CHEMISTRY
A European Journal

A Journal of
ChemPubSoc Europe

Accepted Article

Title: Base Pairing Properties of Double-Headed Nucleotides

Authors: Mick Hornum, Julie Stendevad, Pawan Kumar Sharma, Pawan Kumar, Rasmus Blaaholm Nielsen, Michael Petersen, and Poul Nielsen

To be cited as: *Chem. Eur. J.* 10.1002/chem.201901077

Link to VoR: http://dx.doi.org/10.1002/chem.201901077

Supported by
ACES

WILEY-VCH

10.1002/chem.201901077

WILEY-VCH

FULL PAPER

Base-Pairing Properties of Double-Headed Nucleotides

Mick Hornum,[a] Julie Stendevad,[a] Pawan K. Sharma,[b] Pawan Kumar,[a] Rasmus B. Nielsen,[a] Michael Petersen[a] and Poul Nielsen*[a]

Dedication ((optional))

Abstract: Nucleotides that contain two nucleobases (double-headed nucleotides) have the potential to condense the information of two separate nucleotides into one. This presupposes that both bases must successfully pair with a cognate strand. Herein, we develop and examine in full detail double-headed nucleotides that feature cytosine, guanine, thymine, adenine, hypoxanthine and diaminopurine linked to the C2′-position of an arabinose scaffold. These monomeric units were efficiently prepared via convergent synthesis and incorporated into DNA oligonucleotides by means of the automated phosphoramidite method. Their pairing efficiency were assessed by UV-based melting temperature analysis in several contexts and extensive molecular dynamics studies. Altogether, our results show that these double-headed nucleotides have a well-defined structure and invariably behave as functional dinucleotide mimics in DNA duplexes.

Introduction

Besides its renowned role in heredity, DNA has emerged as a valuable tool for a wide range of purposes. Well-studied examples include therapeutics,[1–3] diagnostics,[4] genetic engineering,[5] forensics,[6] phylogenetics,[7] data storage,[8,9] nanostructures,[10,11] nanomechanical devices,[12,13] etc. In most cases, the versatility of DNA arises from its predictable, yet malleable, structure and programmable assembly. In addition, chemically modified DNA is a very mature field of research, and thousands of synthetic derivatives have been developed to suit specific needs, such as tailored physical properties,[14] enhanced drug pharmacokinetics[15] or in the desire to construct genetic systems based on alternative chemical platforms.[16,17] A unique type of DNA modification is the so-called double-headed nucleotides, which are single nucleotides holding two nucleobases.[18–23] If both nucleobases are oriented towards the duplex core for partaking in base pairing, the double-headed nucleotide is turned into a functional dinucleotide. Such bifurcated structures can function to condense the genetic information in DNA by comprising two Watson–Crick base pairs per each nucleotide unit.[24] This design can be used to e.g. improve cellular uptake, since the same molecular information can be delivered using a shorter sequence and decreased polyanionic charge.

The C2′-*exo* position of β-nucleosides emerges as the most reasonable site for introducing an additional base and enable dual Watson–Crick contacts.[18] In our preceding studies, we have gauged the potential of three designs (double-headed β-D-ribosides,[25] β-D-deoxyribosides[26,27] and β-D-arabinosides;[24,28,29] see Figure 1) in terms of their synthetic feasibility and dinucleotide potential. For convenience, uracil has invariably been used as the C1′-linked base.

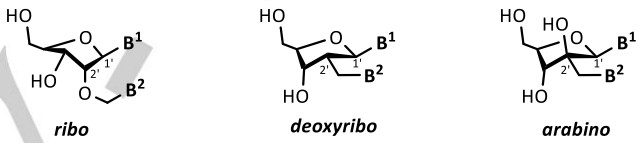


Figure 1. Three double-headed nucleoside variants. B = nucleobases.

Incorporation of the double-headed *ribo* structure turned out to destabilize DNA duplexes[25] indicating that the C2′-linked nucleobase is not well-accommodated. Contrarily, the *deoxyribo* structure displayed potent recognition of complementary sequences indicative of dinucleotide behavior.[26,27] This highlighted the importance of using a shorter linker and/or the importance of the sugar puckering. Nevertheless, this building block could only be obtained following a long synthetic route with poor yielding steps, which foiled its development. In the present work, we focus on the *arabino* structure, which is essentially isosteric to the *deoxyribo* structure, yet we have shown that it can be obtained in remarkably improved yields over fewer steps.[24,28] Favorably, the 2′-OH group is shielded and does not necessitate a protecting group for oligonucleotide synthesis.[28]

In this report, we take advantage of the synthetic ease offered by the *arabino* design to gain a deeper insight into double-headed nucleotides' base-pairing efficiency compared to natural dinucleotides. Accordingly, we set out to investigate a full set of canonical and modified DNA bases linked to the C2′-position (Figure 2). Additionally, we accompaniment the experimental work with atomistic molecular dynamics computations for assessing the double-headed nucleotides' binding mode and impact on the helix's tertiary structure. This study presents the first synthesis of double-headed nucleotides containing hypoxanthine (H) and 2,6-diaminopurine (D). As the nucleobase of inosine found in the wobble position of many tRNAs, H has often been exploited as a universal base in various applications such as hybridization probes[30–32] or primers.[33] In the light of its promiscuous base pairing, we rationalized that H could potentially serve as an indicator for better understanding how the additional nucleobase is positioned in the base stack. The weakly fluorescent D, a selective

[a] Dr. M. Hornum, J. Stendevad, Dr. P. Kumar, R. B. Nielsen, Dr. M. Petersen, Dr. P. Nielsen
Department of Physics, Chemistry and Pharmacy
University of Southern Denmark
Campusvej 55, DK-5230 Odense M
E-mail: pouln@sdu.dk
[b] Dr. P. K. Sharma
Department of Chemistry
Kurukshetra University
Kurukshetra, Haryana 136119, India

Supporting information for this article is given via a link at the end of the document.

10.1002/chem.201901077

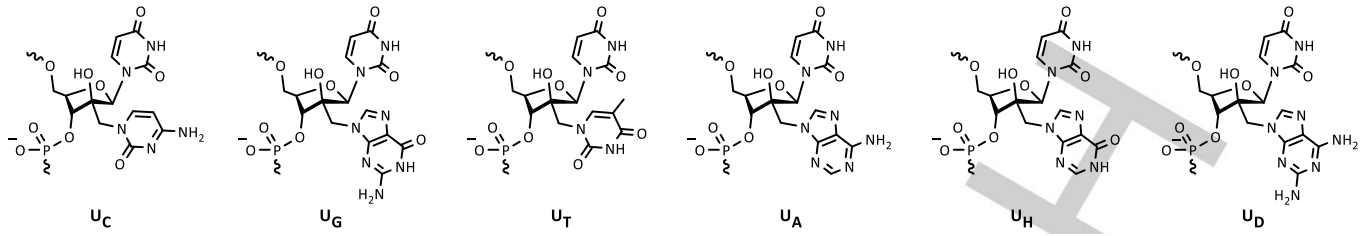


Figure 2. DNA building blocks of this study.

analog of A, has often been introduced to improve binding affinity by potentially allowing three hydrogen bonds to T/U,[34,35] rendering it useful in e.g. primers[36] or hybridization probes.[37] However, this enthalpic gain is sometimes offset by the entropic penalty caused by disruption of the spine of hydration around the minor groove of DNA duplexes.[38] Also, the practical use of D in duplexes is hampered by its susceptibility to depurination during oligonucleotide synthesis.[39] For the present case, depurination of the second nucleobase is irrelevant, and we conjectured that D could be advantageous in improving duplex stability (when paired-up with T) and recognition fidelity. In this regard, the extra 2-amino group compared to A can serve as a useful indicator of base pairing through Watson–Crick faces.

