Double-headed nucleic acids condense the molecular information of DNA to half the number of nucleotides

Beck, Kasper Munch; Sharma, Pawan K.; Hornum, Mick; Risgaard, Nikolaj; Nielsen, Poul

Published in:
Chemical Communications

DOI:
10.1039/D1CC03539H

Publication date:
2021

Document version:
Accepted manuscript

Citation for pulished version (APA):
Beck, K. M., Sharma, P. K., Hornum, M., Risgaard, N., & Nielsen, P. (2021). Double-headed nucleic acids condense the molecular information of DNA to half the number of nucleotides. *Chemical Communications*, *57*(72), 9128-9131. https://doi.org/10.1039/D1CC03539H

COMMUNICATION

Double-headed nucleic acids condense the molecular information of DNA to half the number of nucleotides

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Kasper M. Beck,[a] Pawan K. Sharma,[b] Mick Hornum,[a] Nikolaj A. Risgaard[a] and Poul Nielsen[a]*

Nucleotide monomers that hold two nucleobases each, i.e. double-headed nucleotides, have been shown to form two sets of functional Watson–Crick base pairs when incorporated into dsDNA, and they hereby behave as dinucleotides. To form the basis for fully modified double-headed nucleic acids (DhNA), we have prepared three new DhNA monomers and can now demonstrate that the molecular information of 10 Watson–Crick base pairs can be condensed to highly stable 5-mer DhNA duplexes.

The unique predictability and simplicity of the native nucleic acid structure have inspired the development of numerous synthetic polymers for a wide range of functions within medicinal chemistry,[1,2] bio- and nanotechnology.[3,4] A class of designs with two nucleobases per nucleotide has emerged,[5] including double-headed cytidine monomers that mimic flipped-out cytidines while still participating in Watson–Crick base-pairing,[6] bimodal PNA capable of binding two different complementary DNA strands simultaneously[7] and self-assembling double-headed homoazanucleosides.[8] We have recently focused on double-headed nucleic acid (DhNA) monomers, where the additional nucleobase is attached to the 2′-position of arabino-nucleotides through a methylene linker (Fig. 1). This configuration directs the additional nucleobase towards the duplex core, where it participates in Watson–Crick base-pairing with a nucleobase from the complementary strand. In oligonucleotides (ON's), one DhNA monomer can therefore replace a natural dinucleotide (Fig. 1B). This effectively condenses the molecular information to half the number of nucleotides and might improve key features of future nucleic acid-based therapeutics such as pharmacokinetics and target affinity.

Our previous studies have shown that single incorporations of DhNA monomers (**U_T**, **U_C**, **U_A**, **U_G**, **C_C**, **A_T**) into dsDNA are well-tolerated with only minor effects on the melting temperature (T_m) of the duplex (–3.5 °C to +4.0 °C).[9-12] Intensive molecular dynamics studies have shown that the monomers adopt C2′-*endo* conformations and fit very well into the B-form duplex with the two bases stacking rigidly and with a smaller twist than

