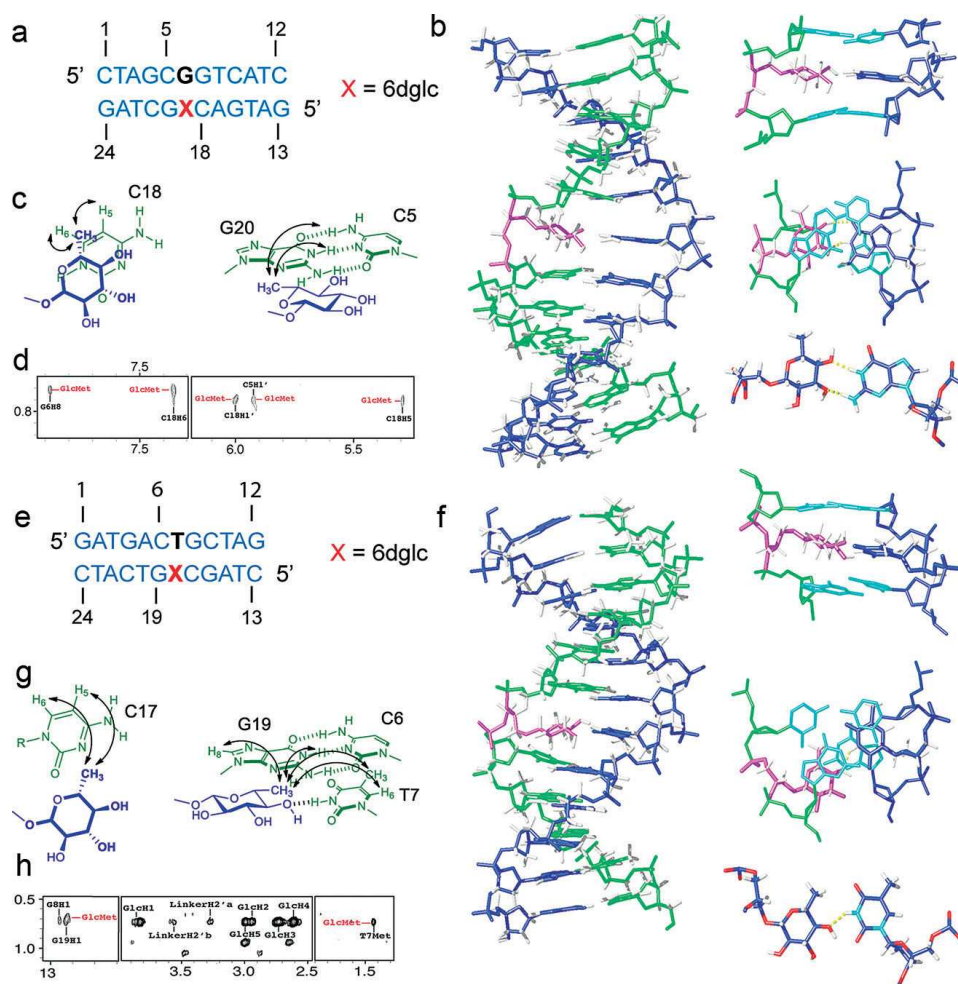


Figure 3. Solution structure of the helix 6dglc-G and helix 6dglc-T. a) Sequences of the helix 6dglc-G. b) Refined solution structure of the helix 6dglc-G and zoom views of the 6dglc-G pair. c) NOE interactions between 6dglc and its 5'-neighboring base C18 and its 3'-neighboring base pair G20-C5. d) 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 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 interstrand 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 3 f), 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 (DE) of the Watson–Crick base pairs A-T and G-C agreed 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).

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	$\square E_{\text{gas}}$	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	-9.1	-5.0	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.

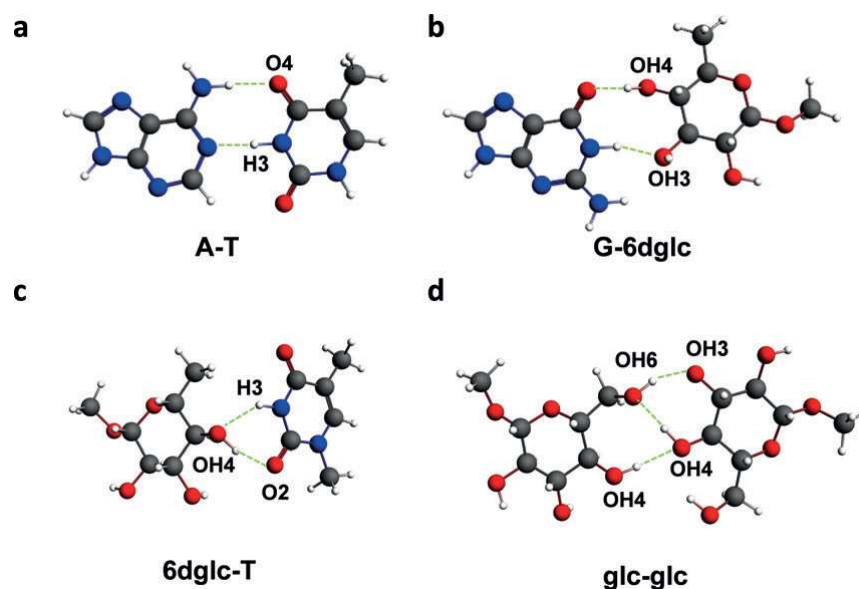


Figure 4. A-T, G-6dglc, 6dglc-T, and glc-glc pairs calculated at BLYPD3(BJ)/TZ2P using COSMO to simulate aqueous solution. Hydrogen bonds are shown in green.

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 $-9.55 \text{ kcal mol}^{-1}$.

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

- [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. Tisé, *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. González, J. C. Morales, *Chem. Eur. J.* 2013, 19, 1920 – 1927; c) R. Lucas, P. Peñalver, I. Gomez-Pinto, E. Vengut-Climent, L. Mtashobya, J. Cousin, O. S. Maldonado, V. Perez, V. Reynes, A. Aviñó, R. Eritja, C. González, 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. Cuniassé, 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. Cuniassé, 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.

Glucose-nucleobase pseudo base pairs as a new binding motif in a DNA context

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

Supplementary Information

Contents:

Supplementary materials and methods

- Synthesis materials and methods	3
- Synthesis of T* , glc and 6dglc phosphoramidites	3-20
- ¹ H and ¹³ C-NMR spectra of new compounds	3-20
- Synthesis of natural and modified oligonucleotide DNA strands	20
- HPLC chromatograms of modified oligonucleotide DNA strands.	21-22
- Maldi-TOF mass of modified oligonucleotide DNA strands.	22
- Synthesis of oligonucleotide GNA strands	23
- HPLC chromatograms of oligonucleotide GNA strands.	24
- Maldi-TOF mass of oligonucleotide GNA strands.	24
- Thermal denaturation methods	25
- NMR methodology	25
- Structure Calculations	26
- DFT quantum chemical methods	26

Supplementary Figures

- Figure S1: Synthesis of flexible T* amidite.	27
- Figure S2: Melting curves for DNA double helices containing glc and 6dglc .	28
- Figure S3. Melting curves for GNA double helices containing 6dglc .	29
- Figure S4. Imino region of the NMR spectra of helix 6dglc-G and helix 6dglc-T .	30
- Figure S5. More significant changes in proton chemical shifts along the sequence for helix 6dglc-G and helix 6dglc-T ,	30-31
- Figure S6. Several regions of NOESY spectra of helix 6dglc-T	32
- Figure S7. Solution structure of helix 6dglc-T	33
- Figure S8. Solution structure of helix 6dglc-G	34
- Figure S9. Structures for pairs containing glc and 6dglc in <i>vacuo</i> and in water, computed at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution.	

Supplementary Tables

- Table S1: ¹H-NMR assignments of **helix 6dglc-T**. 36
- Table S2: ¹H-NMR assignments of **helix 6dglc-G**. 37
- Table S3. Structurally relevant carbohydrate-DNA NOE contacts for **helix 6dglc-G** and **helix 6dglc-T**. 38
- Table S4. NMR structural constraints and calculation statistics. 39
- Table S5. Cartesian coordinates and ADF total energies of all stationary points in this study, computed at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution. 39-55

Supplementary references 55-56

Synthesis materials and methods

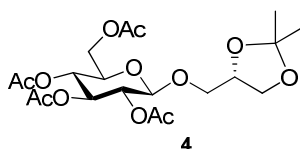
All chemicals were obtained from chemical suppliers (Sigma-Aldrich, Carbosynth) and used without further purification, unless otherwise noted. Anhydrous dichloromethane was dried in an aluminium oxide column machine, PURESOLV (Scharlab). Other anhydrous solvents were dried over molecular sieves (4 Å) for 48 h.

All reactions were monitored by TLC on precoated silica gel 60 plates F₂₅₄ (Merck) and detected by heating after staining with H₂SO₄:EtOH (1:9, v/v), anisaldehyde (450 ml ethanol, 25 ml anisaldehyde, 25 ml H₂SO₄ and 1 ml AcOH) or Mostain (500 ml of 10% H₂SO₄, 25g of (NH₄)₆Mo₇O₂₄•4H₂O, 1g Ce(SO₄)₂•4H₂O). Products were purified by flash chromatography with silica gel 60 (200-400 mesh). Eluents are indicated for each particular case.

NMR spectra were recorded on either a Bruker Avance 300 or ARX 400 MHz or Bruker Avance DRX 500 MHz [300, 400 or 500 MHz (¹H), 75, 100, 125 MHz (¹³C), at room temperature for solutions in CDCl₃ or CD₃OD] spectrometer. Chemical shifts are referred to the solvent signal and are expressed in ppm. 2D NMR experiments (COSY, TOCSY and HMQC) were carried out when necessary to assign the corresponding signals of the new compounds. Low resolution electrospray mass spectral analyses were obtained on a Bruker Esquire 6000 ion trap mass spectrometer. High resolution FAB (+) mass spectral analyses were obtained on a Micromass AutoSpec-Q spectrometer.

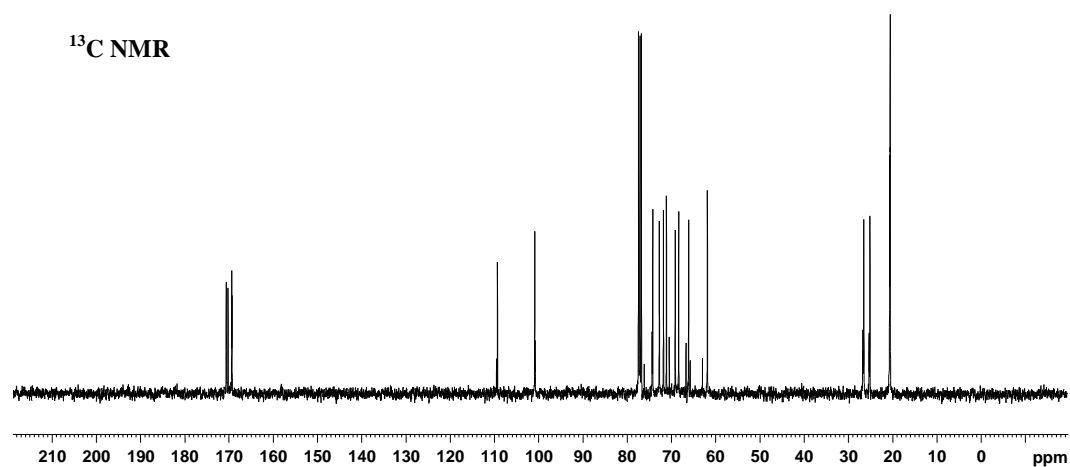
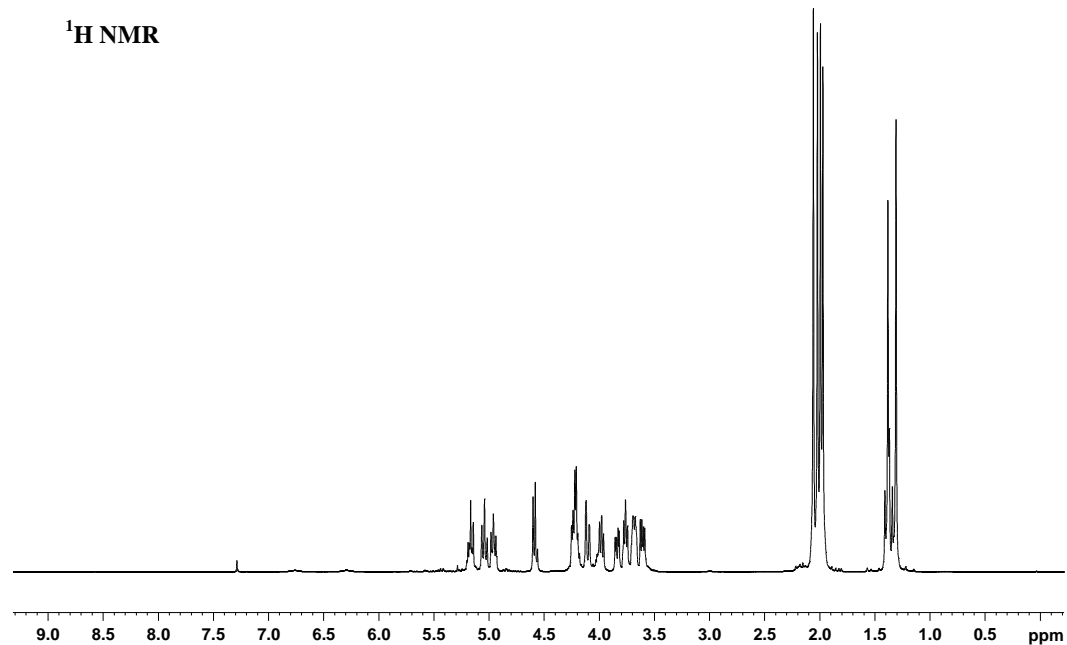
Synthesis of T*, glc and 6dglc phosphoramidites

(S)-2,2-dimethyl-4-(2,3,4,6-tetra-O-acetyl-β-D-glucopyranos-1-ylmethyl)dioxolane (4)

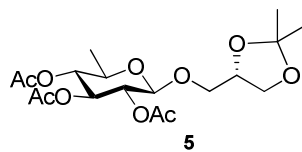


To a solution of 2,3,4,6-tetra-*O*-acetyl- α,β -D-glucopyranosyl trichloroacetimidate **1** (3 g, 6.08 mmol) and (*S*)-(+)-1,2-isopropylidenglycerol **3** (2.26 ml, 18.26 mmol) in anhydrous CH₂Cl₂ (40 ml) BF₃·OEt₂ (77 μ l, 0.609 mmol) was added. The reaction was then stirred for 30 min and NEt₃ (0.5 ml) was then added. The solvent were removed and the crude was purified by flash column chromatography (Hex/EtOAc, from 2:1 to 1:2) to afford **4** (2415 mg, 86%); ¹H NMR (400 MHz, CDCl₃) δ (ppm): 5.17 (t, *J* = 9.44 Hz, 1H, H₃), 5.04 (t, *J* = 9.60 Hz, 1H, H₄), 4.96 (t, *J* = 9.20 Hz, 1H, H₂), 4.59 (d, *J* = 7.96 Hz, 1H, H₁), 4.25-4.19 (m, 2H, H₆, -OCHH-isopropylidene), 4.10 (dd, *J* = 1.6/12.0 Hz, 1H, H_{6'}), 3.99-3.96 (m, 1H, -CH-isopropylidene), 3.84 (dd, *J* = 4.4/10.8 Hz, 1H, -OCHH-isopropylidene), 3.78-3.74 (m, 1H, -OCHH-), 3.70-3.68 (m, 1H, H₅),

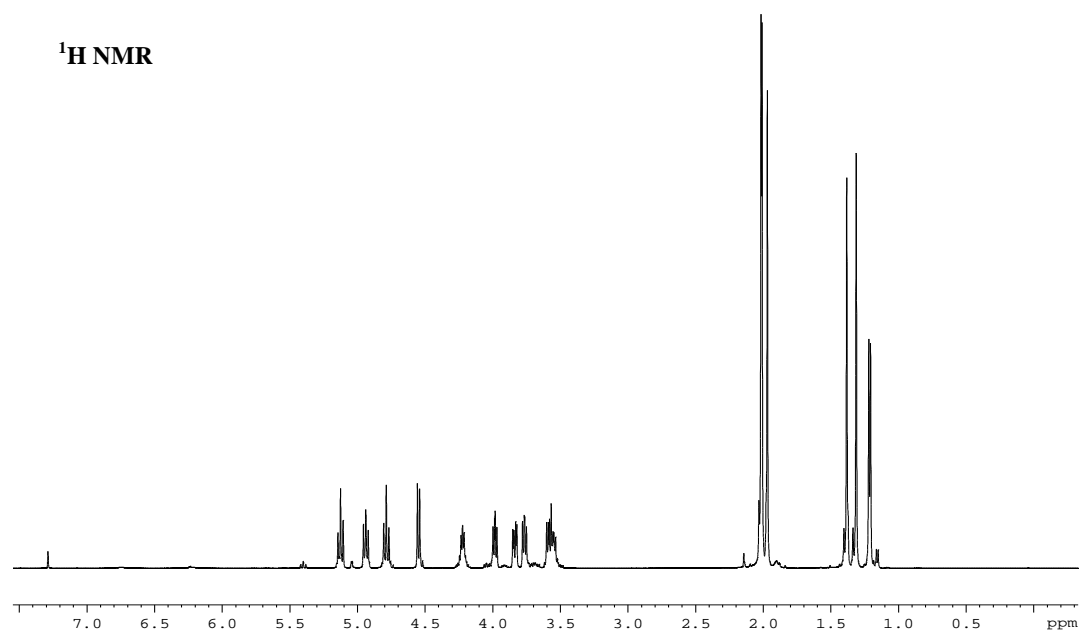
3.63-3.60 (dd, $J = 5.8/10.6$ Hz, 1H, -OCHH-), 2.06, 2.02, 2.00, 1.98 (4s, 12H, 3x-OCOCH₃), 1.41, 1.33 (2s, 6H, 2x-C(CH₃)₂); ¹³C NMR (100 MHz, CDCl₃): $\delta = 170.6, 170.2, 169.4, 169.3$ (CO), 109.3 (-C(CH₃)₂), 100.9 (C₁), 74.3, 72.7, 71.8, 71.1, 69.1, 68.4, 66.1, 61.9, 26.7, 25.1, 20.7, 20.6, 20.5. HRMS (FAB⁺) Calcd. for C₂₀H₃₀NaO₁₂ (M+Na): 485.1635, found; 485.1622.