Results

Synthesis

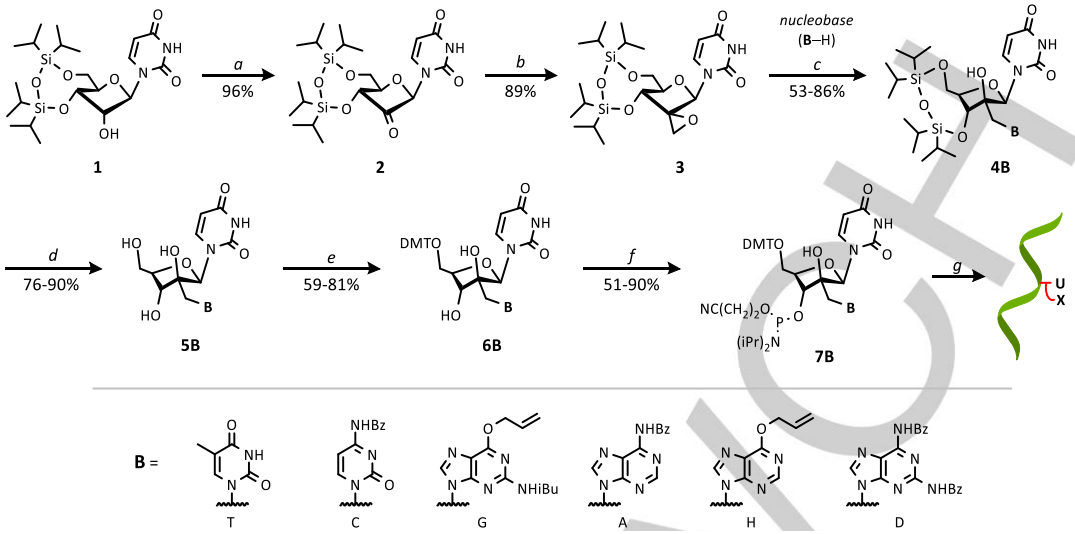
The double-headed nucleosides (Figure 2) were constructed according to our previously described methodology,[24,28] and incorporated into oligonucleotides using conventional phosphoramidite chemistry. The syntheses of the phosphoramidite building blocks are illustrated in Scheme 1. The mild TEMPO-based[40,41] oxidation of 3′,5′-*O*-(tetraisopropyldisiloxane-1,3-diyl)uridine **1** followed by Corey–Chaykovsky epoxidation[42] of the 2′-keto-uridine intermediate **2** were used to obtain the 2′-spiro-epoxy nucleoside **3**. Both reactions proceed stereospecifically and in high yields. This 2′-spiroepoxy nucleoside is the central branch point for the generation of all the double-headed nucleotides; the epoxide is readily opened by nucleophilic bases under alkaline conditions. All six nucleobases are alkylated with good-to-exclusive N-1 (for pyrimidines) and N-9 (for purines) selectivity. Moreover, selectivity was greatly enhanced using protecting groups that were anyway required for oligonucleotide synthesis. Protection of the exocyclic amines were realized with standard benzoyl (C, A and D) and isobutyryl (G) groups. In addition, the G and H bases were found to require installation of 6-*O* allyl groups[43,44] to reduce the nucleophilicity of the N-3 atom, which appeared to otherwise react intramolecularly with the P(III) center once activated during the coupling step of the oligonucleotide synthesis (see the Supporting Information p. 12 for details and discussion). The isolated yields for the introduction of the second nucleobases varied from 53% (**4T**) to 83% (**4H**). After formation of the double-headed structure, the silyl ethers were removed using TBAF to obtain the corresponding sugar-deprotected double-headed nucleotides **5**. In the case of **4D**, the use of $Et_3N{\cdot}3HF$ was superior to TBAF, since residues of the latter adhered strongly to the nucleoside. Excess $Et_3N{\cdot}3HF$ was neatly decomposed into volatiles (acetylene, TMSF and Et_3N) when treated with trimethylsilylacetylene. Finally, the 5′ and 3′ hydroxyls were routinely dimethoxytritylated and phosphitylated. Due to solvation difficulties with **5D** in pyridine and thus reduced reactivity, a DMSO/2,6-lutidine solvent system[45] was used to significantly improve the rate of tritylation. In this way, all nucleoside phosphoramidites **7** were successfully obtained in overall yields of 27% ($\mathbf{U_T}$), 40% ($\mathbf{U_C}$), 25% ($\mathbf{U_A}$), 30% ($\mathbf{U_G}$), 40% ($\mathbf{U_H}$) and 24% ($\mathbf{U_D}$) starting from **3**.

All double-headed phosphoramidites **7** were successfully incorporated into oligonucleotides on an automated solid-phase DNA synthesizer when activated with 1*H*-tetrazole. Coupling efficiencies exceeded 95% for all monomers. After completion of the oligonucleotide synthesis, the *O*-allyl protecting groups on $\mathbf{U_G}$ and $\mathbf{U_H}$ were successfully removed by treating the solid-support bound oligonucleotides with a mixture of $Pd(PPh_3)_4$ and $Et_2NH{\cdot}HCO_3$ in CH_2Cl_2 following a protocol by Eschenmoser et al.[44] All oligonucleotides were deprotected and cleaved from the solid support by treatment with ammonium hydroxide for 24 h at rt. In the case of $\mathbf{U_D}$, we observed slow deprotection of the benzoyl protecting groups as expected.[46] Incubation in ammonium hydroxide at 65 °C[47] or a methylamine/ ammonia mixture[48] caused substantial depyrimidination of the $\mathbf{U_D}$ monomer in our case (see the Supporting Information p. 13 for discussion). Instead, the use of anhydrous ammonia in methanol and incubation at 65 °C for 14 days successfully gave the fully deprotected oligonucleotides. The final crude oligonucleotides were purified by ion-exchange chromatography followed by desalting to provide the final oligonucleotides in high yield and purity.

11-mer duplexes containing a twelfth base pair

To gauge the quality of the double-headed nucleotides' base pairing, they were first analyzed in undecamers containing an A, T and U rich core flanked by CGC/GCG trinucleotides to minimize fraying effects (Figure 3). When any two of the double-headed nucleotides are placed in a so-called +1 interstrand zipper arrangement, the two C2′-linked nucleobases come in spatial contact, and a twelfth base pair may ensue (Figure 3A). On basis of the six different C2′-linked nucleobases, this twelfth base pair ($\mathbf{X}{\cdot}\mathbf{Y}$) can assume a total of $6^2 = 36$ possible combinations. The influence of this twelfth base pair was studied by melting temperature (T_m) experiments and compared with a reference duplex without the C2′-linked bases (Figure 3B, U = 2′-deoxyuridine), as well as duplexes where the double-headed nucleotides are

10.1002/chem.201901077



Scheme 1. *Reagents and conditions*: *(a)* TEMPO, IBD, AcOH, CH_2Cl_2, rt; *(b)* $(CH_3)_3SOI$, NaH, THF, DMSO, rt; *(c)* Bases **B**–H, *and* NaHMDS, THF, 55 °C (**4C**), *or* NaH, DMF, 35 °C (**4G**,**H**), *or* NaH, DMF, 110 °C (**4T**,**D**), *or* KHMDS, THF, 55 °C (**4A**); *(d)* TBAF, THF, rt (**5C**,**G**,**T**,**A**,**H**), *or* $Et_3N{\cdot}3HF$, THF, rt (**5D**); *(e)* DMTCl, pyridine, rt (**6C**,**G**,**T**,**A**,**H**), *or* DMTCl, 2,6-lutidine, DMSO, rt (**6D**); *(f)* 2-Cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite, iPr_2NEt, CH_2Cl_2, rt; *(g)* Incorporation into DNA oligonucleotides. Deprotection conditions: Conc. aq. NH_3, 24 h, rt ($\mathbf{U_T}$,$\mathbf{U_A}$,$\mathbf{U_C}$), *or* $Et_2NH{\cdot}HCO_3$, $Pd(PPh_3)_4$, PPh_3, CH_2Cl_2, 50 °C, *then* conc. aq. NH_3, 24 h, rt ($\mathbf{U_G}$,$\mathbf{U_H}$), *or* conc. NH_3 in MeOH, 14 days, 65 °C ($\mathbf{U_D}$).