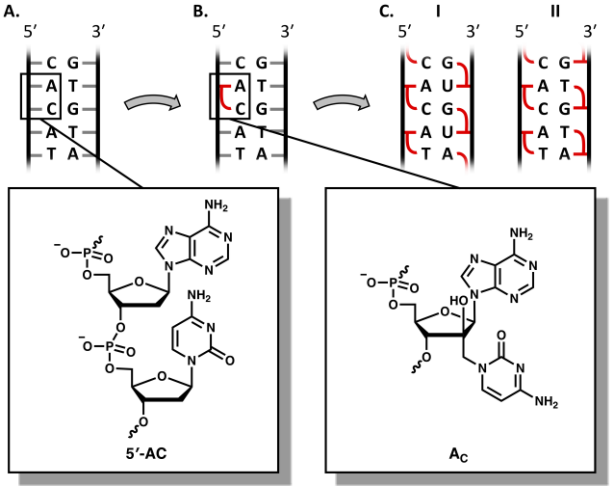


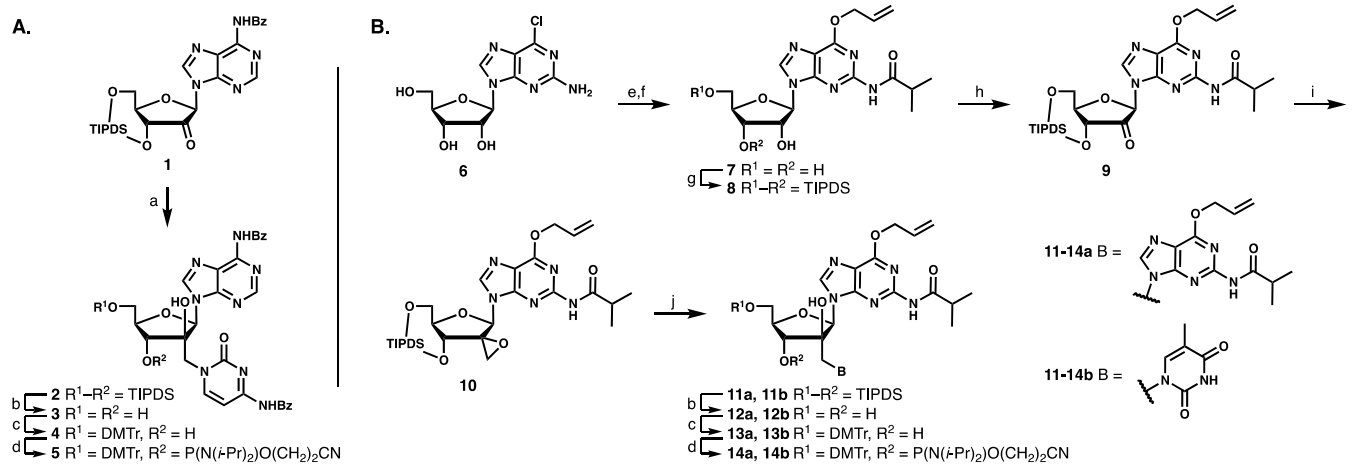
Fig. 1 Excerpt of (A) natural dsDNA, (B) dsDNA with a single DhNA monomer and (C) dsDhNA (two types). Highlighted as examples are the structures of a 5′-AC dinucleotide and the corresponding DhNA monomer **A_C**, respectively.

for a native base step.[13] Introduction of 2–4 DhNA monomers in opposite strands increased duplex stability with up to 14.0 °C by forming either 2′·2′ base pairs or alternating 1′·2′ base pairs (Fig. 1C).[12,13] Finally, we have studied the recognition of DhNAs by DNA polymerases, and found that the information of a template-embedded **U_A** monomer could be transferred to natural DNA by Therminator DNA polymerase under controlled conditions.[12] With its distinct structure and stable self-pairing motifs, the DhNA has the potential to function as an artificial genetic system, i.e. as a xeno nucleic acid (XNA).[2,14]

Herein, we present the synthesis and evaluation of the first DhNA duplexes consisting entirely of DhNA monomers (Fig. 1C). To prepare diverse, non-repeating sequences, we also introduce three new DhNA monomers in the form of **A_C**, **G_G** and **G_T**. The addition of guanosine-based monomers finalizes the preliminary series of DhNA monomers to now include examples of all four nucleobases in both the 1′ and the 2′-position as 9 of the 16 theoretical nucleobase combinations have been realized (considering thymine and uracil to be equal).

The phosphoramidite method was used to prepare ON's containing DhNA monomers. The phosphoramidite equivalent of **A_C** was prepared analogously to that of **A_T** (Scheme 1A). From

[a] *Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230 Odense, Denmark. E-mail: pouln@sdu.dk.*
[b] *Department of Chemistry, Kurukshetra University, Kurukshetra, 136119, India.*
Electronic Supplementary Information (ESI) available: Experimental procedures, MALDI-TOF spectra, HPLC-profiles, CD spectra, UV melting curves, NMR spectra. See DOI: 10.1039/x0xx00000x