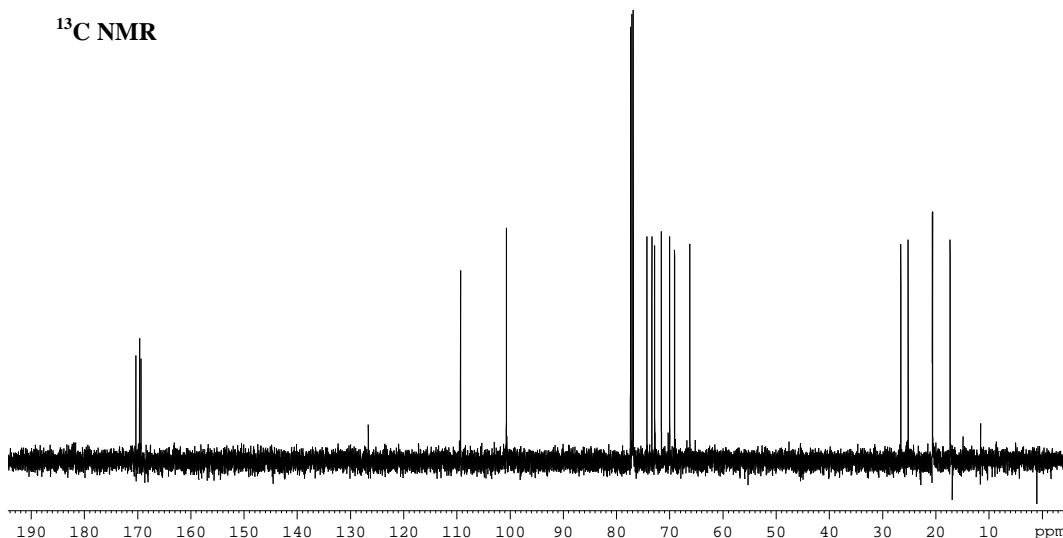
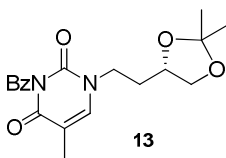


(*S*)-2,2-dimethyl-4-(2,3,4-tri-*O*-acetyl-6-deoxy- β -D-glucopyranos-1-ylmethyl)dioxolane (**5**)



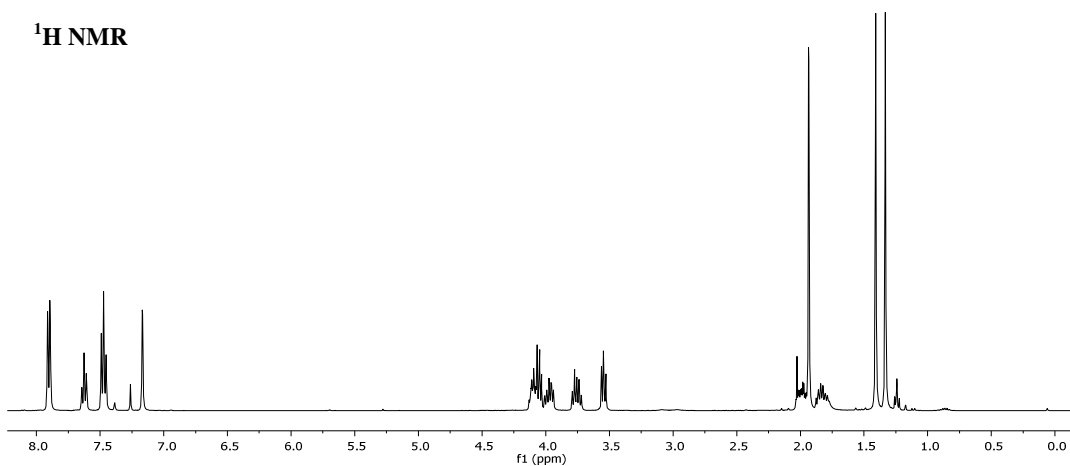
To a solution of 2,3,4-tri-*O*-acetyl-6-deoxyl- β -D-glucopyranosyl trichloroacetimidate **1** (900 mg, 2.07 mmol) and (*S*)-(+)-1,2-isopropylidenglycerol **3** (388 μ l, 3.10 mmol) in anhydrous CH_2Cl_2 (10 ml) $\text{BF}_3\cdot\text{OEt}_2$ (30 μ l, 0.21 mmol) was added. The reaction was stirred for 30 min and NEt_3 (0.5 ml) was then added. The solvent was removed and the crude was purified by flash column chromatography (Hex/EtAcO, from 3:1 to 1:1) to afford **5** (500 mg, 60%); ^1H NMR (500 MHz, CDCl_3) δ (ppm): 5.13 (t, $J=9.5$ Hz, 1H, H_3), 4.94 (t, $J=8.0$ Hz, 1H, H_2), 4.78 (t, $J=9.5$ Hz, 1H, H_4), 4.54 (d, $J=8.0$ Hz, 1H, H_1), 4.24-4.20 (m, 1H, $-\text{CH}_{\text{-isopropylidene}}$), 3.99-3.97 (m, 1H, $-\text{OCHH-}$), 3.84 (dd, $J=4.5/10.5$ Hz, 1H, $-\text{OCHH-}$), 3.76 (dd, $J=6.0/8.0$ Hz, 1H, $-\text{OCHH-}$), 3.60-3.53 (m, 2H, H_5 , $-\text{OCHH-}$), 2.01, 2.00, 1.97 (3s, 9H, $3\times\text{-OCOCH}_3$), 1.38, 1.31 (2s, 6H, $-\text{C}(\text{CH}_3)_2$), 1.22-1.20 (m, 3H, $-\text{CH}_3$); ^{13}C NMR (125 MHz, CDCl_3) δ (ppm): 170.3, 169.6, 169.3 (CO), 109.3 ($\text{C}(\text{CH}_3)_2$), 100.7 (C_1), 74.3, 73.3, 72.8, 71.5, 69.9, 69.0, 66.2, 26.6, 25.2, 20.6, 17.3. HRMS (FAB^+) Calcd. for $\text{C}_{18}\text{H}_{30}\text{O}_{10}\text{Na}$ ($\text{M}+\text{Na}$): 427.1580, found; 427.1588.



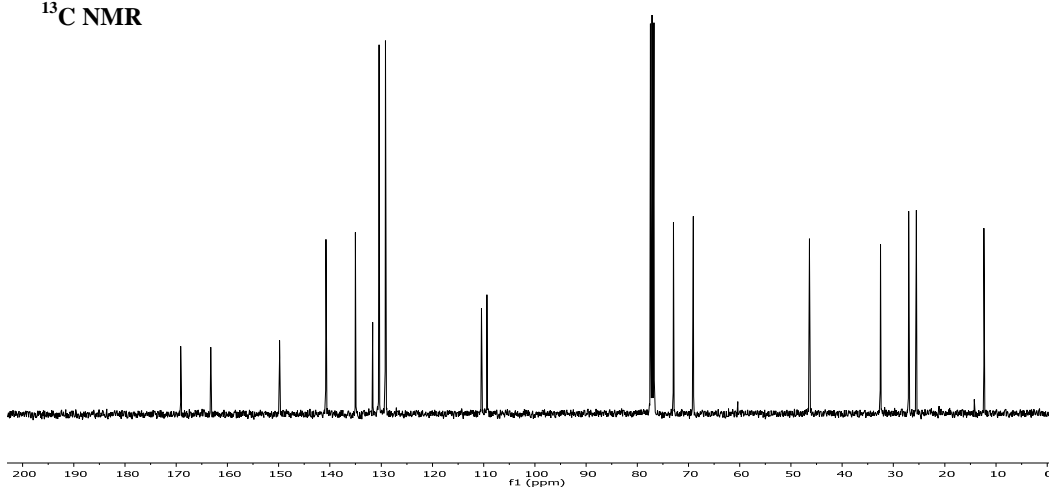
¹³C NMR**(S)-2,2-dimethyl-4-(N-benzoylthymine-1-ylethyl)dioxolane (13)**

To a solution of the benzoylated thymine **11**¹ (600 mg, 2.61 mmol) in dry DMF (10 ml), K₂CO₃ (396 mg, 2.87 mmol) and TBAI (96 mg, 0.261 mmol) were added. Then, (S)-toluene-4-sulfonic acid 2-(2,2'-dimethyl-[1,3]dioxolan-4-yl)-ethyl ester **12**² was added and the reaction mixture was stirred at 70 °C for 2 hours and finally at room temperature overnight. The reaction mixture was washed with water and saturated NH₄Cl, and extracted with ethyl acetate and ether. Solvents were removed and the crude was purified by silica gel column chromatography (Hex/EtOAc from 3:1 to 0:1) to give compound **13** (498 mg, 53 %) as a white foam. ¹H NMR (400 MHz, CDCl₃) δ (ppm): 7.90 (d, *J* = 7.2 Hz, 2H, H_{benzoyl}), 7.62 (t, *J* = 7.2 Hz, 1H, H_{benzoyl}), 7.47 (t, *J* = 7.2 Hz, 2H, H_{benzoyl}), 7.17 (s, 1H, H_{thymine}), 4.13-4.03 (m, 2H, -OCHH_{-isopropylidene}, -CH_{-isopropylidene}), 4.00-3.94 (m, 1H, -CHH-CH₂N-), 3.79-3.72 (m, 1H, -CHH-CH₂N-), 3.56-3.53 (m, 1H, -CHH_{-isopropylidene}), 2.02-1.95 (m, 1H, -CH₂-CHH-CH-O-), 1.93 (s, 3H, -CH₃_{thymine}), 1.87-1.79 (m, 1H, -CH₂-CHH-CH-O-), 1.40 (s, 3H, -CH₃_{isopropylidene}), 1.33 (s, 3H, -CH₃_{isopropylidene}); ¹³C NMR (100 MHz, CDCl₃) δ (ppm): 169.1, 163.2, 149.8, 140.9, 135.0, 131.7, 130.4, 129.1, 110.4, 109.3, 76.8, 72.9, 69.1, 46.4, 32.5, 27.0, 25.5, 12.3. HRMS (ES⁺) Calcd. for C₁₆H₁₈N₂O₅: 358.1529, found; 358.1518.

¹H NMR

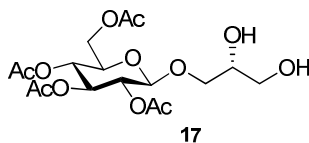


¹³C NMR



General procedure for isopropylidene removal. A solution of the 2,2-dimethyldioxolane derivative (1.93 mmol) was dissolved in 40 ml of acetic acid 80% and the solution was stirred at 80 °C for 1 to 2 hours. The solvent was removed and coevaporated with toluene. The reaction crude was then purified by flash chromatography.

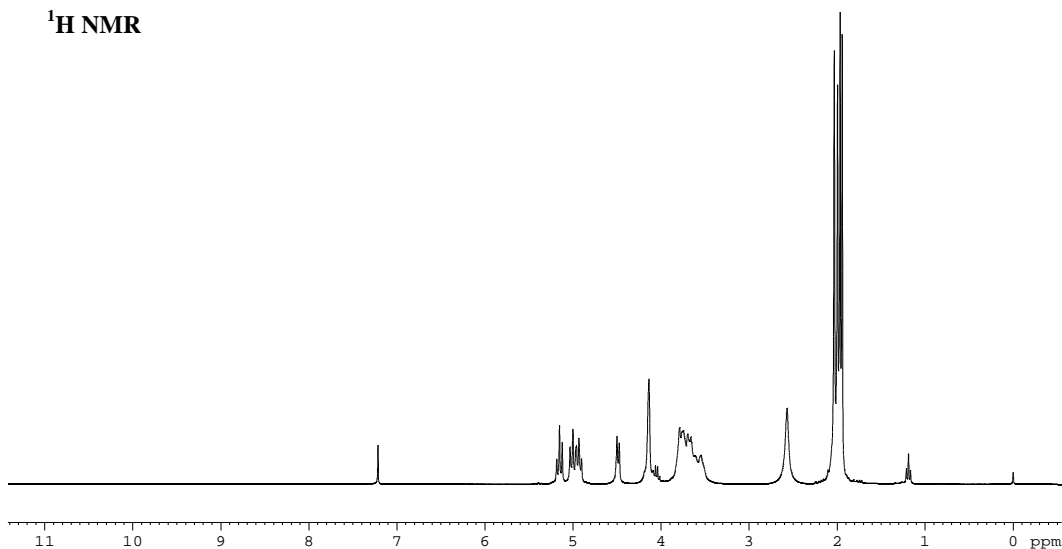
(R)-2,3-dihydroxypropyl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranoside (17)



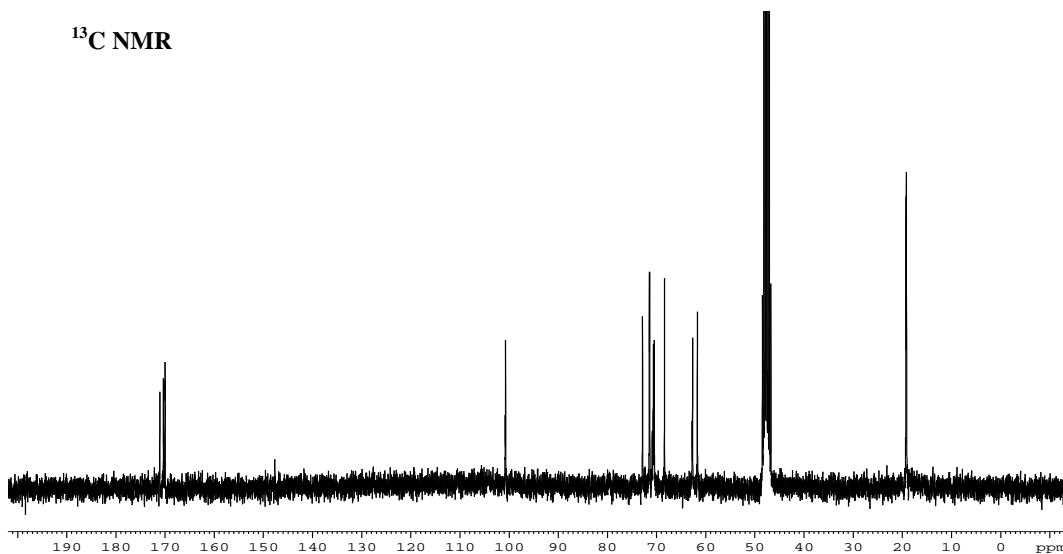
Compound **4** (3.5 g, 7.56 mmol) was dissolved in a mixture of CH₃COOH/H₂O (150 ml, 4;1). The reaction mixture was reacted following the general isopropylidene removal procedure. The

crude was purified by silica gel column chromatography (EtOAc/MeOH, from 1:0 to 10:1) to give **17** (2.8 g, 87%) as a syrup. ^1H NMR (400 MHz, CDCl_3) δ (ppm): 5.15 (t, $J=9.3$ Hz, 1H, H_3), 5.02 (t, $J=9.5$ Hz, 1H, H_4), 4.93 (t, $J=9.3$ Hz, 1H, H_2), 4.48 (d, $J=7.8$ Hz, 1H, H_1), 4.13 (s, 2H, H_6, H_6'), 3.79-3.55 (m, 6H, $-\text{CH}-, 2\times-\text{OCH}_2-$, H_5), 2.57 (br.s, 1H, OH), 2.03, 2.00, 1.97, 1.94 (4s, 12H, 3x- OCOCH_3); ^{13}C NMR (101 MHz, CD_3CD) δ (ppm): 171.0, 170.3, 170.0, 169.9 (CO), 100.8 (C_1), 72.8, 71.5, 71.4, 70.6, 70.5, 68.4, 62.8, 62.7, 61.7, 19.3, 19.2. HRMS (FAB $^+$) Calcd. for $\text{C}_{17}\text{H}_{26}\text{NaO}_{12}$ ($\text{M}+\text{Na}$): 445.1322, found; 445.1329.

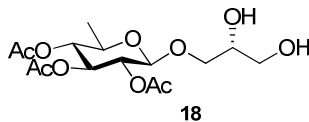
^1H NMR



^{13}C NMR

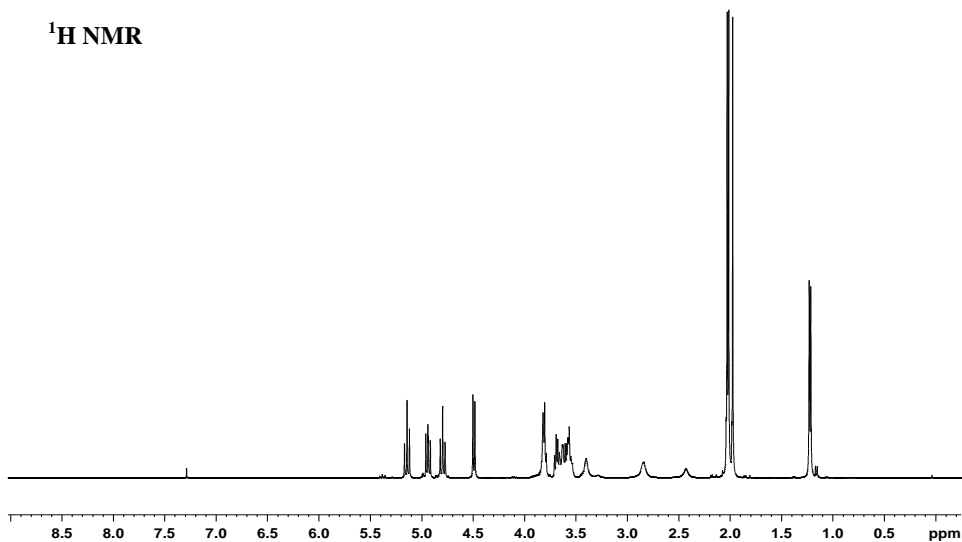


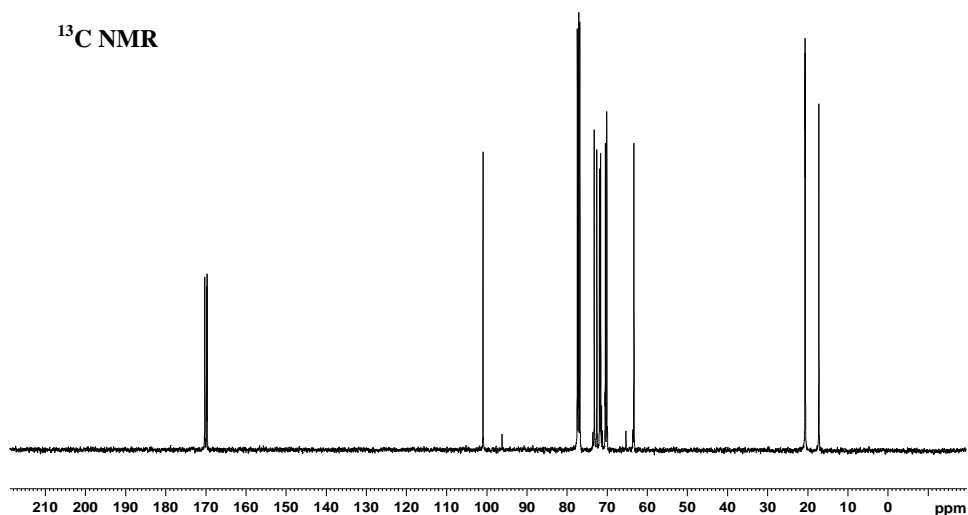
(R)-2,3-dihydroxypropyl 2,3,4-tri-O-acetyl-6-deoxy-β-D-glucopyranoside (18)



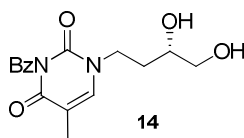
Compound **5** (1 g, 2.45 mmol) was dissolved in a mixture of CH₃COOH (18 ml) and H₂O (2 ml). The reaction mixture was reacted following the general isopropylidene removal procedure. The crude was purified by silica gel column chromatography using as eluent (EtOAc:MeOH, 1:0-10:1) to give **18** (599 mg, 68%) as a syrup. ¹H NMR (CDCl₃, 400 MHz,) δ (ppm): 5.14 (t, *J*=9.6 Hz, 1H, H₃), 4.94 (dd, *J*=8.0/9.6 Hz, 1H, H₂), 4.79 (t, *J*=9.6 Hz, 1H, H₄), 4.49 (d, *J*=8.0 Hz, 1H, H₁), 3.85-3.79 (m, 2H, -OCH₂-), 3.72-3.52 (m, 4H, H₅, -CH-, -OCH₂-), 3.40 (br.s, 1H, OH), 2.84 (br. s, 1H, OH), 2.03, 2.01, 1.98 (3s, 9H, 3 x -OCOCH₃), 1.21 (m, 3H, CH₃); ¹³C NMR (CDCl₃, 125 MHz) δ (ppm): 170.3, 169.8, 169.7 (CO), 100.9 (C₁), 73.2, 72.6, 71.9, 71.6, 70.4, 70.0, 63.4, 20.7, 20.6, 17.2. HRMS (FAB⁺) Calcd. for C₁₅H₂₄NaO₁₀ (M+Na): 387.1267, found; 387.1257.

¹H NMR

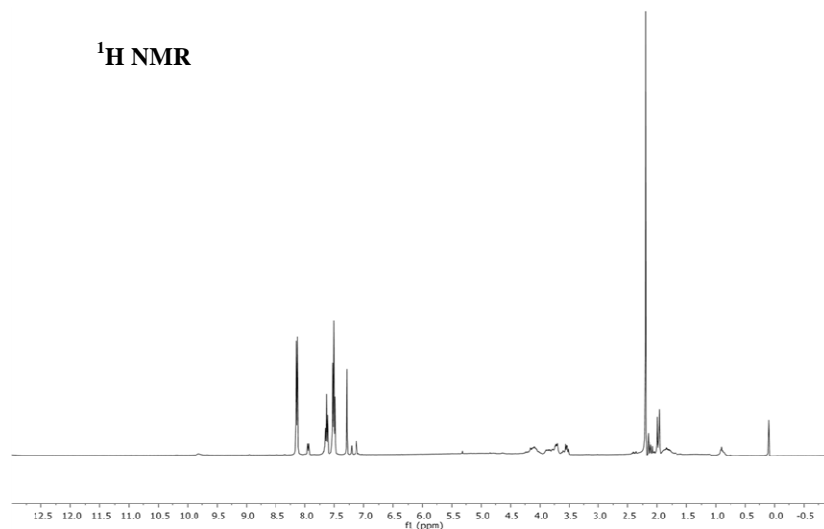


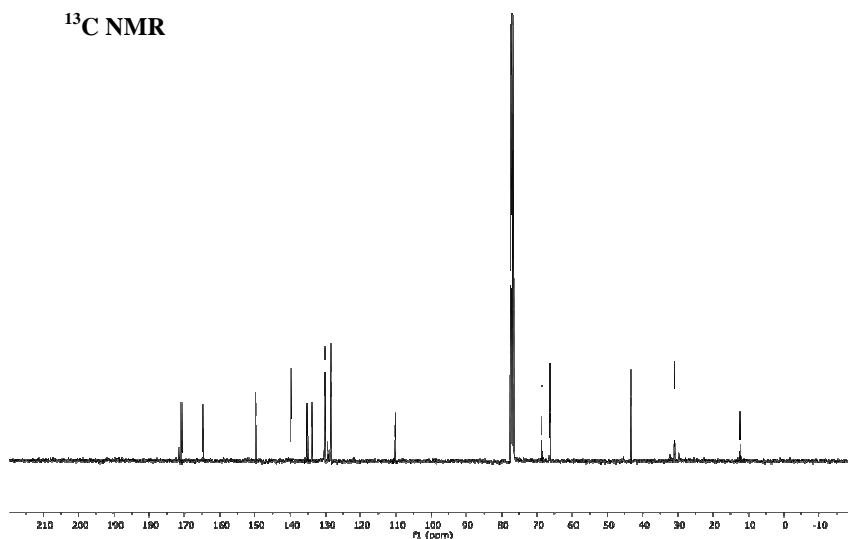


(S)-1-(3,4-dihydroxybutyl)-N-benzylthymine (14)



A solution of compound **13** (1.45 g, 1.93 mmol) was dissolved in 40 ml of acetic acid 80% and was reacted following the general isopropylidene removal procedure. Flash chromatography was used (Hex/EtOAc from 1:6 to 0:1 and EtOAc/MeOH from 20:1 to 10:1) to give compound **14** (72%) as a white solid. ¹H NMR (400 MHz, CDCl₃) δ (ppm): 7.91 (d, *J* = 7.2 Hz, 2H, H_{benzoyl}), 7.68-7.64 (m, 1H, H_{benzoyl}), 7.52-7.48 (m, 2H, H_{benzoyl}), 7.16 (s, 1H, H_{thymine}), 4.16-4.03 (m, 1H, -OCHH-), 3.88-3.70 (m, 2H, -OCHH-, -CH_{isopropylidene}-), 3.60-3.51 (m, 2H, -CH₂-CH₂N-), 2.20 (s, 3H, -CH₃_{thymine}), 2.15-1.75 (m, 2H, -NCH₂CH₂O-); ¹³C NMR (100 MHz, CDCl₃) δ (ppm): 172.2, 165.0, 149.8, 140.9, 136.1, 132.7, 130.4, 129.1, 110.4, 68.4, 66.3, 43.4, 31.5, 12.3. HRMS (FAB⁺) Calcd. for C₁₆H₁₈NaO₂ (M+Na): 341.1113, found: 341.1109.