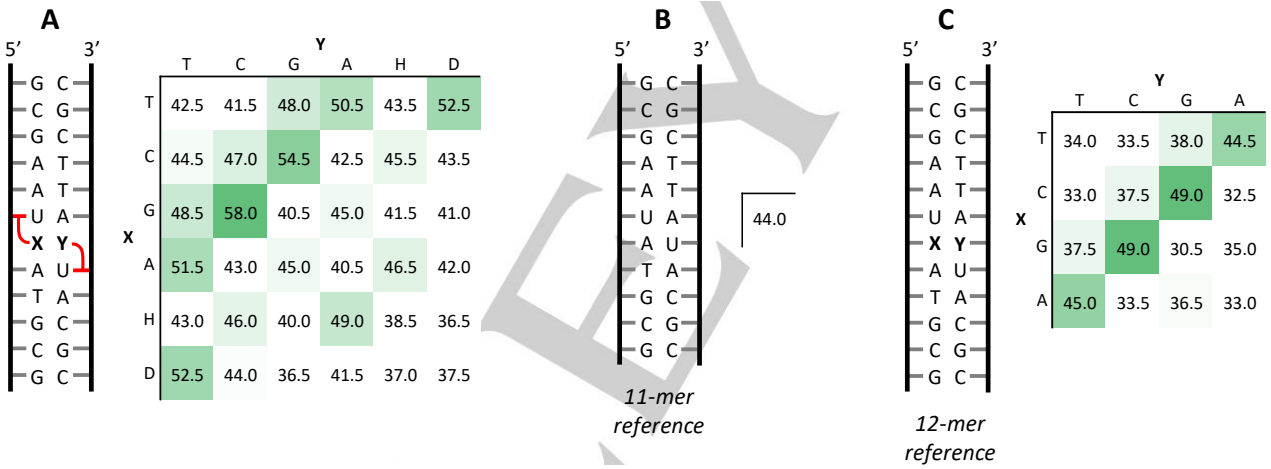


Figure 3. 11-mer DNA duplexes containing a twelfth base pair (**X**·**Y**) arising from the C2′-linked bases (A), and the corresponding regular 11- and 12-mer regular DNA duplexes (B,C). Matrices comprise the measured Tm values (°C) at 1.5 μM concentrations of each DNA strand in a medium salt buffer (2.5 mM Na_2HPO_4, 5.0 mM NaH_2PO_4, 100 mM NaCl, 0.1 mM EDTA) at pH 7.0.

swapped for their corresponding natural deoxydinucleotide pairs (Figure 3C).

For the T_m experiments, DNA duplexes were formed by annealing the constituent oligonucleotides in neutral medium salt buffer using 1.5 μM concentrations of each strand. The duplexes' T_m values were derived from the 260 nm UV absorption melting curves. We observed consistency between duplicate melting curve recordings, and annealing curves were all within 1.0 °C of the corresponding melting curves. The measured T_m values (rounded to nearest half) are listed in the matrices of Figure 3. As indicated, the unmodified 11-mer reference has an experimental T_m value of 44.0 °C (Figure 3B). The 36 different modified 11-mer duplexes have T_m values in the range of 36.5–58.0 °C (Figure 3A). The most stable duplexes ensued by placing C2′-linked C and G in contact (T_m = 54.5–58.0 °C) or C2′-linked T and A/D in contact (T_m = 50.5–52.5 °C). It follows that intercalation of additional base pairs in the duplex is highly thermostabilizing (up to +14 °C), and Watson–Crick base combinations are favored. Non-Watson–Crick pairs generally have a neutral or destabilizing effect relative to the 11-mer reference duplex except for the three classical wobble pairs H·A, H·C and G·T (which are weakly stabilizing by +1.5 to +5.0 °C) as well as the C·C pair (+3.0 °C).

Remarkably, the modified 11-mer duplexes with matched Watson–Crick base pairing were also notably more stable (+5.5 to +9.0 °C) than the corresponding genuine 12-mer duplexes (compare Figure 3A with 3C). Specifically, the T_m value is increased by 5.5–9.0 °C and 6.0–6.5 °C when the twelfth base pair is a C2′-linked G·C or T·A pair, respectively. Hence, it is very

beneficial for the DNA duplex to convert one of the standard base pairs into a C2′-linked base pair. This net stabilizing effect can be rationalized in terms of reduced torsional freedom and electrostatic repulsion of the helix backbone as the number of phosphates is reduced. Indeed, the T_m vales of all the modified 11-mer duplexes—and not only the fully Watson–Crick matched duplexes—are elevated non-specifically by 5.5–11.5 °C when compared to the 12-mer reference. The base pairing fidelity of the twelfth pair is in general slightly compromised by the double-headed nucleotides. The most tolerated non-Watson–Crick base pair is G·T with a discrimination of only 2.5–3.0 °C relative to the A·T matched pair (Figure 3A), while the discrimination is 6.5–7.5 °C in the native duplex (Figure 3C). Relative to the G·C pair, the G·T is discriminated by 6.5–9.5 °C, but 11.0–11.5 °C in the native duplexes. The least tolerated are the purine-purine base pairs G·G, H·H, D·D, D·H, G·D and G·H in the modified duplex. The most discriminating residues are **U$_D$** and **U$_C$** with discriminations of no less than 8.5 °C and 7.5 °C, respectively. **U$_D$** is particularly interesting, since it demonstrates much better base pairing specificity than **U$_A$**, and it binds more strongly to **U$_T$** (by 1–2 °C). Conversely, **U$_H$** is the most promiscuous base, as expected, with a slight binding preference for **U$_A$** (T_m = 46.5–49.0 °C), and least preference for other purines **U$_G$**, **U$_D$** and **U$_H$** (T_m = 36.5–41.5 °C). Masked as a T analog, H discriminates A over G better than T (5.0–9.0 *vs* 2.5–3.0 °C), since it cannot partake in wobble pairing with G.

Double-headed nucleotides behaving as condensed dinucleotides

Next, to explore the actual dinucleotide behavior of the double-headed nucleotides (**U$_C$**, **U$_G$**, **U$_A$**, **U$_T$**, **U$_D$** and **U$_H$**), they were placed across two natural nucleotides. For this study, we turned to DNA duplexes of 14 base pairs with a single perturbation site (Figure 4). An asymmetric, nonpalindromic sequence was chosen to diminish loop formation or mispairing. The measured T_m values from this study are displayed on the horizontal axes in Figure 4. Here, combinations of **X**:**Y** are arranged vertically; sorted in descending order of T_m values. In the modified duplex (Figure 4A), **X** corresponds to a C2′-linked nucleobase whereas in the unmodified duplex (Figure 4B), **X** takes up a normal 2′-deoxyribonucleotide. Thus, the specific base-pairing properties of each of the six double-headed nucleotides can be compared directly with that of the corresponding genuine dinucleotides

As seen in the unmodified duplex (Figure 4B), the base pairing of **X**·**Y** is dictated by Watson–Crick pairing (C·G, T·A) as expected. Likewise, the C2′-linked nucleobases favor Watson–Crick base combinations (Figure 4A). Notably, the duplex T_m values relating to **U$_C$** are all lifted by 1 °C in relation to the natural contexts, thus retaining the same level of mismatch discrimination (14–16 °C). In the case of **U$_G$**, the T_m value of the fully-matched duplex is intact (T_m = 55.0 °C) compared to the native duplexes, and the base-pairing specificity is improved by 1.0–2.0 °C. Accordingly, both **U$_C$** and **U$_G$** appear as flawless analogs of 5'-UC and 5'-UG dinucleotides. While **U$_T$** and **U$_A$** also favor Watson–Crick base pairing, these fully-matched duplexes are slightly destabilized in comparison to 5′-UT or 5′-UA (compare Figure 4A with 4B for X = T,A); the T_m values are lowered from 51.0–51.5 °C to 48.0–48.5 °C. In addition, the stabilities of two duplexes containing the mismatched base pairs T·G and A·G are slightly increased. These two combinations causes inferior discrimination of G. Specifically, the T·G and A·G mismatches are discriminated by only 1.5 °C (down from 6.5 °C) and 1.0 °C (down from 4.5 °C) relative to matched duplexes with T·A and A·T. Less important reductions in the mismatch discrimination are seen in two cases, i.e. the T·T mismatch relative to the T·A match, which is discriminated by 6.0 °C (down from 9.5 °C), as well as the mismatch A·C relative to the matched A·T, which is discriminated by 5.0 °C (down from 8.5 °C). Altogether, **U$_T$** and **U$_A$** clearly behave as dinucleotide mimics with predilection for Watson–Crick base pairing, however, mismatches are moderately tolerated.