Scheme 1 Reagents and conditions: (a) i. NaH, trimethylsulfoxonium iodide, DMSO, THF, 0 °C, 40 min. ii. NaHMDS, 4-*N*-benzoylcytosine, THF, DMF, 65 °C, 18 h, 46% (two steps); (b) $Et_3N{\cdot}3HF$, THF, rt, 3–20 h, 85% **3**, 93% **12a**, 89% **12b**; (c) DMTrCl, DMAP, pyridine, rt, 18–48 h, 50% **4**, 79% **13a**, 74% **13b**; (d) 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite, DIPEA, CH_2Cl_2, rt, 18 h, 79% **5**, 60% **14a**, 87% **14b**; (e) allyl alcohol, NaH, reflux, 60 min; (f) i. TMSCl, pyridine, rt, 60 min. ii. isobutyryl chloride, pyridine, rt, 18 h. iii. NH_3 (aq), pyridine, rt, 10 min, 61% (two steps); (g) $TIPDSCl_2$, pyridine, 0 °C to rt, 18 h, 83%; (h) Dess–Martin periodinane, CH_2Cl_2, 0 °C to rt, 18 h, quant.; (i) NaH, trimethylsulfoxonium iodide, DMSO, THF, 0 °C, 30 min, 60%; (j) **11a**: NaHMDS, 6-*O*-allyl-2-*N*-isobutyrylguanine,[10] THF, DMF, rt, 3 d, 46%; **11b**: K_2CO_3, thymine, DMF, rt, 18 h, 73%. DMTr = 4,4′-dimethoxytrityl, TIPDS = tetraisopropyldisiloxan-1,3-diyl.

the protected ketone **1**,[12] the crude intermediate 2′-spiroepoxide was prepared and directly reacted with deprotonated 4-*N*-benzoylcytosine to give the double-headed nucleoside **2** in 46% overall yield. The silyl protecting group was then removed to give **3**. Subsequent tritylation to form **4** followed by phosphitylation using standard protocols provided the protected phosphoramidite **5**. The 2′-OH remained unprotected as sterical hindrance renders it practically inert.[9,10]

For the design of guanosine-based analogues, we have previously demonstrated *N*9-selective alkylation of 6-*O*-allyl-2-*N*-isobutyrylguanine when reacted with the protected 2′-spiroepoxide of uridine in the synthesis of $\mathbf{U_G}$, as well as convenient deprotection of the guanine after oligonucleotide synthesis.[10] Thus, matching protecting groups on the 1′-guanine are desired (Scheme 1B). Starting from 6-chloroguanosine (**6**), the 6-*O*-allyl group was installed by treatment with deprotonated allyl alcohol. Selective 2-*N*-isobutyryl protection was achieved via transient hydroxyl protection to give nucleoside **7** in 61% overall yield. From **7**, the same methodology as for the uridine[9,10] and adenosine-based[12] DhNA monomers was employed, i.e. 3′,5′-*O*-TIPDS protection to give **8** followed by oxidation using Dess–Martin periodinane to give **9** and then Corey–Chaykovsky epoxidation to give epoxide **10**. Ring opening was readily achieved by treatment with either 6-*O*-allyl-2-*N*-isobutyrylguanine[10] and catalytic amounts of NaHMDS or thymine and excess K_2CO_3 to furnish the double-headed nucleosides **11a** and **11b** in yields of 46% and 73%, respectively. Also the *N*7-isomer of **11a** (Scheme S2, ESI†) was formed in this step in a yield of 27%. The regiochemistry of **11a** was confirmed using ^1H,^{13}C-HMBC spectroscopy. A NOESY spectrum confirmed the expected stereochemistry at C2′ by the presence of dipolar couplings between H1′ and H6′ and between H8 and the 2′-OH. Desilylation of **11a**,**b** afforded **12a**,**b**, which were tritylated to give **13a,b** and phosphitylated to give phosphoramidites **14a**,**b**. Initial strategies to synthesize guanosine-based monomers included the use of monoprotected (2-*N*-amidine or 2-*N*-isobutyryl) guanosines, and the key nucleoside **7** has also been prepared using Vorbrüggen conditions and from 2-*N*-isobutyryl-guanosine using Mitsunobu conditions (Scheme S2, ESI†).