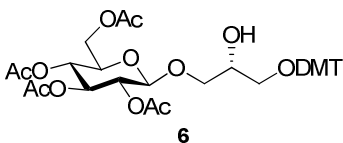




General procedure for primary hydroxyl protection with 4,4-dimethoxytritylgroup.

DIPEA (1.941 mmol) and DMAP (0.097 mmol) were added to a solution of the unprotected compound (0.971 mmol) in pyridine/CH₂Cl₂ (1:1, 10 ml). The temperature was lowered to 0 °C and DMTC1 (1.456 mmol) was slowly added. The reaction was stirred for 15 minutes at 0 °C and 3 h at room temperature. Then, reaction was stopped by adding MeOH and the solvent was removed in *vacuo*. The reaction mixture was purified by flash chromatography.

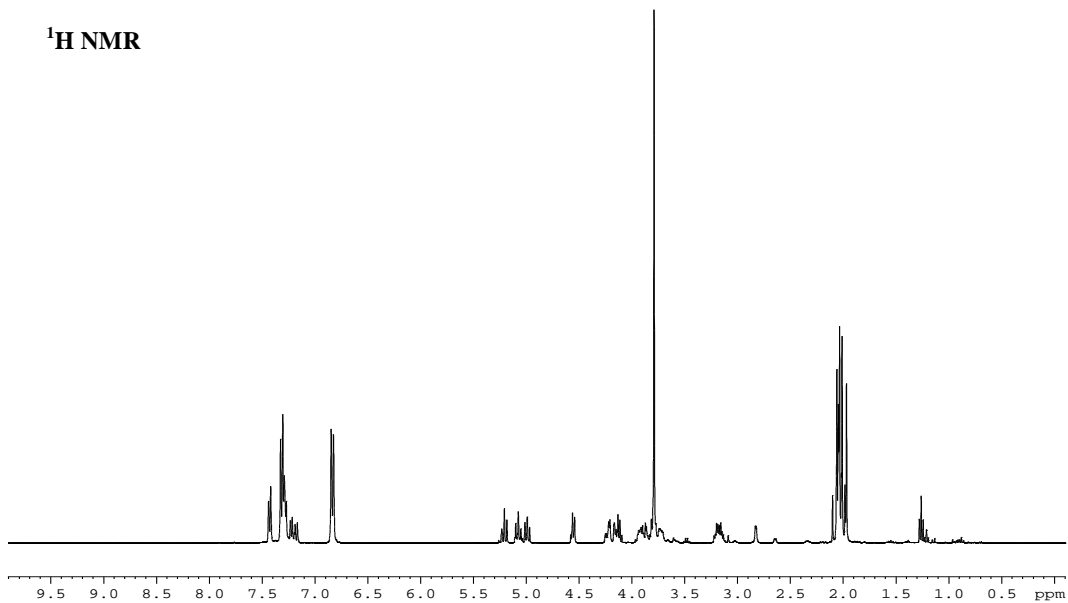
(R)-3-(4,4'-dimethoxytrityloxy)-2-hydroxypropyl 2,3,4,6-tetra-O-acetyl-β-D-glucopyranoside (6)



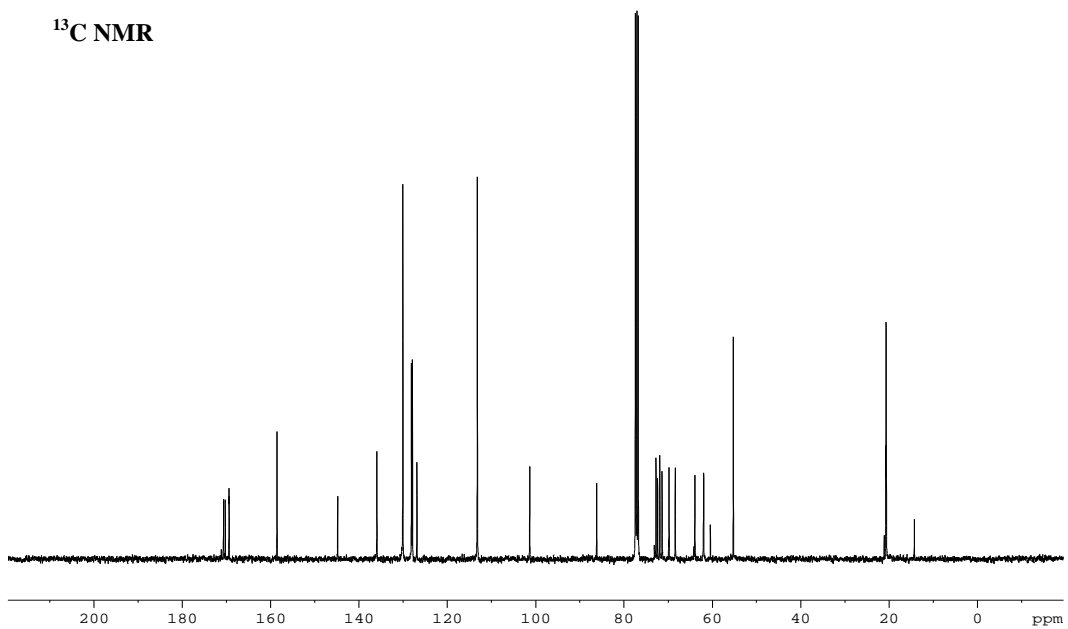
A solution of compound **17** (400 mg, 0.95 mmol) in dry pyridine-CH₂Cl₂ (1:1, 6 ml) was reacted following the general procedure for hydroxyl protection. The crude was purified by silica gel column chromatography (Hex/EtOAc, from 2:1 to 1:2) to give **6** (484 mg, 70%) as a syrup. ¹H NMR (500 MHz, CDCl₃) δ (ppm) : 7.42 (d, 2H, *J* = 7.8 Hz, H_{arom}), 7.33-7.17 (m, 7H, H_{arom}), 6.83 (d, 4H, *J* = 8.7 Hz, H_{arom}), 5.20 (t, 1H, *J* = 9.6 Hz, H₃), 5.08 (t, 1H, *J* = 9.6 Hz, H₄), 4.99 (t, 1H, *J* = 9.3 Hz, H₂), 4.55 (d, 1H, *J* = 7.8 Hz, H₁), 4.22-4.11 (m, 2H, H₆, H_{6'}), 3.94-3.77 (m, 9H, -OCH₂-, 2x-OCH₃, -CH-), 3.74-3.71 (m, 1H, H₅), 3.20-3.15 (m, 2H, -OCH₂-), 2.06, 2.04, 2.01, 1.97 (4s, 12H, 4x-OCOCH₃); ¹³C NMR (100 MHz, CDCl₃): δ (ppm): 170.7, 170.2,

169.5, 167.9 (CO), 158.5, 144.8, 135.9, 134.1, 131.8, 130.0, 129.1, 128.1, 127.8, 126.8, 123.5, 113.2, 101.3 (C₁), 86.0, 73.1, 72.7, 71.9, 71.3, 70.8, 69.9, 68.4, 55.2, 44.7, 20.7, 20.6. HRMS (FAB⁺) Calcd. for C₃₈H₄₄NaO₁₄ (M+Na): 747.2629, found; 747.2639.

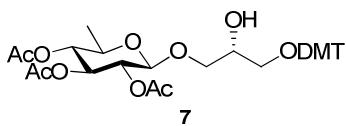
¹H NMR



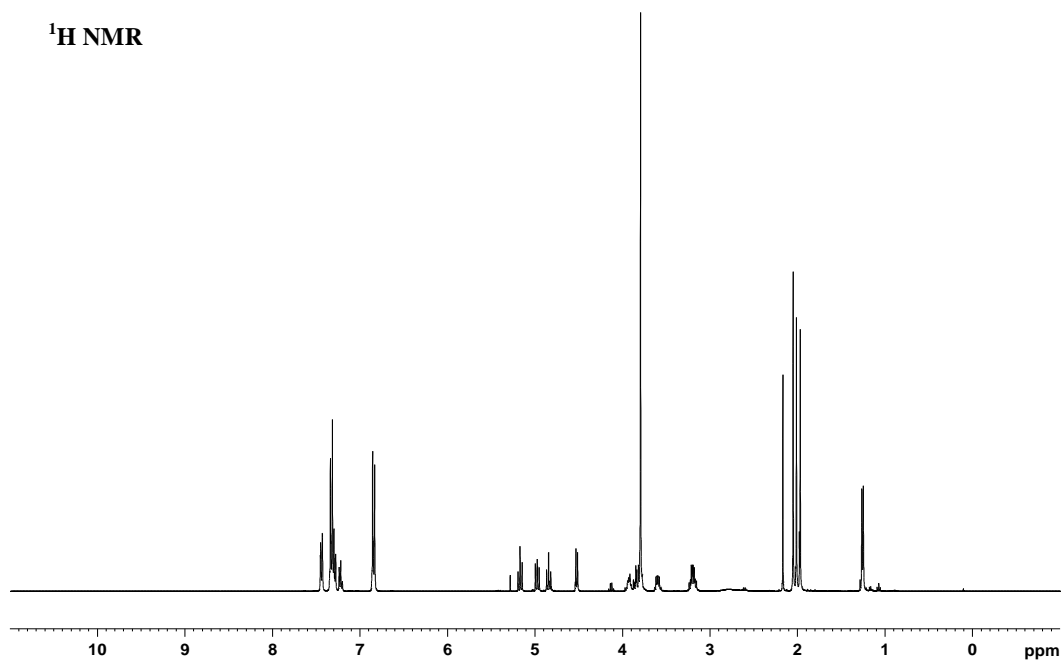
¹³C NMR

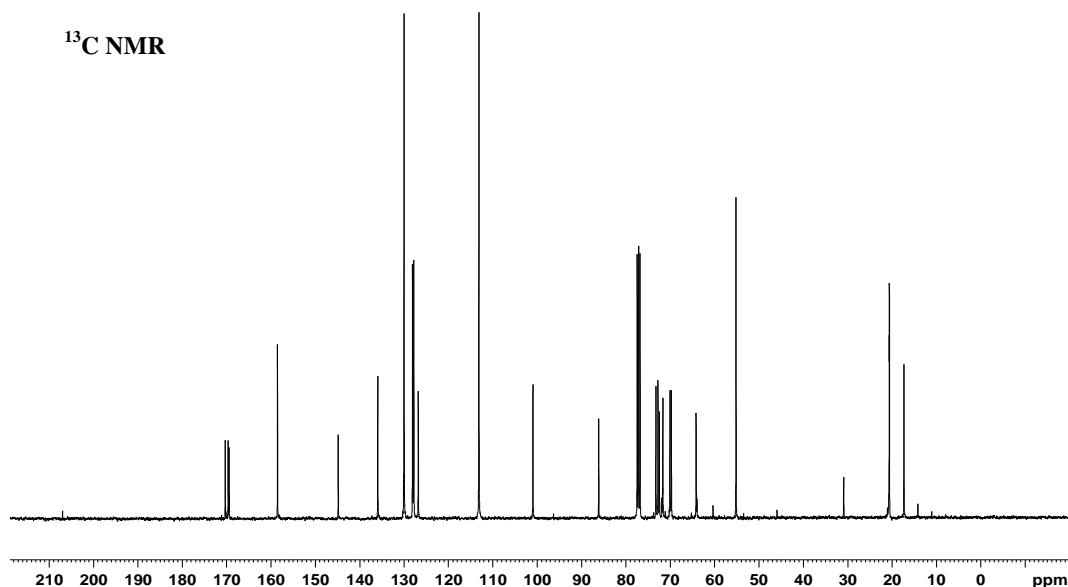


(R)-3-(4,4'-dimethoxytrityloxy)-2-hydroxypropyl 2,3,4-tri-O-acetyl-6-deoxyl-β-D-glucopyranoside (7)

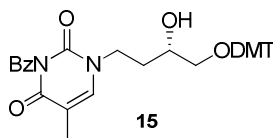


A solution of compound **18** (225 mg, 0.68 mmol) in dry pyridine-CH₂Cl₂ (1:1, 20 ml) was reacted following the general procedure for hydroxyl protection. The crude was purified by silica gel column chromatography (Hex/EtOAc, from 2:1 to 1:2) to give **7** (223 mg, 54%) as a syrup. ¹H NMR (400 MHz, CDCl₃) δ (ppm): 7.44 (d, 2H, *J* = 8.8 Hz, H_{arom}), 7.35-7.20 (m, 7H, H_{arom}), 6.83 (d, 4H, *J* = 8.8 Hz, H_{arom}), 5.17 (t, 1H, *J* = 9.6 Hz, H₃), 4.97 (dd, 1H, *J* = 8.0/9.6 Hz, H₂), 4.86 (t, 1H, *J* = 9.6 Hz, H₄), 4.53 (d, 1H, *J* = 8.0 Hz, H₁), 3.94-3.92 (m, 1H, -CH-), 3.88-3.74 (m, 8H, -OCH₂-, 2x-OCH₃), 3.62-3.58 (m, 1H, H₅), 3.24-3.16 (m, 2H, -OCH₂-), 2.05, 2.01, 1.97 (3s, 6H, 3x-OCOCH₃), 1.25 (d, 3H, *J* = 6.2 Hz, -CH₃); ¹³C NMR (125 MHz, CDCl₃) δ (ppm): 170.3, 169.6, 169.4, 158.5, 144.8, 135.9, 130.5, 128.1, 127.8, 126.8, 113.1, 100.9, 86.1, 73.2, 72.8, 72.5, 71.7, 70.1, 69.8, 64.2, 55.2, 30.9, 20.7, 20.6, 17.3. HRMS (FAB⁺) Calcd. for C₃₆H₄₂NaO₁₂ (M+Na): 689.2574, found; 689.2587.

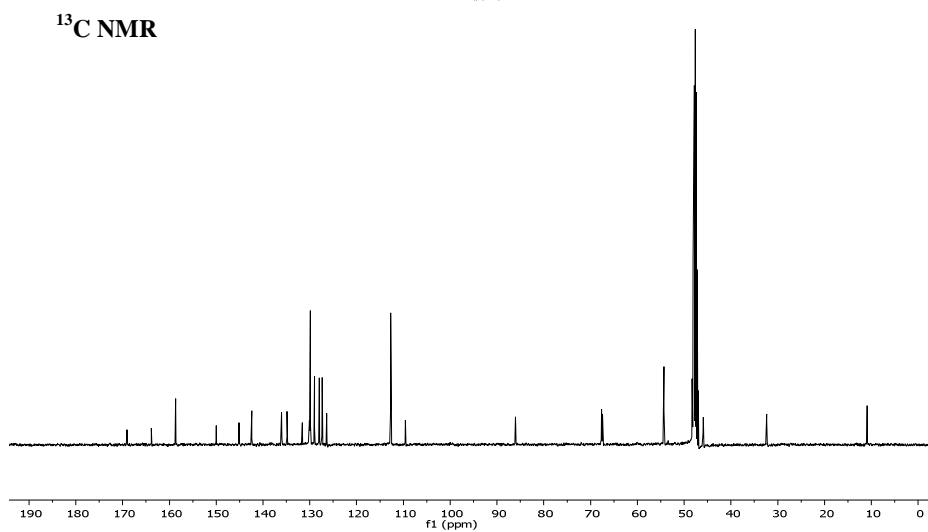
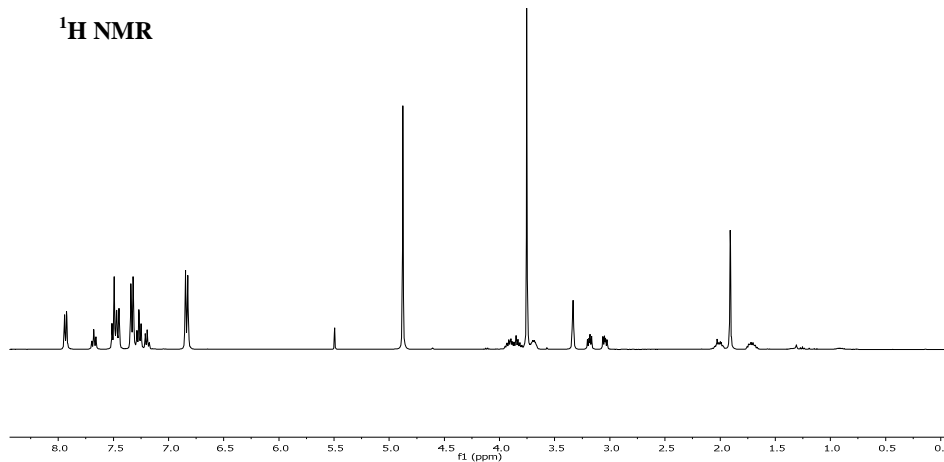




(S)-1-(4-(4,4'-dimethoxytrityloxy)-3-hydroxybutyl)thymine (15)

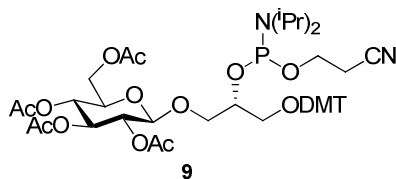


Compound **14** was reacted following the general procedure for hydroxyl protection. The reaction mixture was purified by flash chromatography (Hex/EtOAc, 1:2) to give compound **15** (745 mg, 87%) as a white solid. ¹H NMR (400 MHz, CD₃OD) δ (ppm): 7.93 (d, 2H, *J* = 7.6 Hz, H_{benzoyl}), 7.68 (t, 1H, *J* = 7.6 Hz, H_{benzoyl}), 7.53-7.745 (m, 5H, H_{benzoyl}, H_{thymine}, H_{arom}), 7.34 (d, 4H, *J* = 8 Hz, H_{arom}), 7.29-7.25 (m, 3H, H_{arom}), 7.21-7.17 (m, 1H, H_{arom}), 6.84 (d, *J* = 8 Hz, 4H, H_{arom}), 3.95-3.80 (m, 2H, -CH₂-CH₂N-), 3.75 (s, 6H, 2x-OCH₃), 3.70-3.67 (m, 1H, -OCH₂CH(OH)CH₂-), 3.20-3.17 (m, 1H, -OCHHCH(OH)CH₂-), 3.06-3.02 (m, 1H, -OCHHCH(OH)CH₂-), 2.06-1.98 (m, 1H, -CH(OH)CHHCH₂N-), 1.91 (s, 3H, -CH₃_{thymine}), 1.76-1.68 (m, 1H, -CH(OH)CHHCH₂N-); ¹³C NMR (101 MHz, CD₃OD) δ (ppm): 169.1, 163.8, 158.7, 150.0, 145.1, 142.4, 136.0, 134.9, 131.6, 130.0, 129.9, 129.0, 128.0, 127.4, 126.4, 112.7, 109.6, 86.0, 67.6, 67.4, 54.3, 45.9, 32.4, 10.9. HRMS (EI⁺) Calcd. for C₃₇H₃₆N₂O₇: 620.2523, found; 620.2526.