In the unmodified duplex with H and D (Figure 4B), the H·C and D·T pairs are the thermally most stable configurations with T_m values of 49 °C and 50 °C, respectively. Thus, H and D behave mostly as G and A analogs, respectively, in standard settings. Yet, H display selectivities of only 3.0–5.5 °C, thereby confirming its degeneracy. In the duplex with **U$_H$** (Figure 4A), a slight preference for base pairing to A (T_m = 48 °C) over C (T_m = 47 °C) was observed, which means that a C2′-linked H more take the role of a T analog rather than a G analog. Consequently, the relative binding preference to A versus C is reversed compared to the unmodified duplex, where H has preferential binding to C (T_m = 49 °C) and binding to A is somewhat less stable (T_m = 46 °C). In addition, H·G and H·T mismatches (relative to H·A) are better discriminated with **U$_H$** (5.0 °C) than 5′-UH (2.0–2.5 °C). In the duplex with **U$_D$**, preferential base pairing to T is dominant (T_m = 49.5 °C) like in the reference duplex, and it confirms that C2′-linked D is a potent A analog. The improvement of +1.0 °C in the T_m compared to **U$_A$** matches the typical thermal gain of replacing an A·T pair with D·T,[35] although this gain is not observed in the unmodified duplex. Notably, the improved base-pairing fidelity of **U$_D$** compared to **U$_A$** is distinct. Decreases in the T_m of 5.5–9.5 °C are observed for duplexes mismatches opposite to the C2′-linked D, and mismatches are thus less tolerated than for **U$_A$** (1.0–8.0 °C).

Since **U$_A$** and **U$_D$** carry two complementary bases on the same unit, we also examined their abilities to pair with themselves in what is essentially 14 bp duplexes made from two 13-mer oligonucleotides (Figure 4C). In our preliminary report,[28] we found that this **U$_A$**·**U$_A$** pair increases the T_m by 4.5 °C relative to the regular 14-mer duplex containing 5′-UA:5′-UA (T_m = 51.0 °C), and 6.5–7.5 °C compared to the corresponding singly-modified duplexes.[28] It follows that **U$_A$** binds to itself much more strongly than to the natural dinucleotide. Nonetheless, our present results indicate that **U$_D$** is markedly inferior to **U$_A$** in this regard. In fact, the **U$_D$**·**U$_D$** pair was found to increase the duplex stability by only 1.0 °C (T_m = 52.0 °C, Figure 4C) compared to the reference duplex. This is in the light of the six potential hydrogen bonds in **U$_D$**·**U$_D$** instead of just four in **U$_A$**·**U$_A$**. The explanation for this discrepancy might in fact come from the reinforcement of base pairing, which reduces the flexibility of the DNA or disrupts the hydration of the minor groove. However, it is still of note that the duplex with self-pairing **U$_D$** is more stable than the unmodified duplex with the same number of regular base pairs (by +1 °C), and additionally increases the duplex by 3–4 °C compared to the constituent

10.1002/chem.201901077

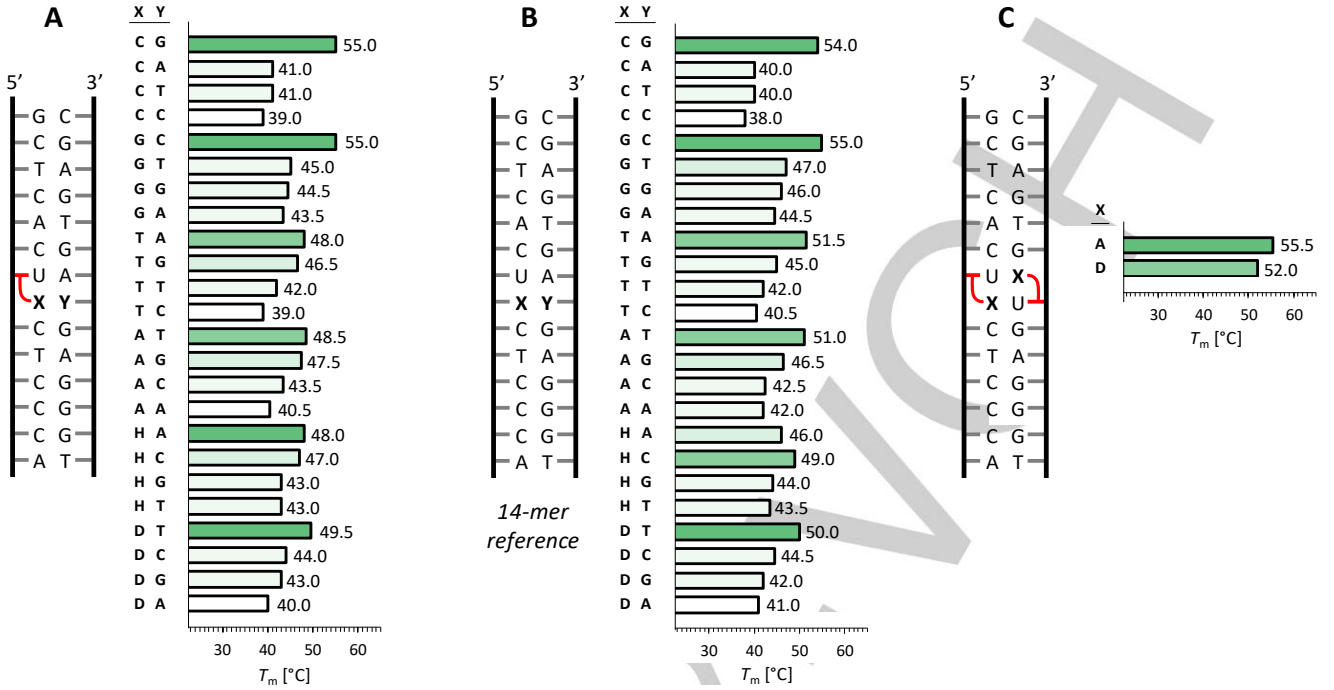


Figure 4. Double-headed nucleotides (**U**$_X$) behaving as condensed dinucleotides in 14-mer DNA across different combinations (3′-A**Y**) of natural dinucleotides (A), and the corresponding native DNA duplexes (B). In addition, data for two duplexes containing self-pairing **U**$_A$ and **U**$_D$ nucleotides are shown to the right (C). Bars representing data from the mismatched base pairs (**X**·**Y**) are shown in lighter gray color. T_m values were measured in medium salt buffer (2.5 mM Na_2HPO_4, 5.0 mM NaH_2PO_4, 100 mM NaCl, 0.1 mM EDTA) at pH 7.0 containing 1.5 μM concentrations of each DNA strand.

singly-modified duplexes containing **U**$_D$:3′-AU (T_m = 48.0 °C) and 5′-UA:**U**$_D$ (T_m = 49.0 °C) (data not shown).