Together with previous monomers ($\mathbf{U_A}$, $\mathbf{U_G}$, $\mathbf{C_C}$, $\mathbf{A_T}$)[9-12], the three new phosphoramidites were used to prepare 11-mer DNA sequences containing one DhNA monomer (ON1–4, Table 1) as well as two 5-mer sequences consisting entirely of DhNA monomers (ON5 and ON6) and finally a 6-mer consisting of four DhNA monomers flanked by DNA nucleotides (ON7). These sequences were all readily prepared by standard solid-phase DNA synthesis and purified under standard conditions. Full deprotection and cleavage from the solid support was achieved by treatment with aq. ammonia at room temperature for all ON's. Deallylation of ON's containing 6-*O*-allyl protected guanines was obtained with a mixture of $Pd(PPh_3)_4$, PPh_3 and $Et_2NH_2{\cdot}HCO_3$ in DCE.[15] As evident from Table 1, MALDI MS confirmed the expected masses of ON1–7. The MW of fully modified 5-mer DhNA sequences are substantially reduced as compared to their 10-mer DNA counterparts with the same nucleobase sequence showing 32.8% and 34.9% higher MW.

Table 1 MALDI-TOF *m/z* values for synthesized oligonucleotides.

		Modified ON		Unmodified ON[a]	
ON	Sequence	MW_{calcd}	MW_{found}	MW_{calcd}	$\%_{increase}$[b]
1	5′-d(CGCT $\mathbf{A_C}$CTACGC)	3416.3	3416.5	3566.4	+4.4%
2	5′-d(GCGTA $\mathbf{G_G}$TAGCG)	3576.4	3576.1	3726.5	+4.2%
3	5′-d(GCGTAG $\mathbf{G_T}$AGCG)	3576.4	3577.0	3726.5	+4.2%
4	5′-d(CGCTA $\mathbf{C_C}$TACGC)	3416.3	3414.8	3566.4	+4.4%
5	5′-d($\mathbf{U_AC_CA_CA_TA_C}$)	2191.6	2190.2	2956.0[c]	+34.9%
6	5′-d($\mathbf{G_TA_TG_TG_GU_A}$)	2333.7	2331.8	3098.1[d]	+32.8%
7	5′-d(G $\mathbf{U_AU_GU_GG_T}$ A)	2455.7	2453.8	3098.1[d]	+26.2%

[a]Corresponding nucleobase sequence using only natural DNA nucleotides; [b]The percentage increase in MW for the unmodified DNA references compared to the respective modified sequences; [c]ON8; [d]ON9.

To verify that the dinucleotide behavior found for the previous DhNA monomers also extends to the new monomers, the three 11-mer sequences (ON1–3, Table 1) containing each of the new DhNA monomers were hybridized to their 12-mer DNA complements, and to sequences with a mismatch opposite to the 1′ or the 2′-base (Fig. 2A). The thermal stability was then determined as the melting temperature (T_m) obtained from the UV_{260} melting curve. The unmodified DNA duplex displayed a T_m of 51.5 °C, and when the $\mathbf{A_C}$ monomer was introduced instead of the central 5′-AC dinucleotide, the T_m value was increased with 2.0 °C. This agrees with our previous studies on the 2′-connected cytosine of $\mathbf{U_C}$ and $\mathbf{C_C}$ that displayed increases in the T_m of 1.0 °C and 4.0 °C, respectively (Fig. 2B).[10,11] Excellent mismatch discrimination was observed for both the 1′ and the 2′-base of $\mathbf{A_C}$ comparable to the observations for the native duplex (Fig. 2A). Unusually low thermal stabilities were found for both the native and the modified C·C mispaired duplex.