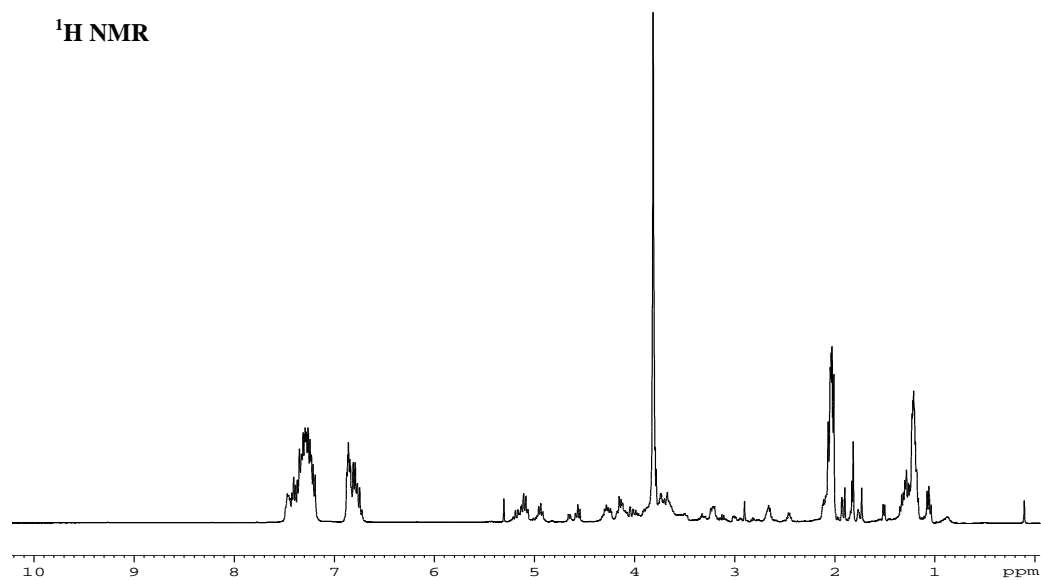


General procedure for synthesis of carbohydrate phosphoramidites. To a solution of DMT protected compound (0.546 mmol) in dry CH₂Cl₂ (5 ml), DIPEA (0.382 ml, 2.185 mmol) and 2-cyanoethyl-*N,N'*-diisopropylamino-chlorophosphoramidite **8** (0.183 ml, 0.819 mmol) were added at room temperature under argon atmosphere. After 1h the solvent was evaporated to dryness. The product was purified by silica gel column chromatography using a mixture of hexane, ethyl acetate and triethylamine.

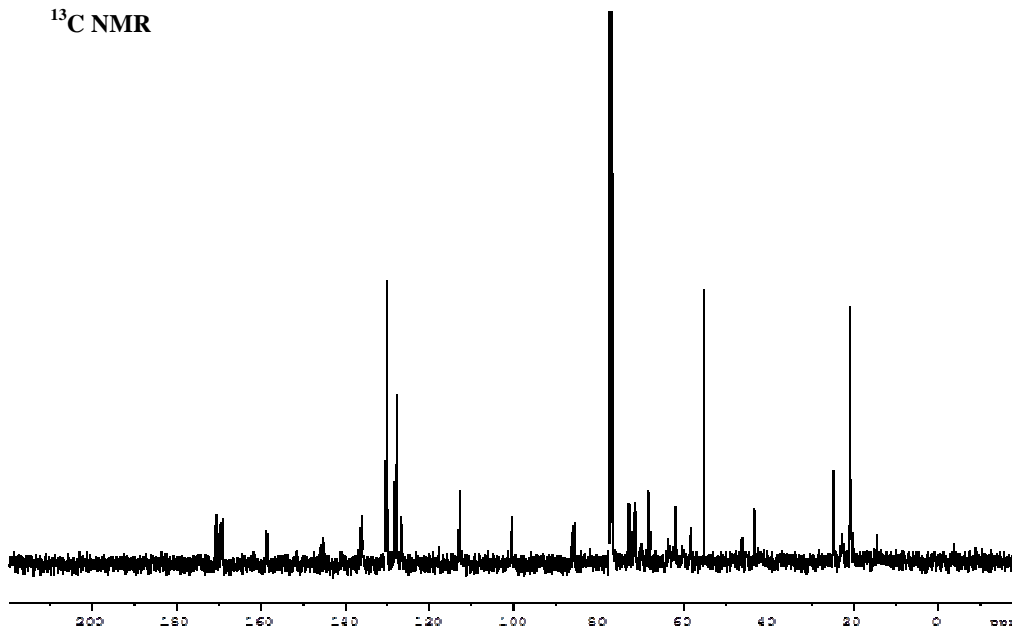
2-(*R*)-1-(4,4'-dimethoxytryloxy)-3-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyloxy)propyl (2-cyanoethyl) (*N,N'*-diisopropyl) phosphoramidite (9**)**



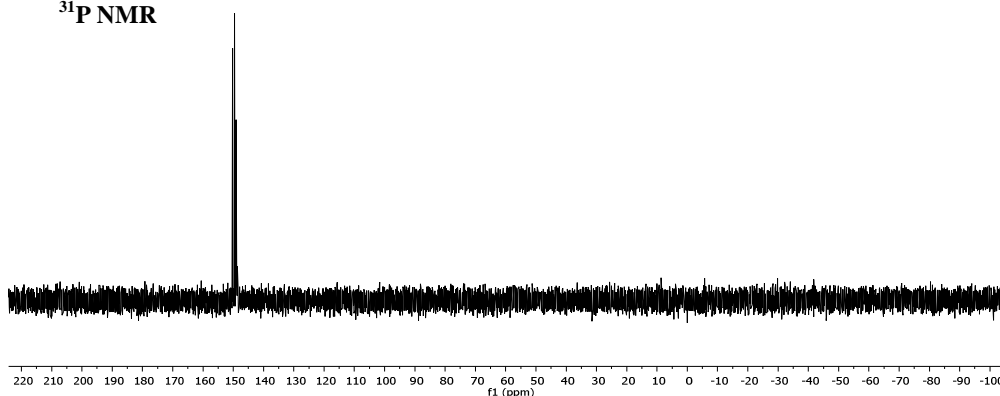
A solution of compound **6** (236 mg, 0.325 mmol) in anhydrous CH_2Cl_2 (5 ml) was reacted following the general procedure for phosphoramidite synthesis. The crude was purified by silica gel column chromatography (Hex/EtOAc, 3:2 with 5% of NEt_3) to give compound **9** (215 mg, 89%) as a white solid. ^1H NMR (400 MHz, CDCl_3) δ (ppm) (mix of isomers): 7.43-7.13 (m, 9H, H_{arom}), 6.85 (m, 4H, H_{arom}), 5.18-4.85 (m, 3H, $\text{H}_2, \text{H}_3, \text{H}_4$), 4.65-4.51 (m, 1H, H_1), 4.29-4.19 (m, 1H, $-\text{CH}-$), 4.29-3.44 (m, 15H, $2\times\text{-OCH}_2-$, $-\text{OCH}_2\text{CH}_2\text{CN}$, $2\times\text{-OCH}_3$, $\text{H}_5, \text{H}_6, \text{H}_6'$), 3.32-2.83 (m, 2H, $2\times\text{CH}(\text{CH}_3)_2$), 2.69-2.60 (m, 2H, $-\text{CH}_2\text{CH}_2\text{CN}$), 2.17-1.69 (m, 12H, $4\times -\text{OCOCH}_3$), 1.32-1.00 (m, 12H, $2\times\text{-CH}(\text{CH}_3)_2$); ^{13}C NMR (101 MHz, CDCl_3) δ (ppm): 171.1, 170.7, 170.2, 169.3, 158.5 (2), 145.1, 136.8, 136.2, 130.4, 130.1, 130.0, 129.1, 128.35, 128.3, 127.8, 127.7, 126.9, 126.7, 118.0, 117.7, 116.2, 113.1, 100.7, 86.4, 76.8, 73.1, 72.6, 71.3, 68.4, 62.4, 61.8, 60.4, 55.2, 55.1, 43.2, 43.1, 24.7, 24.6, 24.5, 24.4, 21.1, 20.6, 20.4, 20.2. ^{31}P NMR (162 MHz, CDCl_3) δ (ppm): 150.2, 149.6. HRMS (FAB $^+$) Calcd. for $\text{C}_{47}\text{H}_{61}\text{N}_2\text{KO}_{15}\text{P}$ (M+K): 963.3447, found; 963.3597.



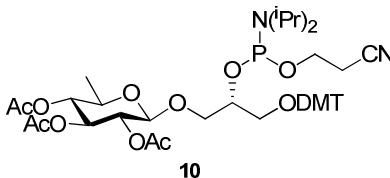
¹³C NMR



³¹P NMR



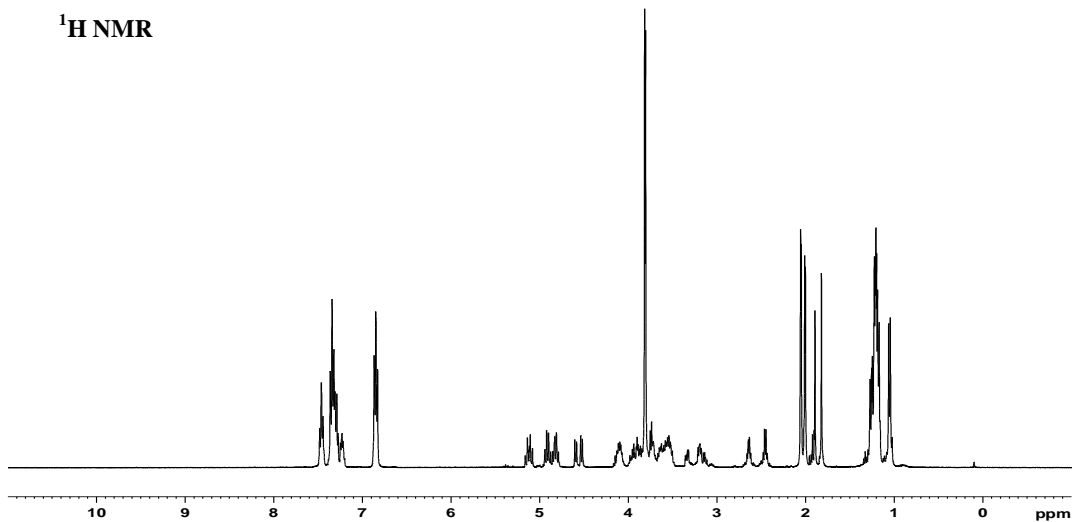
2-(*R*)-1-(4,4'-dimethoxytrytilyloxy)-3-(2,3,4-tri-*O*-acetyl-6-deoxy- β -D-glucopyranosyloxy)propyl (2-cyanoethyl) (*N,N'*-diisopropyl) phosphoramidite (10**)**



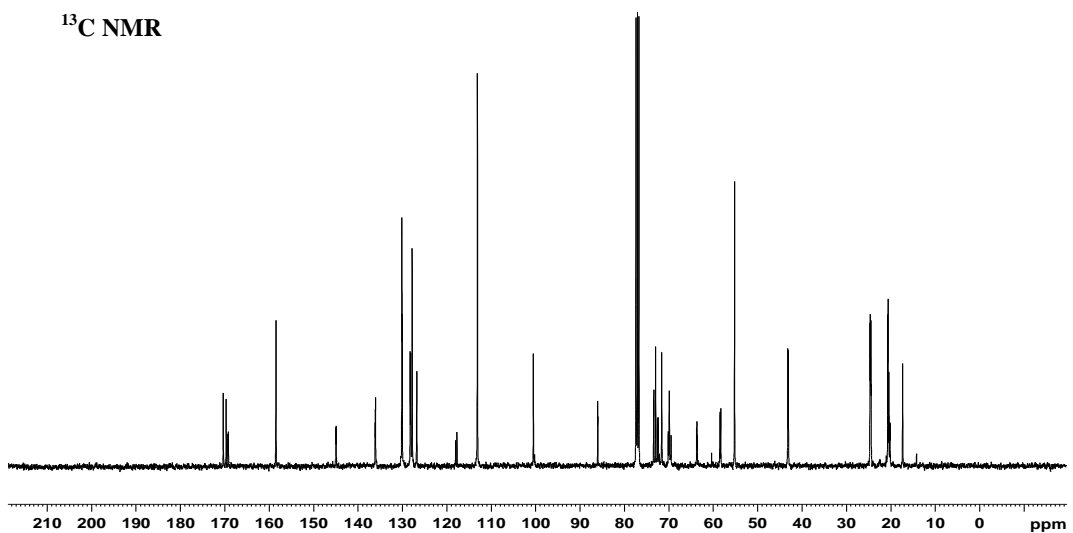
A solution of compound **7** (200 mg, 0.299 mmol) in anhydrous CH₂Cl₂ (5 ml) was reacted following the general procedure for phosphoramidite synthesis. The crude was purified by silica gel column chromatography (Hex/EtOAc, 2:1 with 5% of NEt₃) to give compound **10** (230 mg, 89%) as a glassy solid. ¹H NMR (400 MHz, CDCl₃) δ (ppm) (mix of isomers): 7.48-7.22 (m, 9H, H_{arom}), 6.85 (t, 4H, *J* = 7.0 Hz, H_{arom}), 5.16-5.08 (m, 1H, H₃), 4.94-4.78 (m, 2H, H₄, H₂), 4.55 (dd, 1H, *J*_{H,P} = 26.0 Hz, *J* = 8.0 Hz, H₁), 4.17-4.04 (m, 1H, -CH-), 4.00-3.67 (m, 10H, -

OCH₂-, -OCH₂CH₂CN, 2x-OCH₃), 3.66-3.47 (m, 3H, 2x-CH(CH₃)₂, H₅), 3.33-3.13 (m, 2H, -OCH₂-), 2.66-2.43 (m, 2H,-CH₂CH₂CN), 2.05, 2.00, 1.92 (3s, 9H, 3x-OCOCH₃), 1.29-1.16 (m, 12H, 2x-CH(CH₃)₂), 1.05 (d, 3H, *J* =6.8 Hz, -CH₃); ¹³C NMR (101 MHz, CDCl₃) δ (ppm): 170.3, 169.7, 169.3, 169.2, 158.5, 149.9, 136.1, 130.1, 128.2, 128.1, 127.8, 126.8, 117.8, 113.1, 100.5, 86.1, 73.4, 73.0, 72.6, 72.4, 71.6, 70.2, 69.9, 69.5, 63.7, 58.4, 43.2, 24.7, 24.6, 24.5, 20.7, 20.6, 20.5, 17.3. ³¹P NMR (162 MHz, CDCl₃) δ (ppm): 149.5, 148.9.HRMS (FAB⁺) Calcd. for C₄₅H₅₉N₂NaO₁₃P (M+Na): 889.3652, found: 889.3683.

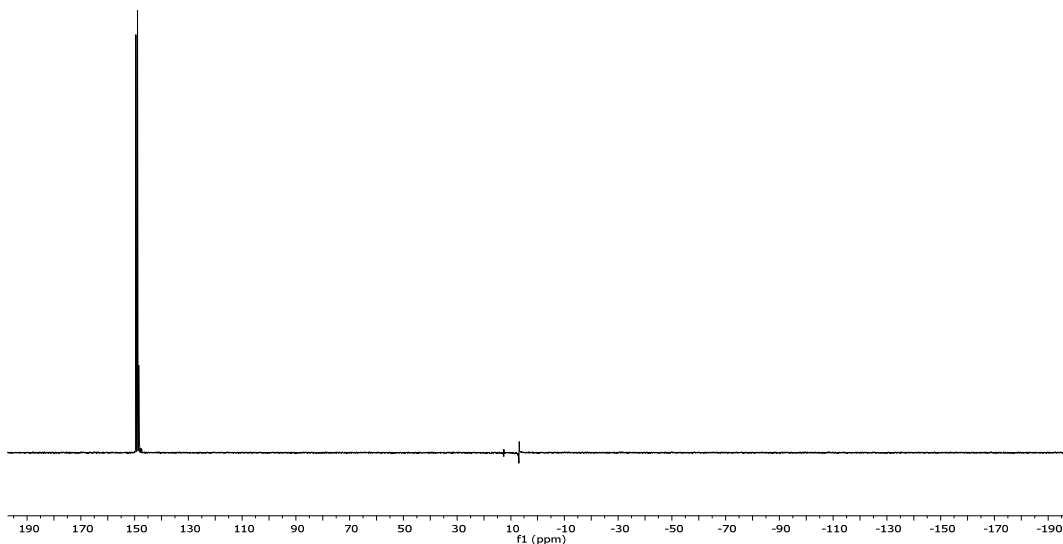
¹H NMR



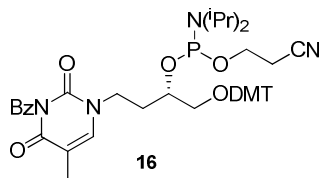
¹³C NMR



³¹P NMR

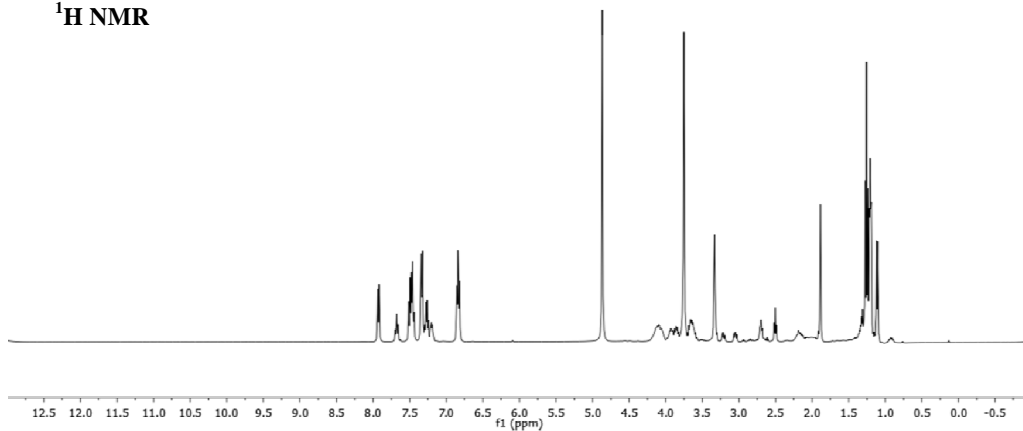


2-(*S*)-1-(4,4'-dimethoxytryloxy)-4-(thyminy)butyl (2-cyanoethyl) (*N,N'*-diisopropyl) phosphoramidite (16**)**

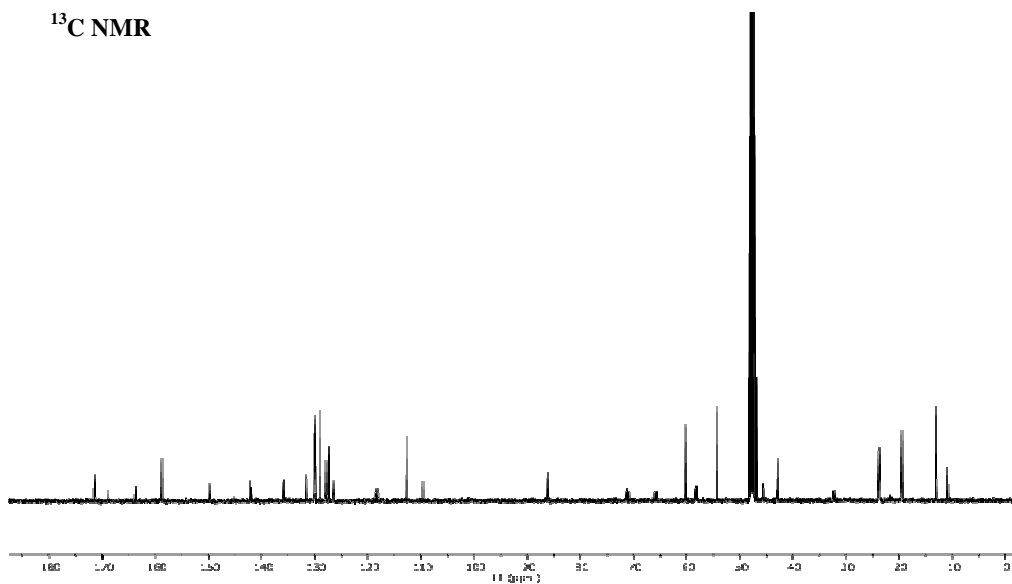


A solution of compound **15** (339 mg, 0.546 mmol) in anhydrous CH₂Cl₂ (5 ml) was reacted following the general procedure for phosphoramidite synthesis. The crude was purified by silica gel column chromatography (Hex/EtOAc, from 1:1 to 1:3, with 5% of NEt₃) to give compound **16** as a white solid (360 mg, 80%). ¹H NMR (400 MHz, CD₃OD) δ (ppm) (mix of isomers): 7.94 (d, 2H, *J* = 8 Hz, H_{benzoyl}), 7.70 (t, 1H, *J* = 7.6 Hz, H_{benzoyl}), 7.53-7.746 (m, 5H, H_{benzoyl}, H_{thymine}, 2xH_{arom}), 7.34 (d, 4H, *J* = 8.4 Hz, H_{arom}), 7.30-7.25 (m, 2H, H_{arom}), 7.23-7.19 (m, 1H, H_{arom}), 6.84 (m, 4H, H_{arom}), 4.14-4.02 (m, 1H, -CH-), 3.98-3.81 (m, 2H, -OCH₂-), 3.76, 3.75 (2s, 6H, 2x-OCH₃), 3.70-3.57 (m, 3H, -CH₂O-, -CHHO-), 3.35-3.30 (m, 1H, -CHHO-), 3.23-3.02 (m, 1H, -CH(CH₃)₂), 2.72-2.67 (m, 1H, -CH(CH₃)₂), 2.52-2.49 (m, 1H, -CHHCH₂N-), 2.23-2.08 (m, 1H, -CHHCH₂N-), 1.88 (s, 3H, -CH₃thymine), 1.35-1.10 (m, 14H, -CH₂CN, 4x-CH₃isopropyl). ¹³C NMR (101 MHz, CD₃OD) δ (ppm): 169.0, 163.8, 163.7, 158.7, 149.9, 145.1, 142.2, 142.0, 135.9, 135.9, 135.8, 134.9, 131.6, 130.1, 130.0, 129.9 (2), 129.0, 128.0, 127.9, 127.4, 126.4, 118.4, 118.1, 112.7, 109.8, 109.6, 86.1, 71.4, 71.2, 71.1, 70.9, 66.0, 65.7, 58.3, 58.2, 58.1, 58.0, 54.3, 47.6, 47.4, 47.2, 47.0, 45.6, 45.5, 43.0, 42.9, 32.4, 32.1, 23.9, 23.8 (2), 23.7, 19.6, 19.5, 11.0. ³¹P NMR (162 MHz, CD₃OD) δ (ppm): 148.6, 148.1. HRMS (ES⁺) Calcd. for C₄₆H₅₃N₄NaO₇P (M+Na): 843.3493, found; 843.3488.