Base-pairing properties of double-headed nucleotides at the atomic level

To gain deeper structural insights into the dinucleotide behavior of the double-headed nucleotides, we selected a subset of the constructed 14-bp DNA duplexes for atomistic molecular dynamics (MD) simulations. Understanding the degree of helix perturbation in combination with the mode of base pairing enabled us to better rationalize the observed trends in the T_m values. Accordingly, we focused our in-silico exploration on the nine duplexes that contain **U**$_A$, **U**$_T$, **U**$_D$ and **U**$_H$ in their capacities as compressed dinucleotides across matching bases (i.e. **U**$_A$:3′-AT, **U**$_T$:3′-AA, **U**$_D$:3′-AT, **U**$_H$:3′-AA and **U**$_H$:3′-AC) or across a G mismatch (i.e. **U**$_A$:3′-AG, **U**$_T$:3′-AG, **U**$_D$:3′-AG and **U**$_H$:3′-AG), respectively. The selected duplexes and numbering of nucleobases are indicated in Figure 5. The double-headed nucleotides are located at base numbers 7 (U_7) and 8 (**X**), and are situated across A_{22} and Y_{21}, respectively. To compare the modified duplexes with idealized B-DNA helix parameters, the corresponding unmodified controls were also subjected to MD simulations. The duplexes were first constructed in canonical B-DNA conformations with explicit representation of water and ions. The MD simulations were performed using the AMBER14 package.[49] After full equilibration of the systems, the duplexes were subjected to 500 ns of Langevin dynamics. To validate the results, two parallel trajectories with different initial velocities were performed. Base pair, base step and backbone conformational parameters (as defined by the 3DNA software package;[50] see Supplementary Figure S33) were monitored during the trajectories to quantitatively probe the conformational behavior of duplexes. All simulated structures remained in canonical B-DNA form over 500 ns. Incorporation of the double-headed nucleotides led to no visual perturbation of the helixes outside of the three central base pairs U_7·A_{22}, **X**·**Y** and C_9·G_{20}. The distributions of helix parameters, sugar puckering and pseudorotational angles for the six central base pairs during the MD simulations are shown in the probability mass functions of Supplementary Figures S34–S51. Supplementary Tables S2–S19 summarize the corresponding translational and rotational mode values for an easier overview.

First, the MD simulations of duplexes 1–5, where the C2′-linked nucleobases are situated across matching bases, were analyzed with a primary focus on establishing the binding geometry and kinematics. The average geometry of the three central base pair residues are shown in Figure 6 (duplexes 1–4) and Supplementary Figure S28 (duplex 5). In all cases, standard base pairing schemes were maintained throughout the simulations, i.e. standard Watson–Crick contacts in the case of **U**$_A$, **U**$_T$ and **U**$_D$ (duplexes 1–3), and standard H·A and H·C wobble contacts in the case of **U**$_H$ (duplexes 4 and 5). In fact, the calculated hydrogen bond occupancies are indifferentiable from the native duplexes (see Supplementary Figure S26). Interestingly, the modified and native base pairs display geometric similarity (isostericity), meaning that the base positions and the interstrand distances are very similar. In all cases, the **X**$_8$·**Y**$_{21}$ base pair of the modified duplexes is neatly

10.1002/chem.201901077

Duplex №	X	Y
1	A	T
2	T	A
3	D	T
4	H	A
5	H	C
6	A	G
7	T	G
8	D	G
9	H	G

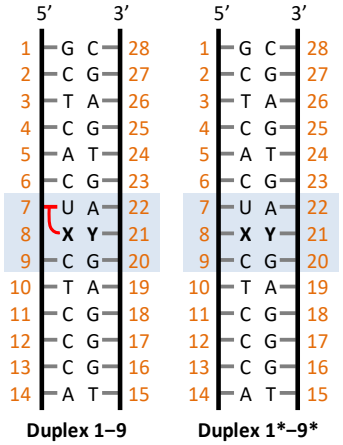


Figure 5. Sequences of the nine DNA duplexes subjected to molecular dynamics studies, and the nucleobase numbering scheme used in analysis.

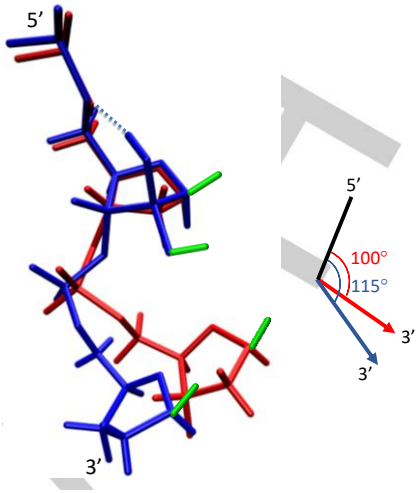


Figure 7. Superposition of the backbone of the 5′-**U**$_T$C (blue) and 5′-UT (red) dinucleotide step from the MD simulations of duplex 2 and 2*, respectively. Bases are omitted for clarity. The internal hydrogen bond interaction of the arabinose sugar is shown as a dashed line. The illustration indicates the projected angle between the two successive C3′-P and P-C4′ vectors. The structure is representative for all simulated double-headed nucleotides.

placed in the stack of nucleobases with ordinary B-DNA rise values of 3.0–3.5 Å. It follows that incorporation of the compressed dinucleotide mimic does not reduce the helical pitch; rather the geometry change is compensated by certain adjustments of the backbone torsion angles in the modified strand (Figure 7).

Specifically, the segment of the 3′-side of the **U**$_X$ nucleotides is elongated to accommodate the extra nucleobase. This happens by changes to the ε, ζ and β+1 torsion angles. The double-headed nucleotides adopt very fixed geometries throughout the simulations. Notably, they feature a robust hydrogen bond from the C2′-hydroxyl group to the intra-residual O5′ atom (indicated with a dashed line in Figure 7). This contact may help to stabilize the sugar conformation in a rigid C2′-*endo* puckering mode (*P* ~ 150°), causing the C2′-linked base to be laterally oriented into the base stack. While this hydrogen bond is not seen in stand-ard arabino nucleic acids,[51] it has been reported in force field simulations of other C2′-substituted arabino nucleic acids.[52] Also common to all the double-headed nucleotides is a relatively high *anti* orientation (χ ~ –105°) of the glycosidic bond compared to the 2′-deoxyuridine residue in the native setting (χ ~ –135°). To achieve the optimal geometry for base pairing, the nucleotide that is presented to the C2′-linked base (i.e. **Y**$_{21}$) also adopts glycosidic angles of –75 to –105° (versus –110 to –135° in native settings).

While the helix width is unaffected by the presence of the double-headed nucleotides, they do have a few substantial effects on the stacking geometry of the central **X**$_8$·**Y**$_{21}$ pair and the immediately adjacent base steps. A striking feature is the reduced *twist* of the **U**$_X$ step (base step 7-8), which with an average value of +10° is invariably lower than the standard B-DNA twist of approx +34° that is observed in the native duplexes (Figure 8A). In addition to the reduced helical twist, specific changes in the buckle (Figure 8B) and propeller (Figure 8C) parameters of the **U**$_X$:3′-A**Y** segment were observed. While the native U$_7$·A$_{22}$ and