Incorporation of $\mathbf{G_G}$ led to a reduction in duplex stability of 4.0 °C, which is somewhat surprising as compared to the only formerly studied 2′-G analogue $\mathbf{U_G}$, which provides neutral effect on stability.[10,11,13] Nevertheless, $\mathbf{G_G}$ generally displayed great mismatch discrimination comparable to what was observed for the native duplex. Finally, incorporation of the $\mathbf{G_T}$ monomer also reduced duplex stability by 4.0 °C, which agrees with our previous findings for other 2′-T analogues (Fig. 2B, $\mathbf{U_T}$ and $\mathbf{A_T}$).[10,12] In the mismatch study of $\mathbf{G_T}$, the 2′-T·G mispair was not sufficiently discriminated with −1.5 °C compared to −6.5 °C for the native mismatch. This drawback was also observed for other 2′-T monomers,[12,13] and work is currently ongoing to design a 2′-T analogue with a better specificity.[16] For the sake of completeness, the modified ON's were also hybridized to their RNA complements, however, as seen before,[10] significant destabilization of 11.0 °C to 13.5 °C were observed compared to the unmodified DNA·RNA hybrids (Fig. S16M+N, ESI†).

With the new monomers established to function as condensed dinucleotides, their interactions with opposite DhNA monomers in various motifs were examined (Fig. 3). Whereas incorporation of $\mathbf{G_G}$ opposite a natural 5′-CC unit led to a decrease in the T_m value of 4.0 °C, its incorporation opposite to $\mathbf{A_C}$ forming a 2′·2′ G·C base pair gave rise to an increase in the

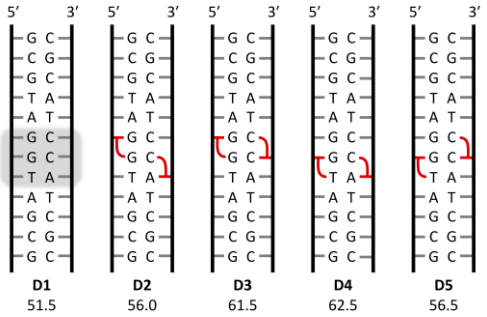


Fig. 3 Structures and melting temperatures (T_m, °C) of dsDNA duplexes containing one DhNA monomer in each strand. See Fig. 2 for conditions.

T_m of 4.5 °C (compare D2 to D1). An even more pronounced stabilizing effect was found for $\mathbf{G_G}$ opposite to the $\mathbf{C_C}$ monomer (D3, +10.0 °C) indicating two strong 1′·2′ G·C base pairs. A similar result was found using the complementary $\mathbf{G_T}$ and $\mathbf{A_C}$ (D4) with an increase of 11.0 °C and even the 1′·1′ base pair in D5 showed an increase in stability of 5.0 °C, which is more than just the additive effects of introducing $\mathbf{C_C}$ and $\mathbf{G_T}$ based on Fig. 2B.

In the study of the fully modified DhNA sequences (ON5–7, Table 1) we found a very low affinity towards DNA complements (D7–D9, Fig. 4), while no duplex formation was observed with RNA complements (Fig. S15, ESI†). This is not unexpected, as earlier studies showed that although dsDNA can adjust the backbone to accommodate single incorporations of DhNA monomers,[13] consecutive incorporations in the same strand decrease duplex stability.[12] The DhNA sequences were then matched to form two fully modified DhNA duplexes: one with alternating 1′·1′ and 2′·2′ pairs (D10) and one consisting only of alternating 1′·2′ pairs (D11). Whereas the native DNA duplex displayed a T_m of 33.0 °C, the two DhNA duplexes were remarkably more stable with T_m values of 48.0 °C and 60.0 °C, respectively. Evidently, the stable motifs of DhNA that was observed individually in dsDNA (Fig. 3) proved functional also in full scale, and we observe a 27.0 °C stability increase for the 5-mer DhNA duplex that holds 10 functional Watson–Crick base pairs as compared to its 10-mer DNA reference. To prove that reduced electrostatic repulsions between the backbones are to some extend responsible for the increased duplex stability, the