¹H NMR



¹³C NMR



³¹P NMR

Synthesis of natural and modified oligonucleotide DNA strands

All natural and modified oligonucleotide DNA strands were synthesized by *Biomers* following standard β -cyanoethylphosphoramidite chemistry on 200 nmol or 1 μ mol scale and using the DMT-off procedure. Oligonucleotide supports were treated with 33% aqueous ammonia for 16 h at 55 °C, then ammonia solutions were evaporated to dryness.

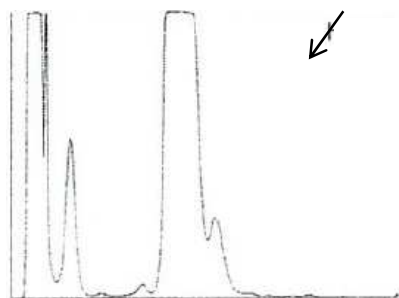
HPLC purification was carried out in a Waters Alliance 2690 RP-HPLC with a UV-Vis Photodiode Array (Waters) and using a Nucleosil 120 C18 (250 x 8 mm, 10 μ m) column; 27 min linear gradient from 0 to 20% B (DMT off conditions); flow rate, 3 ml/ min; solution A was 5% acetonitrile in 0.1 M aqueous triethylammonium acetate (TEAA, pH 6.5) buffer and solution B was 70% acetonitrile in 0.1 M aqueous TEAA (pH 6.5).

HPLC chromatograms of modified oligonucleotide DNA strands

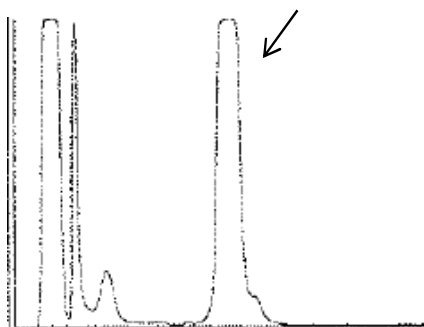
5'-GATGACT*GCTAG



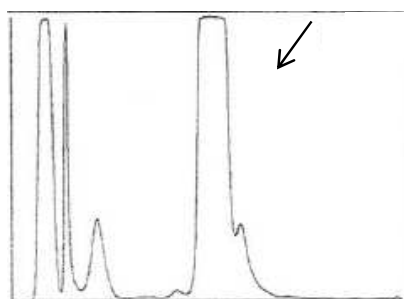
5'-GATGAC-glc-GCTAG



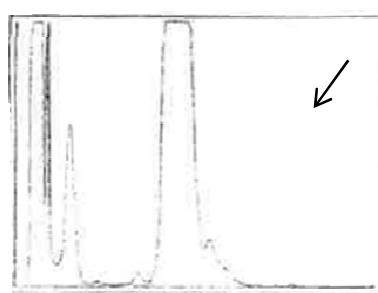
5'-GATGAC-6dglc-GCTAG



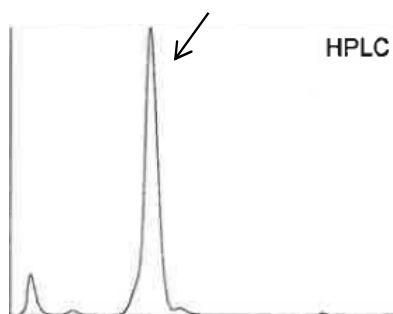
5'-CTAGC-T*-GTCATC



5'-CTAGC-glc-GTCATC



5'-CTAGC-**6dglc**-GTCATC



MALDI-TOF data of modified oligonucleotide DNA strands

Carbohydrate oligonucleotide conjugates	[M-H] calc.	[M-H] exp.
GATGAC- T* -GCTAG (strand 1)	3659	3661
GATGAC- glc -GCTAG (strand 2)	3698	3697
GATGAC- 6dglc -GCTAG (strand 3)	3683	3681
CTAGC- T* -GTCATC (strand 4)	3570	3571
CTAGC- glc -GTCATC (strand 5)	3609	3606
CTAGC- 6dglc -GTCATC (strand 6)	3594	3593

Synthesis of oligonucleotide GNA strands

a. Synthesis of GNA Phosphoramidites.

The synthesis of DMT-phosphoramidite glycidol A and T followed the reported procedure described by Meggers et al.²

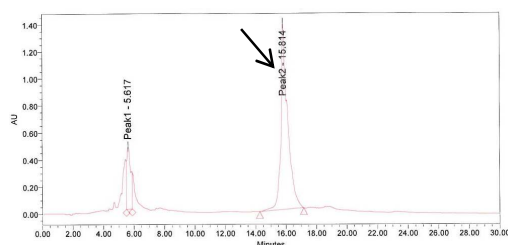
b. GNA Solid support synthesis. To 25 mg (0.049 mmol) of DMT-glycidol T were added succinic anhydride 7.4 mg (0.074 mmol) and 9.0 mg (0.074 mmol) of dimethylaminopyridine in anhydrous DCM (10 ml). The mixture was stirred overnight and then washed with 0.1 M NaH_2PO_4 and dried over MgSO_4 . The solution was concentrated and the succinic derivative was used without further purification for the coupling to the solid support. We then followed the described procedure,² 9.3 mg (0.03 mmol) of 2,2'-dithio-bis-(5-nitropyridine) dissolved in 120 μl ACN: dichloroethane (1:3) was mixed with a solution of the succinic derivative 20 mg (0.03 mmol) and 3.66 mg (0.03 mmol) DMAP dissolved in 500 μl of ACN. 7.8 mg (0.03 mmol) of PPh_3 dissolved in 60 μl of ACN were then added to clear the obtained solution. The mixture was vortexed and added to a vial containing 150 mg (10.7 μmol) of LCAA-CPG 500 Å for 4 hours at room temperature. Then, the resin was washed with ACN and DCM. The unreacted amino groups were acetylated with 1 ml of two capping solutions (acetic anhydride/pyridine/tetrahydrofuran (1:1:8) (capping A), 10% *N*-methylimidazole in tetrahydrofurane (capping B). Trityl analysis yielded a loading of 18 $\mu\text{mol/g}$.

c. GNA Oligonucleotide synthesis and purification. DNA oligonucleotides were synthesized on the 0.2 μmol scale on an Applied Biosystems 394 synthesizer following standard conditions. Modified amidites were dissolved in anhydrous acetonitrile (0.1 M) and all oligonucleotides were synthesized in DMT-ON mode. Then the solid support treated twice with 0.1 M DBU/ACN for five minutes and washed with 1% Et_3N / ACN. The solid support was transferred to a screw-cap glass vial and incubated at room temperature for four hours with 1.5 mL of NH_3 solution (33%). The ammonia solutions were concentrated to dryness and the product was desalted on NAP-10 (Sephadex G-25) columns eluted with water. All the oligonucleotides were purified by HPLC. Semipreparative column: X-bridgeTM OST C_{18} (10x50 mm, 2.5 μm); 30 min linear gradient from 0 % to 70%, flow rate 2 mL/min; solution A was 5% ACN in 0.1 M aqueous TEAA and B 70% ACN in 0.1 M aqueous TEAA. The desired DMT-ON product corresponds to the main peak followed by secondary peak with similar retention time. The pure fractions were combined and evaporated to dryness. The residue that was obtained was treated with 200 μL of 10% AcOH solution and incubated at room temperature for 5 min. The aqueous solution was neutralized with TEAA and, finally, the deprotected oligonucleotide was desalted

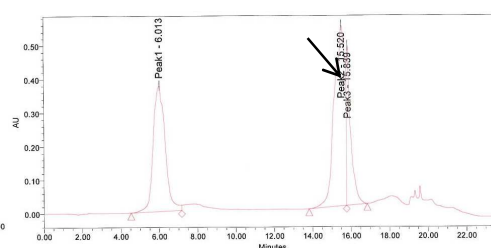
in a NAP-10 column. All oligonucleotides were quantified by absorption at 260 nm and confirmed by MALDI mass spectrometry. Matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectra were recorded on a Voyager-DETMRP spectrometer (Applied Biosystems) in negative mode (2,4,6- trihydroxyacetophenone matrix with ammonium citrate as an additive), (Servei d'Espectrometria de Masses, Universitat de Barcelona).

HPLC chromatograms of oligonucleotide GNA strands

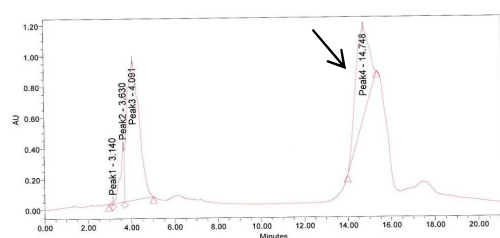
3'- TAAAATTTATATTATTA-2'



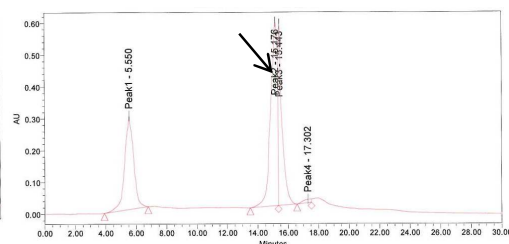
3'- TTAATAATATAAATTTTA-2'



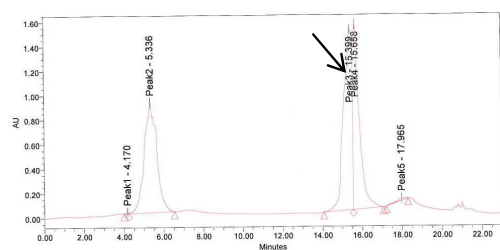
3'- TTAATAATTTAAATTTTA-2'



3'- TAAAATTTA-6dglc-ATTATTA-2'



3'- TTAATAAT-6dglc-TAAATTTTA-2'



MALDI-TOF data of oligonucleotide GNA strands

GNA oligonucleotide sequence	[M-H] calc	[M-H] found
3'- TAAAATTTATATTATTA-2'	4735	4738
3'- TTAATAATATAAATTTTA-2'	4735	4737
3'- TTAATAATTTAAATTTTA-2'	4726	4726

3'- TAAAATTTA 6dglc ATTATTAA-2'	4769	4773
3'- TTAATAAT 6dglc TAAATTTTA-2'	4760	4764

Thermal denaturation methods.

DNA duplexes.

UV-melting curves were measured on a Perkin–Elmer Lambda 750 UV/Vis spectrophotometer. Absorbance of duplexes in a 1:1 stoichiometric ratio were monitored at 260 nm and the heating rate was set to 1.0 °C·min⁻¹ from 10 to 80 °C. The extinction coefficients of the natural oligonucleotide parts, ϵ_{nat} , was calculated using a *oligo calculator* (www.ambion.com). The total absorption coefficient was then calculated by simple addition: $\epsilon = \epsilon_{\text{nat1}} + \epsilon_{\text{nat2}}$. The first derivative of the melting curves was obtained using *Origin 8.0* software. To avoid air water condensation samples were measured in a nitrogen atmosphere.

GNA duplexes.

The absorbance versus temperature curves of duplexes were measured at 0.7 μM strand concentration in 10 mM sodium phosphate buffer at pH 7 implemented with 200 mM NaCl. Thermal UV experiments were performed in Teflon-stopped 1 cm path length quartz cells on a JACSO V-650 spectrophotometer equipped with thermoprogrammer. The samples were heated to 90°C, allowed to slowly cool down to 20°C, and then warmed during the denaturation experiments at a rate of 0.5 °C/min from 10 to 85 °C, monitoring absorbance at 260 nm. Melting temperatures (T_m) were determined by computerfit of the first derivative of absorbance with respect to 1/T.

NMR Spectroscopy.

Samples of all the conjugates and control duplexes were suspended in 500 μL of either D₂O or H₂O/D₂O 9:1 in phosphate buffer 10 mM, 150 mM NaCl, pH 7. NMR spectra were acquired in Bruker Avance spectrometers operating at 600 or 800 MHz and were processed with Topspin software. DQF-COSY, TOCSY, and NOESY experiments were recorded in D₂O. The NOESY spectra were acquired with mixing times of 150 and 300 ms, and the TOCSY spectra were recorded with standard MLEV-17 spin-lock sequence, and 80 ms mixing time. NOESY spectra in H₂O were acquired with 100 ms mixing time. In 2D experiments in H₂O, water suppression was achieved by including a WATERGATE module in the pulse sequence prior to acquisition. Two-dimensional experiments in D₂O were carried out at temperatures ranging from 5 to 25 °C, whereas spectra in H₂O were recorded at 5 °C to reduce the exchange with water. The spectral analysis program Sparky was used for semiautomatic assignment of the NOESY cross-peaks and quantitative evaluation of the NOE intensities. Distance constraints with their

corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

Structure Calculations.

Structures were calculated with the SANDER module of the molecular dynamics package AMBER. Starting models of the conjugate duplexes were built using the program SYBYL. The DNA moieties in the starting models were set to a standard B- canonical structure. These structures were taken as starting points for the AMBER refinement, which started with an annealing protocol *in vacuo* (using hexahydrated Na⁺ counterions placed near the phosphates to neutralize the system). The resulting structures from *in vacuo* calculations were placed in the center of a water-box with around 4000 water molecules and 22 sodium counterions to obtain electroneutral systems. The structures were then refined including explicit solvent, periodic boundary conditions and the Particle-Mesh-Ewald method to evaluate long-range electrostatic interactions. Force field parameters for the carbohydrate moieties were taken from GLYCAM. The TIP3PBOX model was used to describe water molecules. The protocol for the constrained molecular dynamics refinement in solution consisted of an equilibration period of 160 ps using a standard equilibration process, followed by four independent 500 ps runs. Averaged structures were obtained by averaging the last 20 ps of individual trajectories and further energy minimization of the structure. Analysis of the representative structures as well as the MD trajectories was carried out with the program MOLMOL and the analysis tools of AMBER.

DFT quantum chemical methods.

All calculations were performed with the Amsterdam density functional program (ADF 2013.1) (<http://www.scm.com/>).³⁻⁵ Calculations were initiated from the coordinates of the base pairs of interest in the PDB structures (2N9F and 2N9H) we obtained previously from NMR studies and molecular dynamics. The original guanine and thymine bases in the obtained structures were changed to adenine and cytosine, respectively, in order to evaluate the four possible pairs with the natural DNA bases. Bases, base-base pairs, sugar-base pairs and sugar-sugar pairs were optimized at the BLYP-D3(BJ)/TZ2P level of dispersion-corrected density functional theory (DFT) without any symmetrical constraints. Benchmarks have shown that dispersion-corrected DFT reproduces well high level *ab initio* results for weakly interacting systems (that is hydrogen bonding or stacking interactions).⁶ Solvent effects in aqueous solution are described with the COnductor-like Screening MOdel (COSMO), which takes effectively

into account solute–solvent interactions, cavitation, internal energy and entropy effects of the solvent and yields an estimate of the Gibbs free energies.⁷⁻⁹

Figure S1. Synthesis of flexible T* amidite. N-benzyl-thymine **11** was used as starting material.¹ Reagents and conditions: (i) K₂CO₃, TBAI, DMF; (ii) AcOH-H₂O, 80 °C; (iii) DMTCl, DMAP, Py, CH₂Cl₂; (iv) DIPEA, CH₂Cl₂.

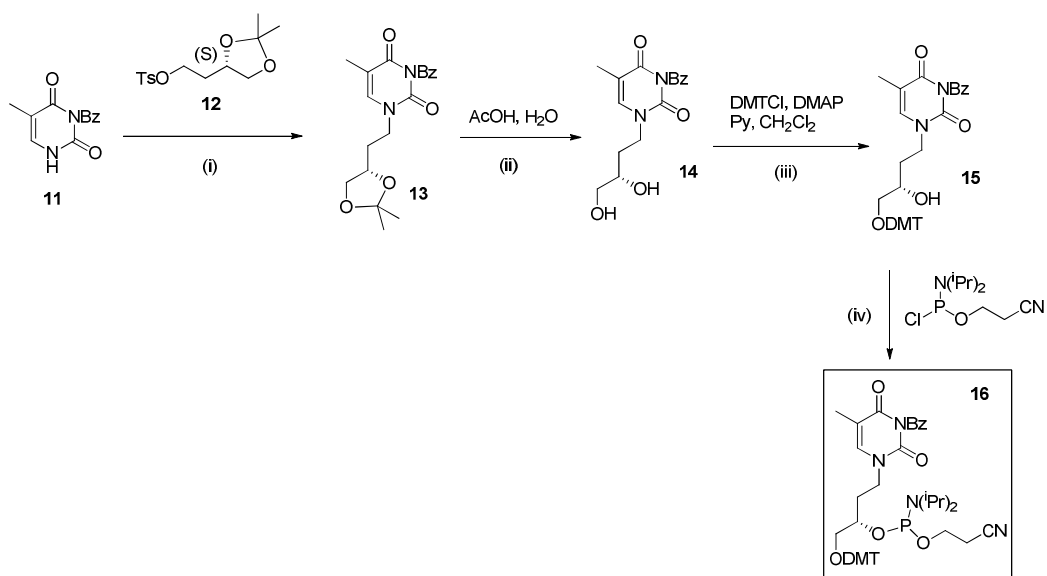


Figure S2: Melting curves for DNA duplexes containing glc-T, glc-G, 6dglc-C and 6dglc-A.