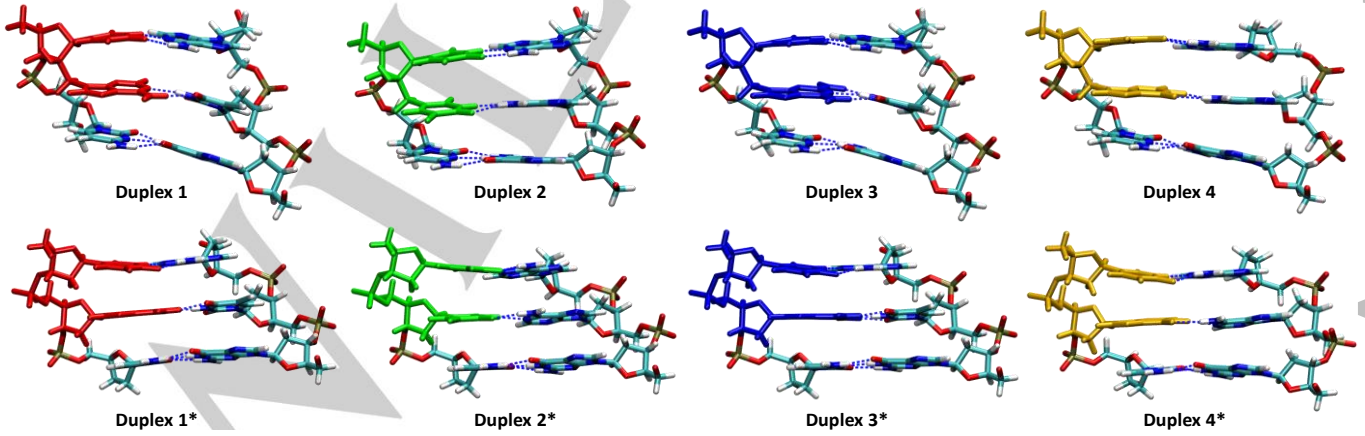


Figure 6. Perpendicular-to-the-helical view of the central 5′-d(U$_7$**X**$_8$C$_9$)/(A$_{22}$**Y**$_{21}$G$_{20}$)-5′ motifs of duplexes 1–4 obtained from the MD simulations (top row). The corresponding native duplexes are marked with an asterisk (*). The models are generated by averaging MD frames over the course of the full trajectory (500 ns or 250k frames) in which all nucleobases are robustly positioned in the helical stack. Potential H-bonds are shown as dotted blue lines. The double-headed nucleotides are colored—**U**$_A$ (red), **U**$_T$ (green), **U**$_D$ (blue), and **U**$_H$ (yellow)—and the corresponding native dinucleotides are colored accordingly.

10.1002/chem.201901077

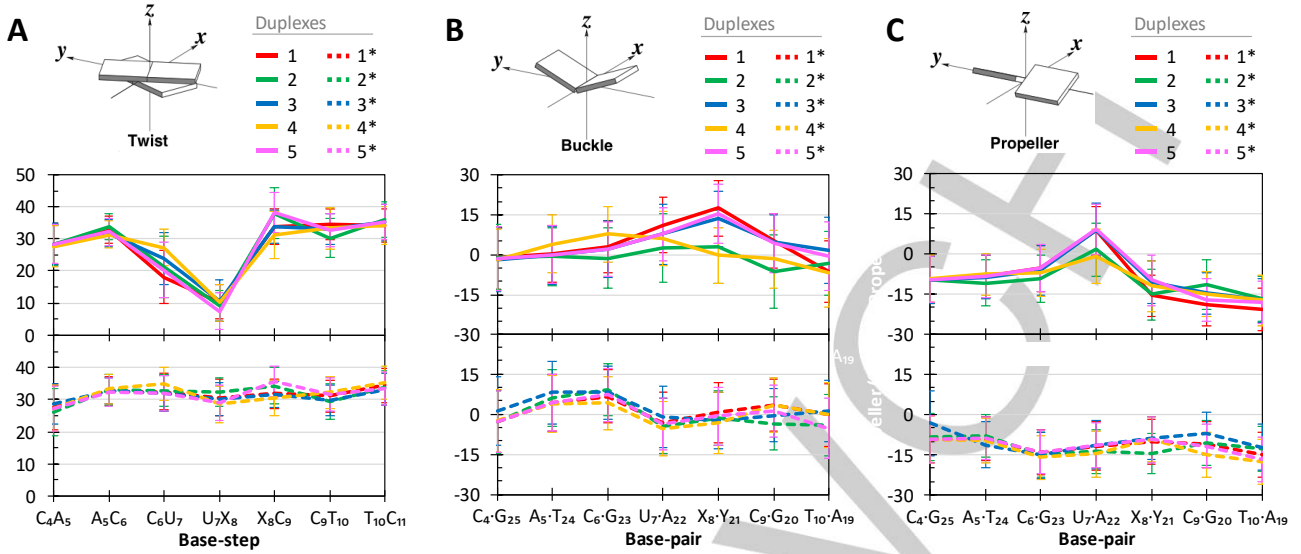


Figure 8. Twist, buckle and propeller angles (°) at central base steps/pairs of the 14-mer duplexes 1–5 (top charts) and 1*–5* (bottom charts). Duplex 1 ($\mathbf{U_A}$:3′-AT), duplex 2 ($\mathbf{U_T}$:3′-AA), duplex 3 ($\mathbf{U_A}$:3′-AT), duplex 4 ($\mathbf{U_H}$:3′-AA) and duplex 5 ($\mathbf{U_H}$:3′-AC). The values are computationally determined mean values based on every fifth frame within the 500 ns simulations. Error bars indicate standard deviations (±σ). The pictorial diagrams are reproduced from the 3DNA protocol.[53] The major groove is at positive *x*.

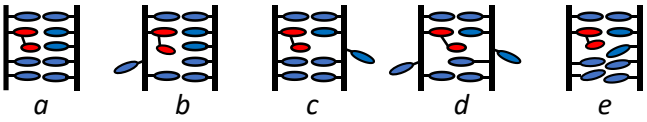
$\mathbf{X_8}$·$\mathbf{Y_{21}}$ base pairs are co-planar along the helix axis (i.e. has a buckle of 0°) and have an average propeller of approx –10°, the double-headed nucleotides cause buckling of the $\mathbf{X_8}$·$\mathbf{Y_{21}}$ pair and change the propeller of the U_7·A_{22} pair. These structural perturbations are larger in the cases where $\mathbf{Y_{21}}$ is a pyrimidine (duplexes 1, 3 and 5) than when it is a purine (duplex 2 and 4). In the former, $\mathbf{Y_{21}}$ adopts pseudorotational angles of 45–75° (C4′-*exo*) instead of the native C2′-*endo* (135–150°) pucker (see the Supporting Information), and the $\mathbf{X_8}$·$\mathbf{Y_{21}}$ pair buckles by approx +15° (Figure 8B). In addition, the U_7·A_{22} pair takes positive propeller values (+10°, Figure 8C). In contrast, when $\mathbf{Y_{21}}$ is a purine, it adopts a Southern pseudorotational angle of ~180° ($^2_3$T pucker), and buckling and propeller-twisting of the base pairs are negligible. For the other base pair and base-step parameters (opening, roll, slide, shift, stagger, etc.) there is largely consensus between the unmodified and modified duplexes, and any disparities are only visible in fine details.

Structural studies of double-headed nucleotides' mispairing to G

Having established the double-headed nucleotides' behavior as functional dinucleotides across matching bases, we turned our attention toward aberrant base pairing with G. As demonstrated by T_m readings, the C2′-linked A and T bases of $\mathbf{U_A}$ and $\mathbf{U_T}$ distinguish a mismatched G nucleotide somewhat poorly. Enhanced fidelity against G is attained by using $\mathbf{U_D}$ and $\mathbf{U_H}$, respectively, as proxies. Several factors could contribute to the higher stability of the $\mathbf{U_A}$:3′-AG and $\mathbf{U_T}$:3′-AG structures, including reduced frequency of base-flipping events as A·G and T·G mispairs are better accommodated by the double-headed design, or the C2′-linked A and T bases engage in favorable interactions with G that are conformationally improbable for classical base pairs. To assess possible explanations, we closely inspected the 500 ns MD simulations of duplexes 6–9 from Figure 5.