5'-d(GATGACXGCTAG)

3'-d(CTACTGYCGATC)

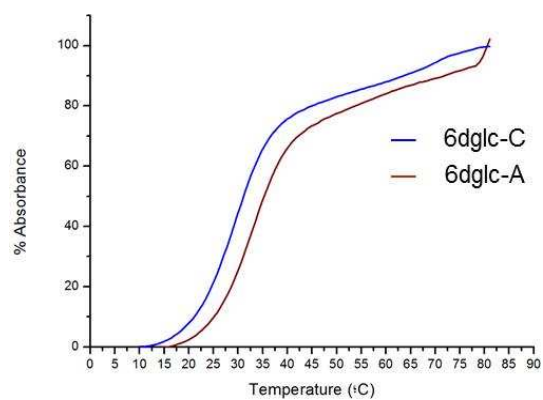
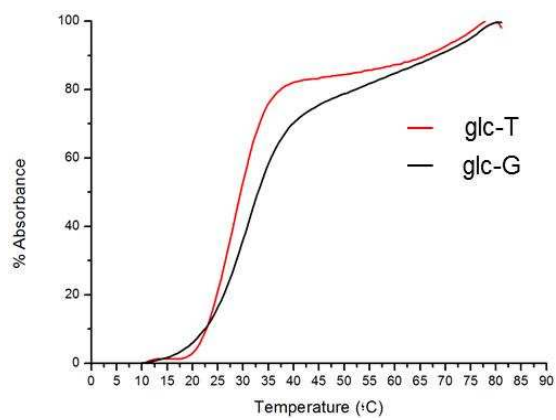
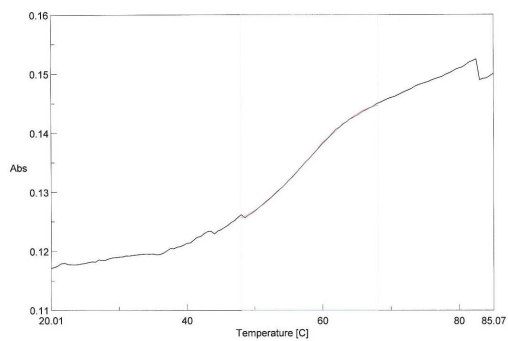


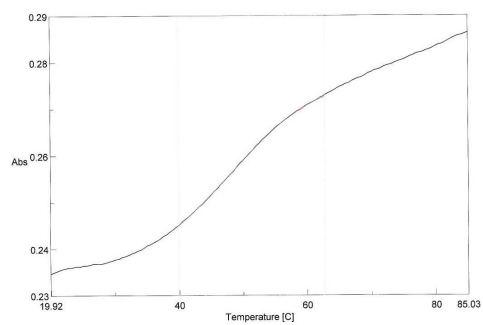
Figure S3: Melting curves for GNA double helices

3'-TAAAATTTAXATTATTAA

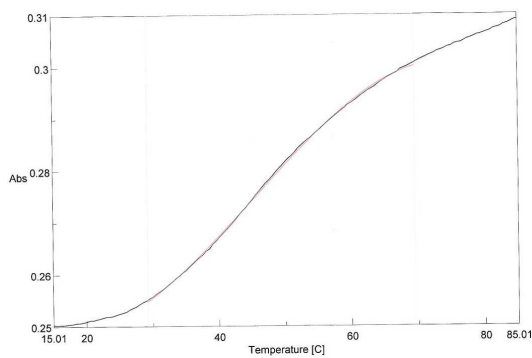
2'-ATTTTAAATYTAATAATT



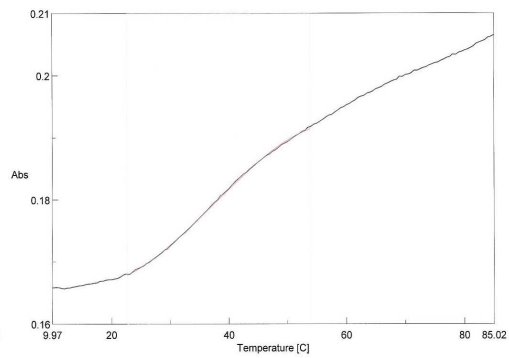
TA pair



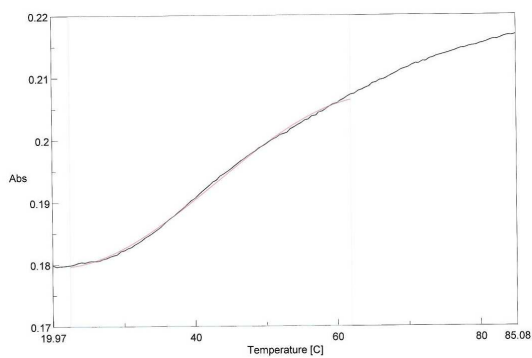
TT pair



6dglc-A pair



T-6dglc pair



6dglc-6dglc pair

Figure S4. Imino region of the NMR spectra of **helix 6dglc-G** and **helix 6dglc-T** at 5 °C (10 mM sodium phosphate, 150 mM NaCl, pH 7).

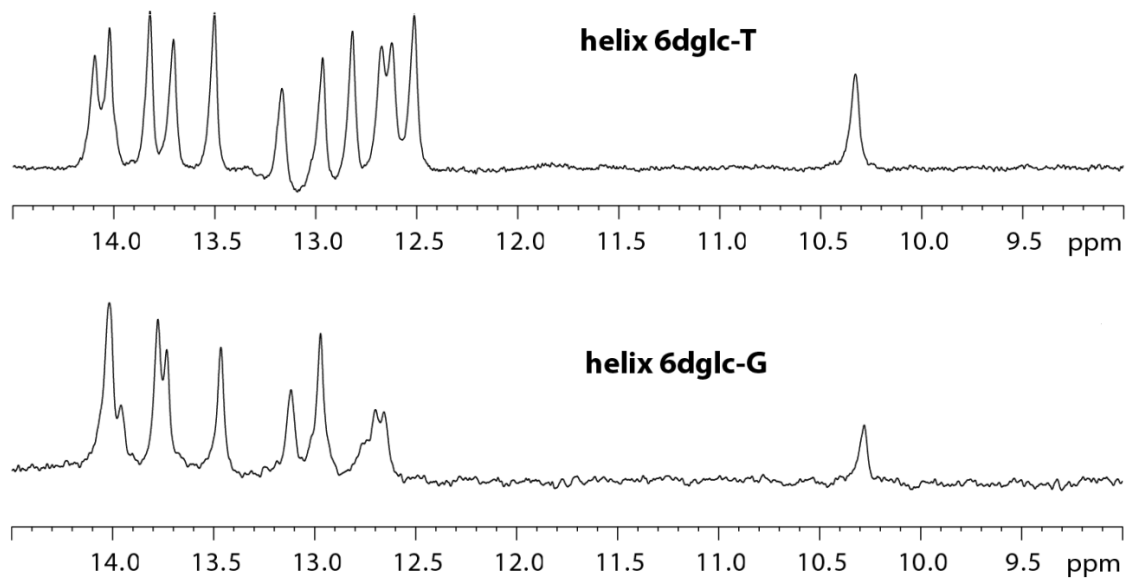
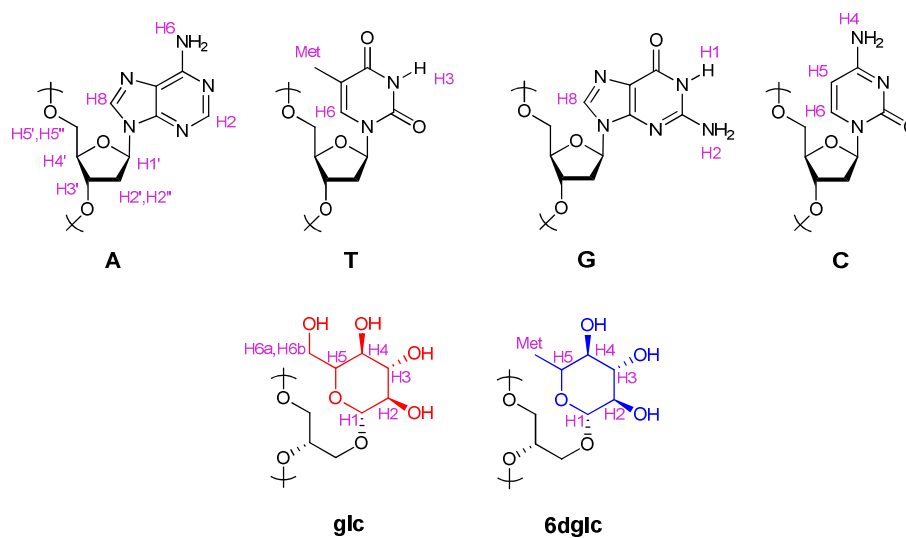
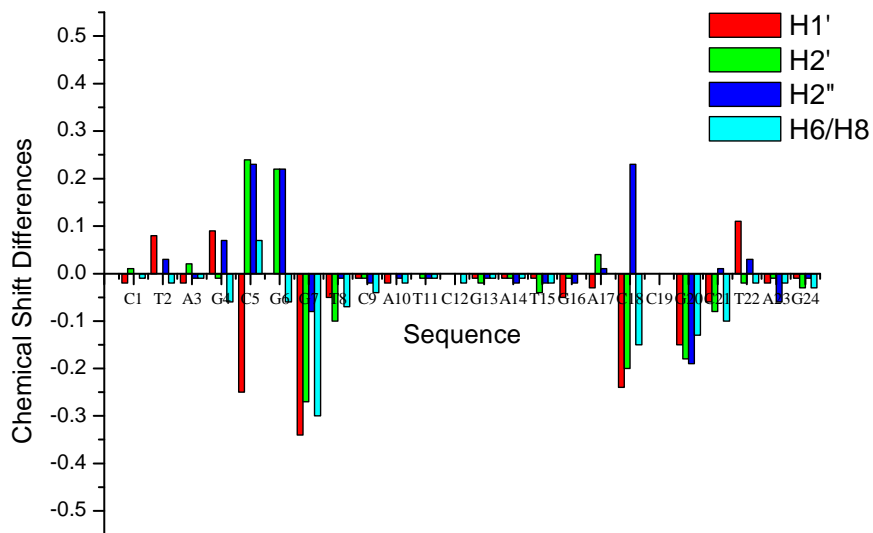


Figure S5. More significant changes in proton chemical shifts along the sequence for **helix 6dglc-G** and **helix 6dglc-T** with respect to double helices containing all natural base pairs.



Helix 6dglc-G

5'-C¹TAGCGGTCATC¹²
3'-G²⁴ATCG-6dGlc-CAGTAG¹³



Helix 6dglc-T

5'-G¹ATGACTGCTAG¹²
3'-C²⁴TACTG-6dglc-CGATC¹³

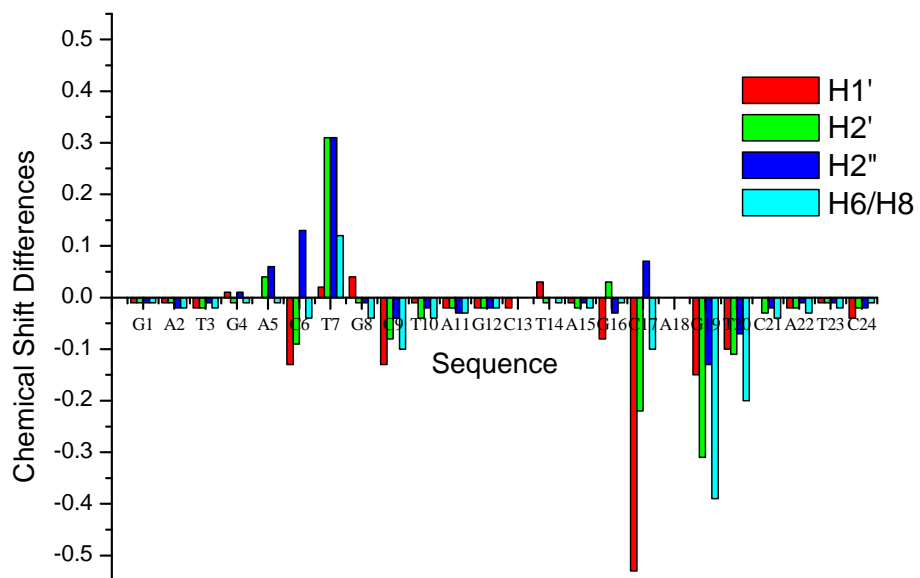


Figure S6.- Several regions of NOESY spectra of **helix 6dglc-T** in H₂O/D₂O 9:1 (left panels) and D₂O (right panels) (mixing time 100 and 250 ms, respectively, T= 5°C). DNA carbohydrate contacts are labeled in red.

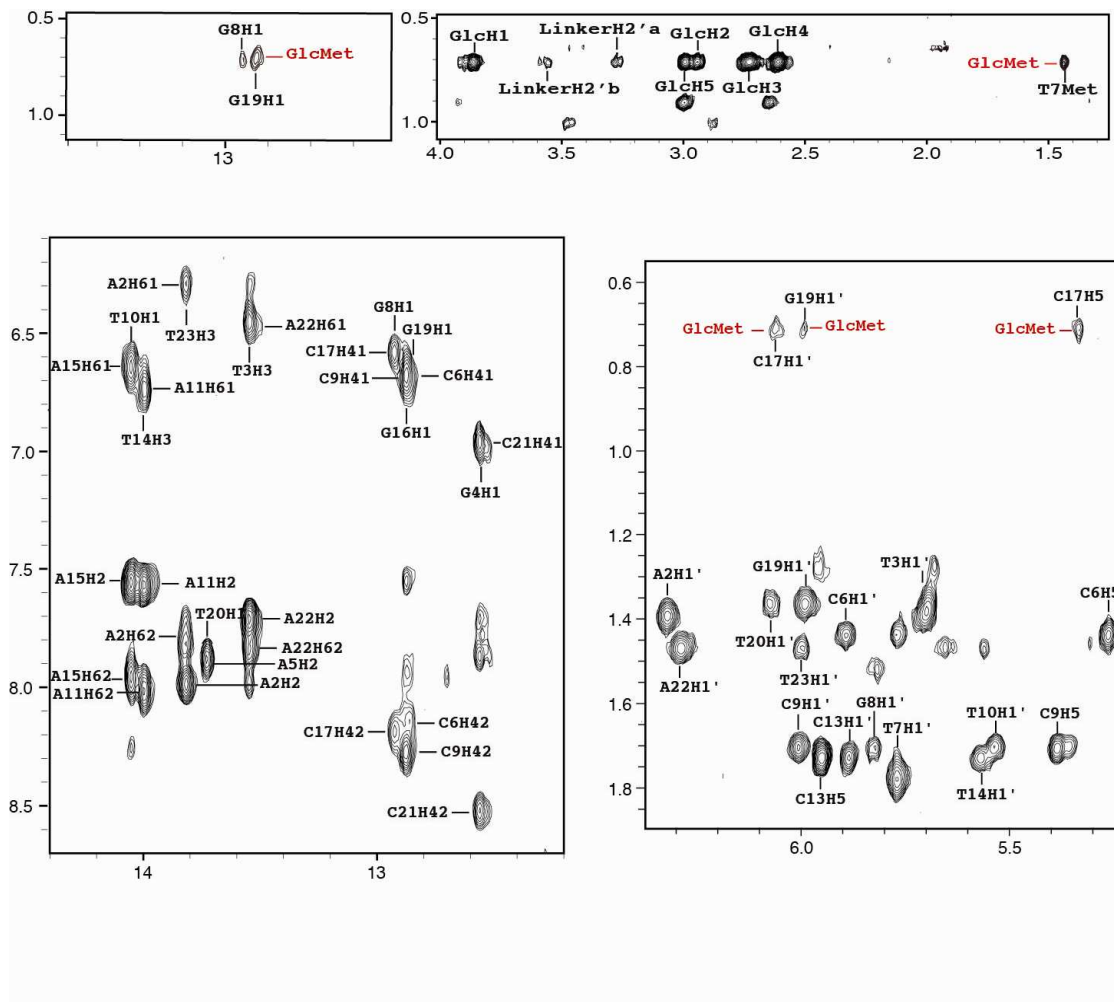


Figure S7. Solution structure of **helix 6dglc-T**. a) Stereoscopic view of the ensemble of the 10 refined structures, b) Stereoscopic view of a representative structure. Color code: modified strand in green; complementary strand in blue; carbohydrate and linker in magenta; hydrogen atoms are shown in grey. c) Two views showing details of the carbohydrate moiety and the surrounding base-pairs. d) Detail of the interaction between carbohydrate and the opposite thymine. Hydrogen bonds are shown in yellow.

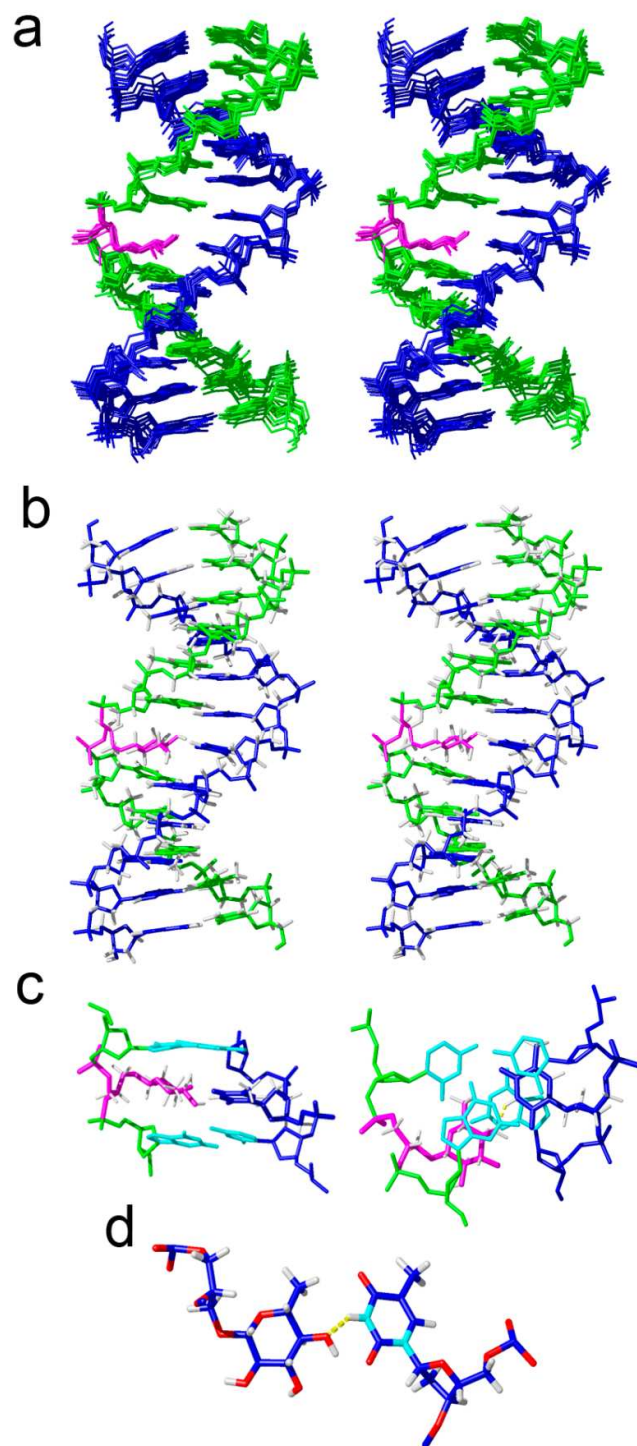


Figure S8. Solution structure of **helix 6dglc-G**. a) Stereoscopic view of the ensemble of the 10 refined structures, b) Stereoscopic view of a representative structure. Color code: modified strand in green; complementary strand in blue; carbohydrate and linker in magenta; hydrogen atoms are shown in grey. c) Two views showing details of the carbohydrate moiety and the

surrounding base-pairs. d) Detail of the interaction between carbohydrate and the opposite guanine. Hydrogen bonds are shown in yellow.

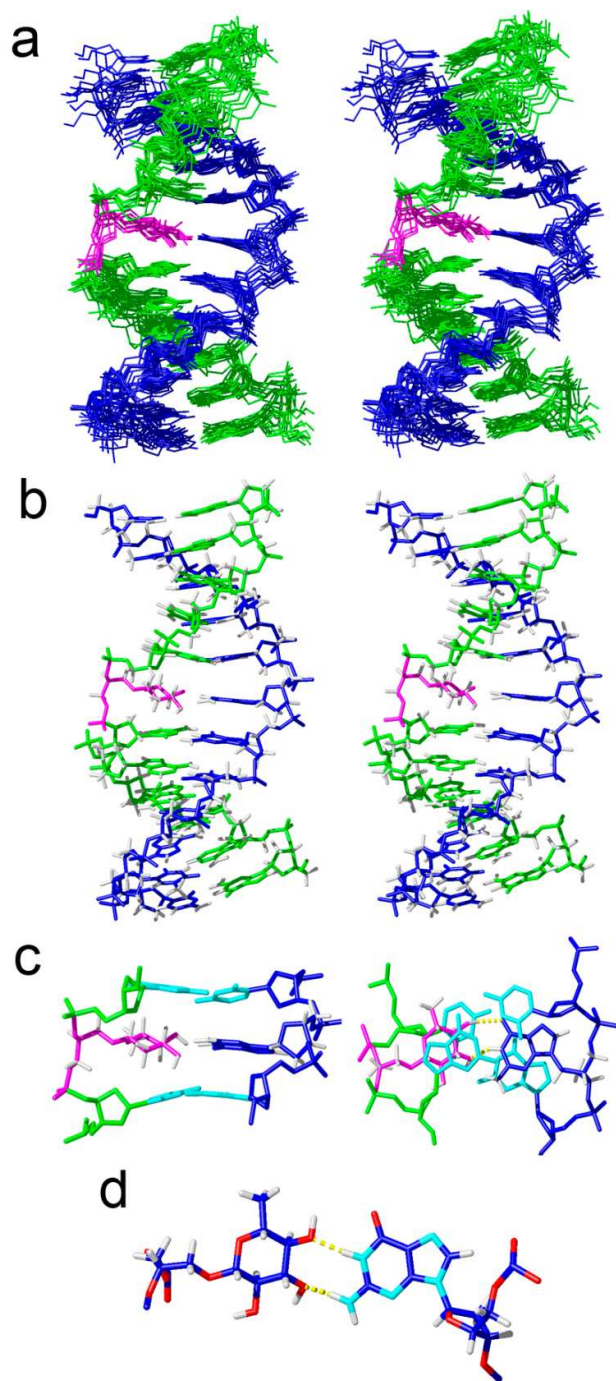


Figure S9. Structures for pairs containing **glc** and **6dglc** in *vacuo* and in water, computed at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution.