In the simulations, structural perturbations were confined to the region of the $\mathbf{U_X}$:3′-AG mismatch leaving the global helix structure unaffected. In contrast to the matched duplexes 1–5, the mismatched duplexes 6–9 displayed more conformational variability. Specifically, the mismatched duplexes adopted five distinct conformations *a*–*e* that are schematized in Table 1 along with the percentage occupancy of each conformation. The edge-to-edge

Table 1. Different conformations *a*–*e* of the mismatched duplexes.



Duplex	Run	Occupancy of conformations				
		a	*b*	*c*	*d*	*e*
6	i	56%	-	31%	13%	-
	ii	84%	-	16%	-	-
7	i	100%†	-	-	-	-
	ii	51%	49%	-	-	-
8	i	95%	-	3%	2%	-
	ii	71%	-	15%	14%	-
9	i	63%	-	-	-	37%
	ii	67%‡	-	-	-	33%
6*–9*	i/ii	100%	-	-	-	-

Schematic representations of the different geometries *a*–*e* of the central five base pairs during the two MD simulations of the modified duplexes 6–9. The table shows the occupancy of conformations *a*–*e* during each of the two trajectories (runs): (*a*) all-in-stack conformation; (*b*) flipping of C_9; (*c*) flipping of G_{21}; (*d*) flipping of both C_9 & G_{21} (cross-strand stacking); (*e*) interstrand H-bonding between neighboring base planes. †Wobble T·G contributes 75% to the conformation, the remaining 25% being T·G pairing where T_8(O4) connects to both G_{21}(N1-H) and G_{21}(C2-NH_2). ‡Hoogsteen configuration makes up 15% of conformation *a*.

conformation (*a*) with all bases in stack dominated the trajectories (i.e. occupancies of 51–100%). Base-flipping of C_9 and/or G_{21} (*b*–*d*), or "twisted" interstrand neighboring base pairing (*e*) were observed to varying degrees (2–49%). These alternative conformations were not observed for the unmodified duplexes 6*–9*. Representative snapshots of duplexes 6–9 in the different conformations (*a*–*e*) are given in Supplementary Figures S29–S32. The MD simulations show that the C2′-linked A, T, D and H bases pair up with G similarly to the native nucleotides, however, the conformational rigidity of the double-headed nucleotides causes alternative base dynamics (see the Supporting Information p. 49 for discussion). That is, the double-headed nucleotides cannot easily absorb the structural stress of the mismatch, which is propagated into neighboring residues instead. It follows that G-mismatches are less likely to retain canonical Watson–Crick structure when double-headed nucleotides are involved. However, our observations do not explicitly explain why this behavior translates into a roughly +1 °C gain in the T_m for $\mathbf{U_A}$:3′-AG, $\mathbf{U_T}$:3′-AG and $\mathbf{U_D}$:3′-AG relative to the natural dinucleotides; no unusual geometry seems to account for this slight increase.

Discussion

The results presented herein now allow for a generalized description of the association properties of double-headed nucleotides. For the first time, we provide exclusive evidence (experimental and computational) to support our hypothesis that both bases of the double-headed nucleotides in fact communicate efficiently with a target strand. Indeed, the *arabino* design appears seamless for this purpose. Our results also support the view that double-headed nucleotides, like natural nucleotides, preferentially accept complementary bases at the Watson–Crick face of the nucleobases and not at the Hoogsteen face. As demonstrated, double-headed nucleotides can associate with either natural dinucleotides or other double-headed nucleotides in the opposite strand. In the latter, the backbone easily adapts to the situation where two modified 11-mer oligonucleotides hybridize to contain 12 base pairs. Hereby, the duplex carries a larger number of Watson–Crick base pairs per phosphate unit, and information can therefore be delivered using a shorter sequence. Notably, such duplexes have higher thermodynamic stability (+5.5 to +9.0 °C) relative to the natural dodecamer. We attribute this stabilization to reduced repulsive Coulombic interactions between the phosphates, and an entropic gain due to the reduced flexibility of the sugar–phosphate backbone. To what extent differential solvation of the backbone affects the entropy term is unknown so far. When double-headed nucleotides are used to replace natural dinucleotides, more energetically neutral effects (–2.5 to +2.0 °C) are observed within the Watson–Crick pairing regime. Nevertheless, such duplexes contain fewer phosphates without perturbing the overall duplex geometry. This feature may be advantageous for e.g. improving membrane permeability whilst retaining biological function. In this regard, we have recently shown that two or three double-headed nucleotides can be incorporated in the same DNA strand without compromising target binding.[29]

With its favorable DNA binding properties, double-headed nucleotides represent an interesting candidate worthy of further exploration. The molecular modeling offers insight into underlying driving forces of base pairing in atomic detail, and into how double-headed nucleotides are housed in DNA duplexes. Knowing such information about new DNA architectures is useful for assessing enzyme compatibility. In this regard, it is of note that double-headed nucleotides do not visibly affect the helix diameter, groove widths, or minor groove faces, which are well-known elements of polymerase activity.[54] Whilst we have established the eligibility and feasibility of double-headed nucleotides as building blocks in duplexes, attention to their functional roles and biological interactions is needed. On a long term, we hope that double-headed nucleotides can unlock an untapped potential in nucleic acid-based drugs or DNA nanotechnology, or lead to a new paradigm in nucleic acid communication. The investigation of double-headed nucleotides' functional properties will be the subject of a following communication.

Conclusions

In this study, we have developed a full set of double-headed nucleotides containing six C2′-linked nucleobases (A, T, C, G, D and H). With the aim to shed light on the base-pairing properties of this interesting new class of xenobiotic nucleotides, we investigated their fidelity of binding in several motifs, and we utilized molecular modeling to obtain structural insight into their base-pairing geometry. We found that the double-headed nucleotides adopt C2′-*endo* sugar puckers and fit very well into the geometry of the B-form duplex. The two bases of the double-headed nucleotides were found to stack very rigidly with a smaller twist than the corresponding native base steps. Double-headed nucleotides hybridize to complementary targets neatly with their Watson–Crick faces using a similar profile as natural DNA. And like normal bases, the C2′-linked bases associate preferentially with G bases, giving rise to the strongest Watson–Crick base pair and the most stable mismatches. C is the most discriminating base in double-headed nucleotides forming the strongest Watson–Crick base pair and the weakest mismatches. Double-headed nucleotides containing D and H were found to strongly bind to T and A, respectively, and neatly discriminate against G; they could be a useful tool in designs where fidelity is important. All in all, the results reported herein provide new insights into the potential of inserting double-headed nucleotides in DNA, and provide valuable information for upcoming designs of new double-headed nucleotides.

Acknowledgements

The authors gratefully acknowledge support for this work by the Danish National Research Foundation, the Danish Council for Independent Research | Natural Sciences (FNU). Computation/simulation for the work described in this paper was supported by the Danish e-Infrastructure Cooperation (DeIC) National HPC Centre, SDU.

Keywords: oligonucleotides • double-headed nucleotides • 2,6-diaminopurine • hypoxanthine

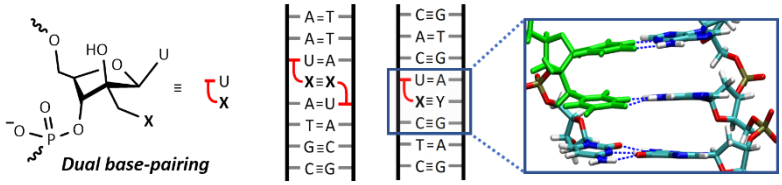
[1] C. Chen, Z. Yang, X. Tang, *Med. Res. Rev.* **2018**, *38*, 829–869.
[2] W. B. Wan, P. P. Seth, *J. Med. Chem.* **2016**, *59*, 9645–9667.
[3] G. F. Deleavey, M. J. Damha, *Chem. Biol.* **2012**, *19*, 937–954.
[4] T. R. Gingeras, R. Higuchi, L. J. Kricka, Y. M. D. Lo, C. T. Wittwer, *Clin. Chem.* **2005**, *51*, 661–671.
[5] J. D. Watson, A. A. Caudy, R. M. Myers, J. A. Witkowski, *Recombinant DNA: Genes and Genomes—A Short Course*, W. H. Freeman & Co., New York, **2007**.
[6] J. M. Butler, *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers*, Elsevier Academic Press, New York, **2005**.
[7] E. O. Wiley, B. S. Lieberman, *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, John Wiley & Sons, Hoboken, **2011**.
[8] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E. M. LeProust, B. Sipos, E. Birney, *Nature* **2013**, *494*, 77–80.
[9] G. M. Church, Y. Gao, S. Kosuri, *Science* **2012**, *337*, 1628–1628.
[10] N. C. Seeman, *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
[11] O. I. Wilner, I. Willner, *Chem. Rev.* **2012**, *112*, 2528–2556.
[12] N. C. Seeman, *Trends Biochem. Sci.* **2005**, *30*, 119–125.
[13] A. Kuzuya, Y. Ohya, *Acc. Chem. Res.* **2014**, *47*, 1742–1749.
[14] W. Xu, K. M. Chan, E. T. Kool, *Nat. Chem.* **2017**, *9*, 1043–1055.
[15] R. S. Geary, D. Norris, R. Yu, C. F. Bennett, *Adv. Drug Deliv. Rev.* **2015**, *87*, 46–51.
[16] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, P. Holliger, *Science* **2012**, *336*, 341–344.
[17] A. W. Feldman, F. E. Romesberg, *Acc. Chem. Res.* **2018**, *51*, 394–403.
[18] P. K. Sharma, P. Kumar, P. Nielsen, *Aust. J. Chem.* **2016**, *69*, 1094–1101.
[19] P. Kielkowski, H. Cahová, R. Pohl, M. Hocek, *Bioorg. Med. Chem.* **2016**, *24*, 1268–1276.
[20] D. S. Chatzileontiadou, V. Parmenopoulou, S. Manta, A. L. Kantsadi, P. Kylindri, M. Griniezaki, F. Kontopoulou, A. Telopoulou, H. Prokova, D. Panagopoulos, E. Boix, N. A. Balatsos, D. Komiotis, D. D. Leonidas, *Bioorg. Chem.* **2015**, *63*, 152–165.
[21] S. L. Pedersen, P. Nielsen, *Org. Biomol. Chem*. **2005**, *3*, 3570-3575.
[22] T. Umemoto, J. Wengel, A. S. Madsen, *Org. Biomol. Chem.* **2009**, *7*, 1793–1797.
[23] T. Wu, K. Nauwelaerts, A. van Aerschot, M. Froeyen, E. Lescrinier, P. Herdewijn, *J. Org. Chem.* **2006**, *71*, 5423–5431.
[24] M. Hornum, P. K. Sharma, C. Reslow-Jacobsen, P. Kumar, M. Petersen, P. Nielsen, *Chem. Commun.* **2017**, *53*, 9717–9720.
[25] P. Kumar, C. S. Madsen, P. Nielsen, *Bioorg. Med. Chem. Lett.* **2013**, *23*, 6847–6850.
[26] C. S. Madsen, S. Witzke, P. Kumar, K. Negi, P. K. Sharma, M. Petersen, P. Nielsen, *Chem. Eur. J.* **2012**, *18*, 7434–7442.
[27] P. Kumar, P. K. Sharma, C. S. Madsen, M. Petersen, P. Nielsen, *ChemBioChem* **2013**, *14*, 1072–1074.
[28] P. Kumar, P. K. Sharma, P. Nielsen, *J. Org. Chem.* **2014**, *79*, 11534–11540.
[29] K. Beck, C. Reslow-Jacobsen, M. Hornum, C. Henriksen, P. Nielsen, *Bioorg. Med. Chem. Lett.* **2019**, *29*, 740–743.
[30] F. H. Martin, M. M. Castro, F. Aboul-ela, I. Tinoco, *Nucleic Acids Res.* **1985**, *13*, 8927–8938.
[31] A. V. Aerschot, B. Peeters, H. Vanderhaeghe, *Nucleosides Nucleotides* **1987**, *6*, 437–439.
[32] F. Seela, Y. Chen, *Nucleic Acids Res.* **1995**, *23*, 2499–2505.
[33] H. Liu, R. Nichols, *BioTechniques* **1994**, *16*, 24–26.
[34] F. B. Howard, H. T. Miles, *Biochemistry* **1984**, *23*, 6723–6732.
[35] S. Gryaznov, R. G. Schultz, *Tetrahedron Lett.* **1994**, *35*, 2489–2492.
[36] M. J. Cairns, T. Thomas, C. E. Beltran, D. Tillett, *BMC Genomics* **2009**, *10*, 344.
[37] U. B. Voss, A. Chollet, A. D. B. Malcolm, *Biochem. Soc. Trans.* **1989**, *17*, 913–913.
[38] F. B. Howard, C. Chen, J. S. Cohen, H. T. Miles, *Biochem. Biophys. Res. Commun.* **1984**, *118*, 848–853.
[39] T. Brown, A. G. Craig, *Nucleosides Nucleotides* **1989**, *8*, 1051–1051.
[40] A. De Mico, R. Margarita, L. Parlanti, A. Vescovi, G. Piancatelli, *J. Org. Chem.* **1997**, *62*, 6974–6977.
[41] S. Lemaire, I. Houpis, R. Wechselberger, J. Langens, W. A. A. Vermeulen, N. Smets, U. Nettekoven, Y. Wang, T. Xiao, H. Qu, *J. Org. Chem.* **2011**, *76*, 297–300.
[42] E. J. Corey, M. Chaykovsky, *J. Am. Chem. Soc.* **1965**, *87*, 1353–1364.
[43] Y. Hayakawa, M. Hirose, R. Noyori, *J. Org. Chem.* **1993**, *58*, 5551–5555.
[44] S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculca, N. Windhab, R. Micura, M. Stanek, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **2003**, *86*, 4270–4363.
[45] P. Børsting, K. E. Nielsen, P. Nielsen, *Org. Biomol. Chem.* **2005**, *3*, 2183–2190.
[46] A. Canol, M. F. Goodman, R. Eritja, *Nucleosides Nucleotides* **1994**, *13*, 501–509.
[47] C. Rosenbohm, D. S. Pedersen, M. Frieden, F. R. Jensen, S. Arent, S. Larsen, T. Koch, *Bioorg. Med. Chem.* **2004**, *12*, 2385–2396.
[48] V. Boudou, K. Rothenbacher, A. V. Aerschot, P. Herdewijn, *Nucleosides Nucleotides* **1999**, *18*, 1429–1431.
[49] D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, P. A. Kollman. *Amber14*; University of California, San Francisco, **2014**.
[50] X.-J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
[51] N. Martín-Pintado, M. Yahyaee-Anzahaee, R. Campos-Olivas, A. M. Noronha, C. J. Wilds, M. J. Damha, C. González, *Nucleic Acids Res.*, **2012**, *40*, 9329–9339.
[52] D. Venkateswarlu, D. M. Ferguson, *J. Am. Chem. Soc.* **1999**, *121*, 5609–5610.
[53] X.-J. Lu, W. K. Olson, *Nat. Protoc.* **2008**, *3*, 1213–1227.
[54] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.

10.1002/chem.201901077

WILEY-VCH

Entry for the Table of Contents

FULL PAPER



Doubling the information of a nucleotide: Double-headed nucleotides constructed on arabinose sugars exhibit excellent dinucleotide nature for all combinations of nucleobases; both with respect to maintaining helix geometry and base pairing mode. Accordingly, this study lays the groundwork for condensing genetic information in DNA onto shorter strands.

*Mick Hornum, Julie Stendevad, Pawan K. Sharma, Pawan Kumar, Rasmus B. Nielsen, Michael Petersen, and Poul Nielsen**

Page No. – Page No.

Base-pairing properties of double-headed nucleotides