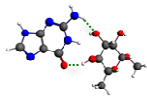
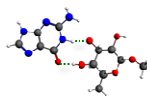
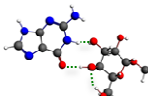

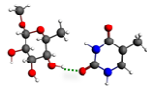
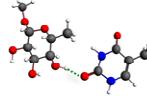
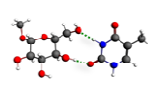
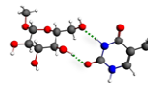
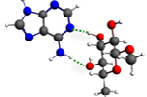
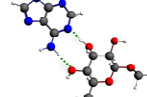
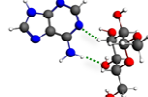
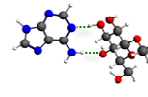
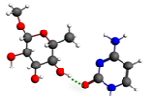
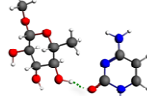
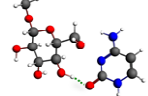
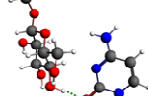
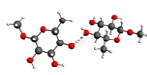
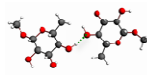
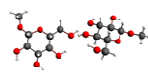
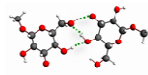
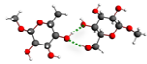
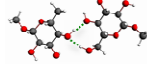
X-Y	gas	water	X-Y	gas	water
6dglc-G			glc-G		
6dglc-T			glc-T		
6dglc-A			glc-A		
6dglc-C			glc-C		
6dglc-6dglc			glc-glc		
6dglc-glc					

Table S1. ¹H-NMR assignments of **helix 6dglc-T** (10 mM sodium phosphate, 150 mM NaCl, pH 7, T = 5°C).[†]

Oligonucleotide strands

helix 6dglc-T	H1'	H2'/H2''	H3'	H4'	H5'/H5''	H5/Met/H2	H6/H8	Imino/amino
G1	5.70	2.66/2.83	4.88	4.23	3.72	---	7.96	n.o.
A2	6.32	2.76/2.99	5.08	4.49	4.23/4.16	7.97	8.37	6.30/7.81
T3	5.71	1.98/2.37	4.88	4.17	4.30	1.40	7.15	13.55
G4	5.54	2.69/2.75	4.93	4.35	n.a.	n.a.	7.89	12.56
A5	6.21	2.66/2.84	4.77	4.47	4.24/4.15	7.87	8.19	n.o.
C6	5.90	2.01/2.34	4.77	4.14	n.a.	5.27	7.32	n.o.
T7	5.77	1.78/2.17	4.76	4.11	4.02	1.44	7.16	10.48
G8	5.83	2.68	4.93	4.34	4.05/3.96	---	7.94	12.92
C9	6.01	2.12/2.49	4.77	4.24	4.15/4.08	5.39	7.53	6.68/8.28
T10	5.54	2.14/2.39	4.87	4.24	4.15/4.07	1.71	7.49	14.04
A11	6.11	2.93/2.76	5.09	4.45	4.15/4.05	7.55	8.27	6.74/8.02
G12	6.03	2.47/2.27	4.66	4.21	4.30/4.13	---	7.77	n.o.
C13	5.89	2.19/2.56	4.69	4.11	3.81	5.96	7.89	7.22/7.99
T14	5.57	2.26/2.53	4.90	4.18	4.04	1.73	7.62	14.00
A15	6.11	2.80/2.96	5.10	4.45	4.19/4.06	7.55	8.27	6.65/7.56
G16	5.77	2.47/2.65	4.97	4.38	4.21	---	7.70	12.87
C17	6.06	2.15/2.28	4.81	n.a.	n.a.	5.33	7.38	6.58/8.17
6dglc18	---	---	---	---	---	---	---	---
G19	6.00	2.76/2.84	4.92	4.33	4.02	---	7.98	n.o.
T20	6.08	2.19/2.54	4.89	4.25	4.19/4.11	1.36	7.39	13.73
C21	5.57	2.20/2.48	4.88	n.a.	n.a.	5.70	7.61	6.96/8.52
A22	6.29	2.73/2.97	4.87	4.46	4.16	7.70	8.40	6.46/7.85
T23	6.00	2.03/2.47	4.86	4.16	n.a.	1.47	7.25	13.82
C24	6.26	2.26	4.59	4.00	n.a.	5.65	7.58	n.o.

6dglc 18

H1	H2	H3	H4	H5	Met
3.86	2.95	2.74	2.62	2.99	0.71

Linker

H2'/H2''	H5'/H5''	H3'
3.28/3.56	3.92	4.31

[†] n.a. Not assigned. n.o. Not observed. No guanine amino protons could be identified.

Table S2. ¹H-NMR assignments of **helix 6dglc-G** (10 mM sodium phosphate, 150 mM NaCl, pH 7, T = 5°C).[†]

Oligonucleotide strands

helix 6dglc-G	H1'	H2'/H2''	H3'	H4'	H5'/H5''	H5/Met/H2	H6/H8	Imino/amino
C1	5.89	2.17/2.55	4.69	4.11	3.80	5.96	7.88	7.23/8.01
T2	5.50	2.26/2.48	n.a.	4.16	4.01	1.74	7.61	13.82
A3	6.10	2.91/2.78	n.a.	4.43	4.14/4.07	7.48	8.26	6.67/7.95
G4	5.59	2.52	n.a.	4.35	n.a.	---	7.77	12.75
C5	5.92	1.58/2.09	4.74	4.09	n.a.	5.44	7.12	6.47/8.13
G6	n.a.	2.48	n.a.	n.a.	n.a.	---	7.89	10.32
G7	6.00	2.77/2.82	n.a.	4.40	n.a.	---	7.93	12.70
T8	6.06	2.19/2.52	n.a.	4.23	4.15	1.34	7.32	n.o.
C9	5.56	2.18/2.47	4.86	4.17	4.08	5.69	7.60	6.92/8.49
A10	6.28	2.71/2.95	n.a.	4.45	4.15	7.66	8.38	6.44/7.82
T11	5.99	2.02/2.47	4.84	4.16	n.a.	1.46	7.24	n.o.
C12	6.25	2.25	4.58	4.00	4.20	5.69	7.58	n.o.
G13	5.70	2.65/2.82	4.88	4.23	3.71	---	7.95	12.68
A14	6.31	2.75/2.98	n.a.	4.48	4.23/4.15	7.99	8.36	6.27/7.78
T15	5.71	1.98/2.37	4.88	4.17	n.a.	1.37	7.13	13.51
G16	5.57	2.68/2.76	n.a.	n.a.	n.a.	---	7.87	12.51
A17	6.24	2.62/2.89	n.a.	n.a.	n.a.	7.86	8.17	n.o.
C18	6.00	2.14	4.84	n.a.	n.a.	5.30	7.35	6.59/8.04
6dglc19	---	---	---	---	---	---	---	---
G20	6.08	2.84/2.92	5.07	4.22	3.89/3.83	---	8.03	13.02
C21	5.94	2.07/2.43	4.80	4.18	n.a.	5.44	7.49	6.81/8.38
T22	5.42	2.12/2.33	4.84	4.11	4.02	1.71	7.44	13.84
A23	6.10	2.91/2.78	5.13	4.43	4.14/4.02	n.a.	8.26	6.78/8.02
G24	6.02	2.26/2.48	4.65	4.19	4.11	---	7.77	n.o.

6dglc 19

H1	H2	H3	H4	H5	Met
3.89	2.92	3.21	2.75	3.00	0.72

Linker

H2'/H2''	H5'/H5''	H3'
3.27	3.84	4.23

[†] n.a. Not assigned. n.o. Not observed. No guanine amino protons could be identified.

Table S3. Structurally relevant carbohydrate-DNA NOE contacts for **helix 6dglc-G** and **helix 6dglc-T**.

Carbohydrate-DNA NOEs	helix 6dglc-G	helix 6dglc-T
DNA-Glc	G20H1-Met: w C18H6-Met: w C18H5-Met: w C18H1'-Met: w C18Q2'-Met: w C5H41-Met: w C5H42-Met: w	G19H1'-Met: w G19H8-Met: vw G19H3-Met: w G8H3-Met: w G8H8-Met: vw C17H6-Met: vw C17H5-Met: w C17H1'-Met: w C17H1'-H2: vw T7Met-Met: m T7H6-Met: vw
DNA- Linker	G20H8-LinkH2': vw C18H1'-LinkH5': vw	G19H8-LinkH3': w G19H8-LinkH5': w G19H8-LinkH2': vw G19H8-LinkH2'': vw C17H1'-LinkH5': w C17H6-LinkH5': vw
Linker-Glc	LinkH5'-Met: vw LinkH3'-Met: w LinkH2'-Met: w	LinkH3'-Met: vw LinkH2'-Met: vw LinkH2''-Met: vw LinkH2'-H5: w LinkH2''-H5: w LinkH2'-H3: vw LinkH2''-H3: vw LinkH2'-H1: w LinkH2''-H1: w LinkH3'-H5: vw
Linker-Linker	LinkH2'- LinkH5': w LinkH2'- LinkH3': m	LinkH2'- LinkH5': w LinkH2''- LinkH5': w LinkH2'- LinkH3': w LinkH2''- LinkH3': w

Table S4. NMR structural constraints and calculation statistics.

Experimental distance constraints	helix 6dglc-G	helix 6dglc-T
Total number	265	409
Intra-residue	131	197
Sequential	104	157
Inter-strand	16	25
r.m.s.d.		
Backbone heavy atoms	$1.5 \pm 0.7 \text{ \AA}$	$0.8 \pm 0.3 \text{ \AA}$
Base heavy atoms	$1.1 \pm 0.4 \text{ \AA}$	$0.7 \pm 0.3 \text{ \AA}$
All heavy atoms	$1.3 \pm 0.6 \text{ \AA}$	$0.6 \pm 0.2 \text{ \AA}$
Residual violations		
Sum of violations (\AA)	10.1	24.9 \AA
Average & range	(10.7 – 9.5)	(25.5 – 24.3)
Maximum violation (\AA)	0.32	0.58
Average NOE energy (kcal/mol)	39.3	125.1
Range of NOE energies (kcal/mol)	37.2 – 42.0	122.1–126.3
Number of NOEs		
Glc-DNA	7	12
Linker-DNA	2	6
Glc-Linker	3	8
Linker-Linker	2	4

Table S5. Cartesian coordinates (in \AA) and ADF total energies (in kcal/mol) of all stationary points in this study, computed at BLYP-D3(BJ)/TZ2P using COSMO to simulate aqueous solution.**6dglc-G**

6dglc-G, bond energy: -5781.60 kcal/mol

1.C	1.553373	-0.693963	2.001937
2.H	0.961107	0.225173	2.035264
3.H	0.945801	-1.517504	2.397427
4.H	2.436595	-0.566815	2.636743
5.O	0.093093	0.112348	-0.489178

6.H	-0.891799	-0.024390	-0.495512
7.O	0.107444	-1.886991	-2.656886
8.H	-0.284002	-2.738252	-2.380522
9.O	2.702894	-3.059352	-3.032877
10.H	3.349215	-3.785294	-2.942433
11.O	2.748176	-2.219078	0.580544
12.O	-2.622667	0.027482	-0.694938
13.C	-2.419648	-0.086916	-4.282840
14.N	-1.469612	-0.553041	-5.137389
15.H	-0.754864	-1.179362	-4.782242
16.H	-1.733238	-0.657916	-6.109583
17.N	-3.517771	0.504447	-4.750371
18.C	-4.337165	0.946206	-3.766511
19.N	-5.537934	1.598134	-3.920432
20.H	5.948947	-2.688677	-0.553804
21.H	5.282471	-3.149560	1.047390
22.N	-5.229415	1.426918	-1.696879
23.O	4.095123	-3.678410	-0.611368
24.H	-6.972916	2.371402	-2.509941
25.C	3.309859	-2.520158	-0.711637
26.H	3.918683	-1.663135	-1.051643
27.C	2.179445	-2.806762	-1.712865
28.H	1.626282	-3.686218	-1.344807
29.C	1.239245	-1.605395	-1.786552
30.H	1.757877	-0.768303	-2.267395
31.N	-2.153929	-0.227379	-2.938269
32.H	-1.295044	-0.738505	-2.677526
33.C	-6.031180	1.862398	-2.654878
34.C	5.395971	-3.456015	0.002259
35.C	-4.155367	0.845972	-2.374476
36.C	-2.976348	0.214059	-1.890369
37.C	0.768500	-1.150217	-0.388661
38.H	0.094775	-1.914990	0.026216
39.C	1.971619	-0.986985	0.568012
40.H	2.605919	-0.170957	0.184788
41.H	-5.980025	1.839386	-4.800553
42.H	5.925402	-4.408450	-0.051234

glc-G

glc-G, bond energy: -5925.82 kcal/mol

1.C	1.531473	-0.875436	2.106333
-----	----------	-----------	----------

2.O	0.875554	0.399589	2.300914
3.H	0.873622	-1.701502	2.416134
4.H	2.423918	-0.882688	2.738885
5.O	-0.014257	-0.032391	-0.290804
6.H	-0.984922	-0.209814	-0.456170
7.O	0.105474	-1.895593	-2.609949
8.H	-0.263837	-2.764397	-2.358178
9.O	2.757094	-2.902871	-3.051671
10.H	3.409316	-3.627713	-3.013911
11.O	2.773746	-2.267786	0.606967
12.O	-2.656155	-0.148616	-0.762705
13.C	-2.367371	-0.038184	-4.342920
14.N	-1.401836	-0.459170	-5.203648
15.H	-0.705060	-1.119744	-4.877553
16.H	-1.644544	-0.498224	-6.186015
17.N	-3.450926	0.588308	-4.798245
18.C	-4.290143	0.976358	-3.808844
19.N	-5.483197	1.644131	-3.950400
20.H	5.985581	-2.552510	-0.526645
21.H	5.324146	-3.133061	1.037905
22.N	-5.228584	1.335777	-1.734707
23.O	4.172581	-3.608469	-0.661240
24.H	-6.945857	2.341983	-2.529324
25.C	3.346374	-2.479523	-0.704453
26.H	3.919773	-1.581020	-0.993089
27.C	2.225544	-2.755656	-1.720537
28.H	1.716106	-3.680165	-1.405132
29.C	1.223041	-1.599933	-1.728481
30.H	1.693621	-0.710929	-2.162935
31.N	-2.133472	-0.260209	-3.004002
32.H	-1.285035	-0.792125	-2.747397
33.C	-6.004382	1.835075	-2.682780
34.C	5.459553	-3.372412	-0.021919
35.C	-4.141965	0.789869	-2.421330
36.C	-2.979525	0.119086	-1.952732
37.C	0.742782	-1.260580	-0.303456
38.H	0.125097	-2.089914	0.072341
39.C	1.949969	-1.079740	0.644486
40.H	2.535293	-0.206967	0.311950
41.H	-5.902679	1.940839	-4.824450
42.H	6.025515	-4.299210	-0.126008
43.H	0.285233	0.508883	1.520163

6dglc-T

6dglc-T, bond energy: -5472.38 kcal/mol

1.H	3.193367	1.449423	-0.360756
2.C	6.309905	0.268397	-0.467993
3.C	3.566909	-0.560656	-0.380522
4.H	-4.286484	0.878373	-1.798914
5.H	4.391489	-2.427264	-0.413511
6.N	4.622933	-1.439199	-0.416949
7.C	7.746799	0.716355	-0.519789
8.N	3.945002	0.764093	-0.385093
9.H	8.419853	-0.145817	-0.537032
10.H	7.993623	1.337345	0.349912
11.C	5.941553	-1.041224	-0.461942
12.H	6.661026	-1.851104	-0.491523
13.C	-3.025128	5.600927	-0.626714
14.H	-3.651591	5.451549	0.262129
15.H	-3.488563	6.334421	-1.288474
16.O	-2.936694	4.364418	-1.389220
17.C	-2.440101	3.283256	-0.649792
18.H	-2.896563	3.253228	0.355515
19.C	-2.721069	1.982684	-1.406881
20.H	-2.261722	2.068771	-2.404685
21.C	-2.093250	0.796070	-0.656566
22.O	-1.008473	3.440470	-0.513114
23.H	-2.601777	0.681573	0.309136
24.C	-0.598108	1.023472	-0.396067
25.H	-0.059682	1.035139	-1.356971
26.C	-0.398944	2.396205	0.296615
27.H	-0.903774	2.360200	1.275493
28.C	1.067210	2.762466	0.481918
29.H	1.559890	2.014671	1.111918
30.H	1.162605	3.736440	0.972087
31.C	5.257648	1.274163	-0.425213
32.O	2.379991	-0.937047	-0.348943
33.O	5.432650	2.505135	-0.423007
34.O	-4.148544	1.807679	-1.530649
35.H	1.567777	2.815650	-0.494401
36.O	-0.107500	-0.042941	0.434751
37.H	0.809691	-0.271380	0.149699
38.O	-2.370576	-0.436760	-1.362635
39.H	-1.869808	-0.424791	-2.201234
40.H	-2.029580	5.944811	-0.326444
41.H	7.933948	1.324859	-1.413120

glc-T

glc-T, bond energy: -5617.61kcal/mol

1.H	3.417264	1.705725	0.063497
2.C	6.518743	0.432748	-0.121675
3.C	3.825846	-0.026863	-0.964578
4.H	-4.371765	0.516456	-0.877370
5.H	4.648887	-1.667552	-1.862628
6.N	4.871691	-0.856467	-1.294749
7.C	7.930033	0.692800	0.334072
8.N	4.179012	1.053708	-0.193295
9.H	8.606297	-0.084531	-0.033620
10.H	7.987155	0.718176	1.429539
11.C	6.170285	-0.634224	-0.889800
12.H	6.887207	-1.373845	-1.226662
13.C	-3.474356	5.462786	-0.495675
14.H	-3.758409	5.361599	0.559573
15.H	-4.217774	6.060184	-1.025704
16.O	-3.469633	4.156641	-1.139508
17.C	-2.631551	3.228067	-0.514234
18.H	-2.755294	3.262564	0.582929
19.C	-2.931736	1.826933	-1.049109
20.H	-2.805123	1.850426	-2.143126
21.C	-1.944703	0.816348	-0.440731
22.O	-1.259409	3.559314	-0.840195
23.H	-2.128216	0.752017	0.639225
24.C	-0.483283	1.239406	-0.660915
25.H	-0.255411	1.201416	-1.738038
26.C	-0.296222	2.701415	-0.177765
27.H	-0.457051	2.735364	0.913303
28.C	1.075541	3.314982	-0.497235
29.O	2.118210	2.889188	0.410551
30.H	0.973671	4.406660	-0.469464
31.C	5.468564	1.364591	0.269025
32.O	2.655212	-0.246239	-1.343507
33.O	5.642193	2.384407	0.963318
34.O	-4.289930	1.475818	-0.712234
35.H	1.410953	3.019151	-1.495750
36.O	0.361650	0.330939	0.060122
37.H	1.209391	0.213424	-0.438569
38.O	-2.231247	-0.511737	-0.938349
39.H	-2.006869	-0.534120	-1.888735
40.H	-2.489815	5.936007	-0.572358

41.H	1.832554	3.060725	1.325935
42.H	8.284046	1.666370	-0.027402

6dglc-A

6dglc-A, bond energy: -5616.87 kcal/mol

1.C	1.932190	-1.132459	2.376285
2.H	1.171714	-0.446386	2.763926
3.H	1.772750	-2.122491	2.821599
4.H	2.916529	-0.760216	2.679817
5.O	-0.576901	-0.808133	0.692904
6.H	-0.911375	-1.029367	1.579040
7.O	-0.732394	-2.559613	-1.607162
8.H	-2.126026	-1.791863	-4.536102
9.O	1.728921	-2.706194	-3.072252
10.H	2.552377	-3.162137	-3.330283
11.O	2.891862	-2.111584	0.401146
12.N	-2.577137	0.619622	-0.878824
13.C	-2.623582	-0.931705	-4.095340
14.C	1.842881	-1.210651	0.858082
15.H	2.010344	-0.214206	0.417448
16.H	-5.331493	1.746853	-5.650225
17.N	-3.463031	-0.251908	-4.889393
18.C	-4.027154	0.793577	-4.242105
19.N	-4.935869	1.715529	-4.717206
20.H	5.425316	-1.453334	-1.781941
21.H	5.550712	-2.216736	-0.163047
22.N	-4.544825	2.295430	-2.574224
23.O	4.038157	-2.988904	-1.410516
24.H	-5.901457	3.408388	-3.806194
25.C	2.978160	-2.145272	-1.038825
26.H	3.148169	-1.119906	-1.413505
27.C	1.685152	-2.726371	-1.628702
28.H	1.581297	-3.759845	-1.264679
29.C	0.469292	-1.912236	-1.175227
30.H	0.535259	-0.911987	-1.633037
31.N	-2.313704	-0.683746	-2.804583
32.H	6.072038	-3.117528	-1.619698
33.C	-5.210732	2.586681	-3.684473
34.C	5.354171	-2.396397	-1.225357
35.C	-3.794827	1.167611	-2.907929
36.C	-2.886579	0.376346	-2.165009

37.C	0.482921	-1.741081	0.353620
38.H	0.280290	-2.714878	0.819008
39.H	-2.974149	1.426873	-0.416231
40.H	-1.883973	0.056283	-0.381002
41.H	-1.323139	-1.861717	-2.031931

glc-A

glc-A, bond energy: -5760.59 kcal/mol

1.C	2.029877	-0.824807	2.307484
2.O	1.116610	0.150877	2.863258
3.H	1.936468	-1.787655	2.831823
4.H	3.041971	-0.439122	2.460154
5.O	-0.638745	-0.961206	0.963178
6.H	-1.366984	-1.570606	1.180659
7.O	-0.857025	-2.792849	-1.271278
8.H	-2.681132	-2.424561	-4.011743
9.O	1.483873	-2.902152	-2.957505
10.H	2.307238	-3.330407	-3.259137
11.O	2.889651	-1.845194	0.319627
12.N	-2.173408	0.735060	-0.978068
13.C	-2.924328	-1.398381	-3.747669
14.C	1.792688	-1.035164	0.806503
15.H	1.792465	-0.055648	0.300543
16.H	-5.106955	1.481484	-5.715041
17.N	-3.666886	-0.712366	-4.627587
18.C	-3.909658	0.548127	-4.201386
19.N	-4.633245	1.545527	-4.820835
20.H	5.114580	-1.150333	-2.171324
21.H	5.476017	-1.713180	-0.505560
22.N	-3.898626	2.438796	-2.885700
23.O	3.958514	-2.777796	-1.509124
24.H	-5.096361	3.569575	-4.259062
25.C	2.846036	-2.021740	-1.119061
26.H	2.859485	-1.026265	-1.597740
27.C	1.577143	-2.792631	-1.522110
28.H	1.621983	-3.791733	-1.063052
29.C	0.338505	-2.049247	-1.013231
30.H	0.286314	-1.081460	-1.537183
31.N	-2.432732	-0.974903	-2.562671
32.H	5.966439	-2.706302	-1.912345
33.C	-4.595637	2.648945	-3.995432

34.C	5.207373	-2.028240	-1.519356
35.C	-3.459700	1.119382	-2.999383
36.C	-2.677623	0.301335	-2.148991
37.C	0.479786	-1.775583	0.494159
38.H	0.475623	-2.729612	1.036977
39.H	-2.368032	1.680030	-0.673991
40.H	-1.624744	0.128683	-0.367538
41.H	-1.492110	-2.155868	-1.728082
42.H	0.229215	-0.080050	2.513135

6dglc-C

6dglc-C, bond energy: -5215.92 kcal/mol

1.H	4.220123	3.481263	-0.883081
2.C	6.115123	0.799535	-0.313176
3.C	3.586722	-0.301652	-0.738606
4.H	-4.016383	0.623446	-1.688860
5.H	4.576892	-2.094620	-0.421461
6.N	4.712700	-1.089482	-0.469970
7.O	-1.914030	-0.406217	-1.394477
8.N	3.737604	1.041323	-0.816650
9.H	-1.458220	-0.308907	-2.252756
10.H	-2.416919	5.931475	-0.188635
11.C	5.947940	-0.551086	-0.261312
12.H	6.750609	-1.248695	-0.057915
13.C	-3.373276	5.453637	-0.426256
14.H	-3.900710	5.195952	0.501239
15.H	-3.986844	6.126983	-1.027164
16.O	-3.168840	4.260958	-1.233398
17.C	-2.472808	3.243401	-0.565097
18.H	-2.856301	3.118745	0.463439
19.C	-2.609050	1.938082	-1.352084
20.H	-2.223908	2.116314	-2.369029
21.C	-1.776308	0.841595	-0.670949
22.O	-1.075381	3.604530	-0.508931
23.H	-2.208521	0.636324	0.316551
24.C	-0.311631	1.266817	-0.477453
25.H	0.173449	1.360111	-1.461652
26.C	-0.260193	2.645856	0.228413
27.H	-0.682599	2.523387	1.239372
28.C	1.145072	3.224290	0.312428
29.H	5.924699	3.406707	-0.531164

30.H	1.136398	4.168068	0.868237
31.C	4.950700	1.586911	-0.603816
32.O	2.476153	-0.885469	-0.895269
33.N	5.044272	2.931563	-0.675786
34.O	-4.002069	1.559540	-1.409218
35.H	1.539495	3.403438	-0.695115
36.O	0.365399	0.281957	0.315929
37.H	1.154061	-0.059543	-0.189667
38.H	7.084260	1.255769	-0.148762
39.H	1.812231	2.517693	0.813370

glc-C

glc-C, bond energy: -5358.52 kcal/mol

1.H	1.117460	2.616950	2.807804
2.C	5.798065	1.142028	-0.632465
3.C	3.453603	-0.371973	-0.588999
4.H	-3.849285	0.947453	-2.136733
5.H	4.703593	-1.848348	0.148934
6.N	4.688190	-0.888236	-0.180743
7.O	-1.942304	-0.354955	-1.625446
8.N	3.404551	0.901141	-1.050141
9.H	-1.354114	-0.262067	-2.399733
10.H	-1.849685	5.926788	-0.079530
11.C	5.833183	-0.148601	-0.199336
12.H	6.733101	-0.643687	0.143050
13.C	-2.810618	5.593267	-0.485183
14.H	-3.503409	5.369422	0.336060
15.H	-3.233225	6.367503	-1.127562
16.O	-2.635894	4.418501	-1.326208
17.C	-2.194751	3.288439	-0.626434
18.H	-2.751824	3.171092	0.320221
19.C	-2.363596	2.049287	-1.508041
20.H	-1.811893	2.225492	-2.445271
21.C	-1.777943	0.820007	-0.795450
22.O	-0.789609	3.459043	-0.324212
23.H	-2.377803	0.611543	0.099563
24.C	-0.322147	1.056275	-0.363130
25.H	0.307200	1.180393	-1.257893
26.C	-0.242392	2.365229	0.462001
27.H	-0.835496	2.243567	1.383883

28.C	1.191329	2.781192	0.839358
29.O	1.640317	2.211929	2.092133
30.H	1.225271	3.877632	0.881388
31.C	4.527675	1.646286	-1.068399
32.O	2.439851	-1.121689	-0.508269
33.N	4.420528	2.914786	-1.515134
34.O	-3.767268	1.856053	-1.786739
35.H	1.898108	2.423661	0.086873
36.O	0.136048	-0.061827	0.405917
37.H	0.997093	-0.385356	0.019917
38.H	6.691399	1.755096	-0.655760
39.H	3.529929	3.257819	-1.852232
40.H	5.227353	3.521293	-1.571873

6dglc-6dglc

6dglc-6dglc, bond energy: -6571.16 kcal/mol

1.H	-1.864818	-2.231176	2.269872
2.H	-0.210374	1.820896	-0.939832
3.H	-6.401935	1.544664	-0.740885
4.C	-1.098019	1.323729	-1.343265
5.H	-1.185270	-0.975844	0.232862
6.O	-5.430792	-0.313663	-0.578363
7.C	-2.055335	0.958975	-0.216563
8.O	-4.902625	-1.080105	2.106952
9.H	-4.550625	1.122602	0.624161
10.H	-3.673430	-1.818968	0.578917
11.H	-4.497737	-1.584943	2.838980
12.C	-6.103940	0.700215	-1.375524
13.O	-3.247296	0.394574	-0.831731
14.C	-2.556882	-0.431654	1.816331
15.H	-2.335226	1.865870	0.344118
16.H	0.457128	0.320553	0.939474
17.H	-2.768522	0.455598	2.425914
18.C	-1.475754	-0.070498	0.789861
19.C	-4.314128	0.170274	0.116501
20.H	-5.454032	1.054715	-2.182418
21.O	-2.133975	-1.436117	2.769001
22.H	-6.989573	0.219579	-1.794275
23.H	-1.584081	2.013050	-2.041457
24.H	-0.795973	0.424538	-1.896850
25.C	-3.858468	-0.882784	1.130143

26.O	-0.340961	0.461117	1.498329
27.H	3.672831	1.606682	-1.374778
28.C	3.157992	-0.285888	-0.481044
29.H	3.201424	-1.354614	-0.741490
30.C	3.933266	-0.070084	0.828613
31.H	3.899090	0.999571	1.088303
32.C	3.417398	-0.910858	1.987100
33.H	2.379662	-0.645896	2.214875
34.H	3.458258	-1.977271	1.733660
35.H	4.024516	-0.732262	2.880625
36.H	6.764904	0.864994	-2.845962
37.C	8.169992	0.377545	0.505087
38.O	7.285333	-0.054605	-0.567375
39.H	8.151618	1.471446	0.591440
40.H	7.874403	-0.077056	1.456456
41.O	5.315936	-0.457112	0.581200
42.O	3.008225	0.372502	-2.835703
43.C	3.773194	0.542043	-1.619235
44.H	3.136259	-0.542895	-3.152363
45.C	5.957848	0.369877	-0.417092
46.H	9.170760	0.042927	0.227045
47.C	5.256119	0.207608	-1.779833
48.O	5.820982	1.098027	-2.761464
49.H	5.371961	-0.840618	-2.097352
50.H	5.916237	1.424371	-0.090838
51.O	1.776935	0.110934	-0.305424
52.H	1.399736	0.177152	-1.204002

6dglc-glc

dglc-glc, bond energy: -6715.32 kcal/mol

1.H	-1.448289	-2.378187	1.870696
2.H	-0.448132	2.083572	-1.128434
3.H	-6.488170	1.341717	-0.275126
4.C	-1.300901	1.494341	-1.481041
5.H	-1.072554	-0.921194	-0.107895
6.O	-5.429554	-0.470766	-0.404066
7.C	-2.110195	0.975058	-0.299616
8.O	-4.557514	-1.432892	2.123365
9.H	-4.484683	0.908674	0.814440
10.H	-3.464132	-1.959293	0.415294
11.H	-4.045317	-1.960375	2.766859

12.C	-6.239616	0.571891	-1.016987
13.O	-3.323308	0.377261	-0.833322
14.C	-2.300253	-0.603859	1.648594
15.H	-2.384549	1.811853	0.362519
16.H	0.489921	0.648779	0.491811
17.H	-2.497304	0.216146	2.350266
18.C	-1.365835	-0.087233	0.549973
19.C	-4.264321	0.015355	0.202956
20.H	-5.710918	1.027976	-1.860615
21.O	-1.711247	-1.645835	2.460689
22.H	-7.151255	0.084897	-1.367417
23.H	-1.924125	2.150122	-2.097262
24.H	-0.938066	0.666086	-2.102090
25.C	-3.641636	-1.086851	1.064278
26.O	-0.199832	0.479260	1.183676
27.H	4.170620	2.020246	-1.179188
28.C	3.141125	0.218752	-0.593189
29.H	2.926420	-0.773278	-1.019790
30.C	3.822395	0.022663	0.774616
31.H	3.973768	0.999881	1.258681
32.C	3.077076	-0.913090	1.752783
33.O	2.123879	-0.239197	2.596680
34.H	2.600481	-1.733398	1.199586
35.H	3.835307	-1.340088	2.416341
36.H	7.118069	0.758596	-2.558196
37.C	8.086549	-0.328213	0.764849
38.O	7.220556	-0.486597	-0.395782
39.H	8.278067	0.736637	0.947370
40.H	7.631848	-0.783739	1.650828
41.O	5.103843	-0.618543	0.530499
42.O	3.351905	1.179001	-2.828361
43.C	4.037013	1.002952	-1.566609
44.H	3.290451	0.307801	-3.266638
45.C	5.995942	0.180309	-0.285132
46.H	9.019783	-0.837828	0.520307
47.C	5.406932	0.343599	-1.697782
48.O	6.241201	1.184651	-2.516514
49.H	5.306595	-0.657715	-2.144900
50.H	6.126134	1.168459	0.190526
51.O	1.901072	0.957952	-0.443761
52.H	1.662631	1.255915	-1.344380
53.H	1.255047	-0.169190	2.137671

glc-glc

glc-glc, bond energy: -6858.65 kcal/mol

1.H	-2.694748	-2.681878	1.996749
2.O	0.181719	1.599032	-0.850018
3.H	-6.047719	2.170269	-0.971447
4.C	-0.883888	0.740587	-1.341678
5.H	-1.537590	-1.492917	0.172757
6.O	-5.492105	0.142480	-0.987040
7.C	-1.970189	0.586706	-0.270538
8.O	-5.414528	-0.930746	1.637198
9.H	-4.426202	1.246827	0.398917
10.H	-4.201415	-1.807033	0.169046
11.H	-5.199904	-1.555593	2.356782
12.C	-5.877831	1.358266	-1.689807
13.O	-3.191175	0.328508	-1.009462
14.C	-2.973200	-0.758081	1.618282
15.H	-2.059787	1.533760	0.285203
16.H	0.222431	-0.428491	1.175256
17.H	-3.072341	0.118744	2.270365
18.C	-1.728268	-0.568541	0.739989
19.C	-4.370699	0.300550	-0.167807
20.H	-5.105588	1.653075	-2.408213
21.O	-2.857542	-1.877388	2.526691
22.H	-6.806293	1.125143	-2.213610
23.H	-1.343254	1.200503	-2.223681
24.H	-0.478577	-0.240292	-1.625680
25.C	-4.254894	-0.892815	0.780345
26.O	-0.631796	-0.290460	1.632728
27.H	3.640681	1.997853	-0.641972
28.C	3.195996	-0.014384	0.019659
29.H	3.018801	-0.997086	-0.445214
30.C	4.351078	-0.154917	1.038019
31.H	4.517807	0.820145	1.523505
32.C	4.076151	-1.206714	2.120776
33.O	3.100485	-0.749865	3.087253
34.H	3.747590	-2.145505	1.651247
35.H	5.001769	-1.395472	2.671177
36.H	6.183271	1.434175	-2.970329
37.C	8.358755	0.172784	-0.372136
38.O	7.174254	0.008507	-1.203751
39.H	8.477874	1.226901	-0.092525
40.H	8.286966	-0.446875	0.527807

41.O	5.547488	-0.570020	0.340731
42.O	2.559446	1.103246	-2.109101
43.C	3.580431	0.994748	-1.078879
44.H	2.523790	0.258920	-2.599028
45.C	5.991271	0.441487	-0.595576
46.H	9.204942	-0.149948	-0.980512
47.C	4.934036	0.628528	-1.696988
48.O	5.313938	1.681693	-2.601906
49.H	4.842729	-0.328064	-2.234859
50.H	6.150231	1.387308	-0.048894
51.O	2.018880	0.390743	0.757776
52.H	1.369505	0.867259	0.170436
53.H	2.366256	-0.355878	2.573535
54.H	0.905782	1.577871	-1.523790

A-T

A-T, bond energy: -4516.66 kcal/mol

1.N	0.953074	-0.173131	-0.001782
2.C	1.564765	-1.374502	-0.000672
3.N	2.882503	-1.630488	0.000722
4.C	3.611603	-0.490825	0.000327
5.C	3.107869	0.821593	-0.000430
6.C	1.699547	0.968609	0.000111
7.N	4.983389	-0.345077	0.000527
8.C	5.250532	1.007233	-0.001243
9.N	4.150404	1.749330	-0.001528
10.N	1.068048	2.156517	0.001525
11.H	0.896100	-2.232059	-0.001104
12.H	5.669077	-1.091957	0.000568
13.H	6.263281	1.383489	-0.002692
14.H	1.606540	3.012183	0.002432
15.H	0.041802	2.196175	0.000416
16.N	-1.886937	-0.124556	0.001772
17.C	-2.501750	-1.358569	0.000858
18.N	-3.883540	-1.299298	0.001245
19.C	-4.590540	-0.119273	-0.000437
20.C	-3.976279	1.095569	0.000128
21.C	-2.521902	1.116118	0.001880
22.O	-1.834684	2.166124	0.002302
23.O	-1.872956	-2.430038	0.000015
24.C	-4.730193	2.399465	-0.000827
25.H	-0.832719	-0.134286	0.002401
26.H	-4.376571	-2.186195	-0.000563
27.H	-5.668913	-0.228913	-0.002827
28.H	-4.473985	2.999888	0.880865
29.H	-5.809605	2.221945	-0.001526
30.H	-4.472756	2.999403	-0.882462

G-C

G-C, bond energy: -4428.07 kcal/mol

1.N	-3.138905	-0.533417	0.301547
2.O	0.787734	1.926324	0.293581
3.N	-1.217080	0.776625	0.296857
4.C	-2.596844	0.700532	0.299541
5.N	-3.388049	1.781137	0.300913

6.C	-2.697272	2.943164	0.300383
7.C	-1.303262	3.122259	0.298829
8.C	-0.470731	1.966679	0.296040
9.N	-3.218551	4.217641	0.302738
10.C	-2.152640	5.100847	0.302832
11.N	-0.983022	4.483219	0.300384
12.H	-2.565481	-1.390165	0.296459
13.H	-4.147561	-0.606763	0.300839
14.H	-0.670332	-0.115516	0.294715
15.H	-4.202521	4.462177	0.303016
16.H	-2.304616	6.170343	0.302962
17.O	-1.631742	-2.939090	0.291093
18.N	2.286200	-0.508448	0.295328
19.N	0.309433	-1.697873	0.292016
20.C	2.421262	-2.917040	0.302948
21.C	-0.373894	-2.872597	0.294217
22.N	0.359215	-4.063542	0.300143
23.C	1.723791	-4.085148	0.304152
24.C	1.663773	-1.697264	0.296687
25.H	3.504588	-2.902767	0.307497
26.H	-0.167683	-4.931232	0.301200
27.H	2.189401	-5.062552	0.309840
28.H	1.746423	0.369027	0.293555
29.H	3.296705	-0.469697	0.297951

References

1. Stevens, K. & Madder, A. Furan-modified oligonucleotides for fast, high-yielding and site-selective DNA inter-strand cross-linking with non-modified complements. *Nucleic Acids Res* **37**, 1555-1565 (2009).
2. Schlegel, M.K. & Meggers, E. Improved phosphoramidite building blocks for the synthesis of the simplified nucleic acid GNA. *J Org Chem* **74**, 4615-4618 (2009).
3. te Velde, G., Bickelhaupt, F. M., Baerends, E. J., Fonseca Guerra, C., van Gisbergen, S. J. A., Snijders, J. G. & Ziegler, T. Chemistry with ADF. *J. Comput. Chem.* **22**, 931-967 (2001).
4. Fonseca Guerra, C., Snijders, J.G., te Velde, G. & Baerends, E.J. Towards an order-N DFT method. *Theor. Chem. Acc.* **99**, 391-403 (1998).
5. Computer code ADF, Version 2013.01: E. J. Baerends, J. Autschbach, A. Bérces, J. A. Berger, F. M. Bickelhaupt, C. Bo, P. L. de Boeij, P. M. Boerrigter, L. Cavallo, D. P. Chong, L. Deng, R. M. Dickson, D. E. Ellis, M. van Faassen, L. Fan, T. H. Fischer, C. Fonseca Guerra, S. J. A. van Gisbergen, J. A. Groeneveld, O. V. Gritsenko, M. Grüning, F. E. Harris, P. van den Hoek, C. R. Jacob, H. Jacobsen, L. Jensen, E. S. Kadantsev, G. van Kessel, R. Klooster, F. Kootstra, E. van Lenthe, D. A. McCormack, A. Michalak, J. Neugebauer, V. P. Nicu, V. P. Osinga, S. Patchkovskii, P. H. T. Philipsen, D. Post, C. C.

- Pye, W. Ravenek, P. Romaniello, P. Ros, P. R. T. Schipper, G. Schreckenbach, J. Snijders, M. Solà, M. Swart, D. Swerhone, G. te Velde, P. Vernooijs, L. Versluis, L. Visscher, O. Visser, F. Wang, T. A. Wesolowski, E. M. van Wezenbeek, G. Wiesenekker, S. K. Wolff, T. K. Woo, A. L. Yakovlev, T. Ziegler, *Theoretical Chemistry*, (SCM, Vrije Universiteit: Amsterdam, The Netherlands,).
6. Fonseca Guerra, C., van der Wijst, T., Poater, J., Swart, M. & Bickelhaupt, F.M. Adenine versus guanine quartets in aqueous solution: dispersion-corrected DFT study on the differences in π -stacking and hydrogen-bonding behavior. *Theor. Chem. Acc.* **125**, 245-252 (2010).
 7. Klamt, A. & Schuurmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2*, 799-805 (1993).
 8. Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J Phys Chem* **99**, 2224-2235 (1995).
 9. Pye, C.C. & Ziegler, T. An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package. *Theor. Chem. Acc.* **101**, 396-408 (1999).
 10. Fyngenson, D.K. & Goodman, M.F. Appendix. Gel kinetic analysis of polymerase fidelity in the presence of multiple enzyme DNA encounters. *J Biol Chem* **272**, 27931-27935 (1997).
 11. Moran, S., Ren, R.X. & Kool, E.T. A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc Natl Acad Sci U S A* **94**, 10506-10511 (1997).