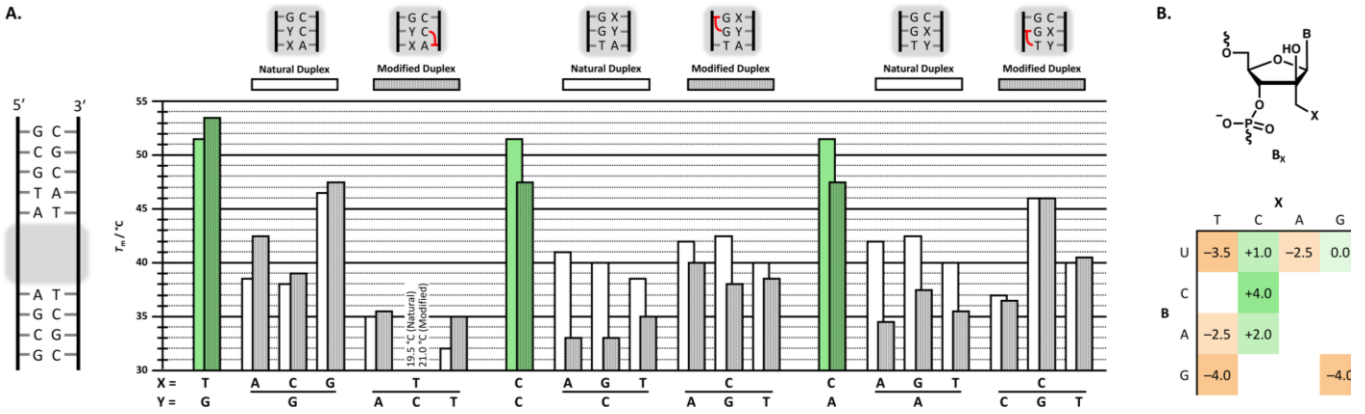


Fig. 2 (A) Structures and melting temperatures (T_m, °C) of matched and mispaired DNA duplexes containing $\mathbf{A_C}$, $\mathbf{G_G}$ or $\mathbf{G_T}$ measured at 1.5 µM concentrations of each DNA strand in a medium salt buffer (2.5 mM Na_2HPO_4, 5.0 mM NaH_2PO_4, 100 mM NaCl, 0.1 mM EDTA, pH 7.0). (B) Structure of DhNA monomers $\mathbf{B_X}$ and the change in melting temperature (ΔT_m, °C) associated with the incorporation of $\mathbf{B_X}$ into a 13 or 14-mer DNA duplex in exchange for the corresponding dinucleotide.[9-13]

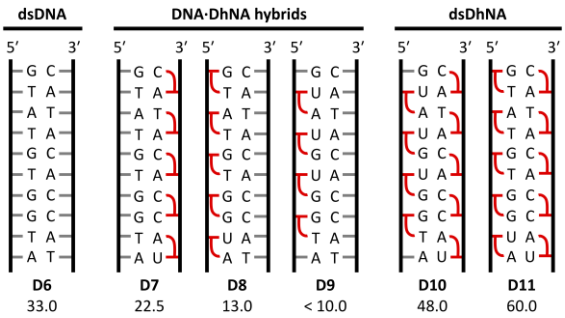


Fig. 4 Structures and melting temperatures (T_m, °C) of the dsDNA, DNA·DhNA hybrid and dsDhNA versions of the duplex sequence. See Fig. 2 for conditions.

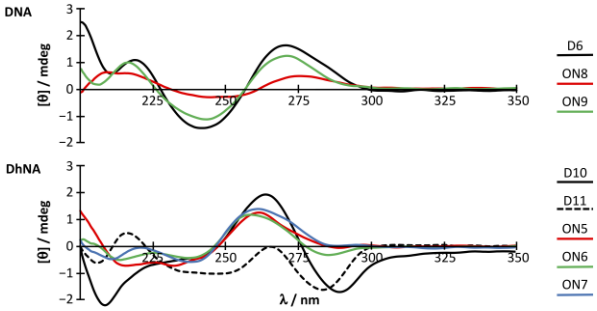


Fig. 6 CD spectra of single and double-stranded DNA and DhNA (referring to Fig. 4 and Table 1).

stabilities of the DhNA duplexes were determined at varying concentrations of NaCl from 0–500 mM (Fig. 5). Due to the lower number of anionic phosphates (18 in D6 compared to 9 in D10 and 8 in D11), the stability of the DhNAs should be less dependent on effective shielding. Indeed, the T_m values for the DhNAs were reduced by only 9.0–9.5 °C, when the NaCl concentration was decreased from 100 mM to 0 mM, whereas the DNA and RNA references were destabilized by 13.5–14.0 °C.

Finally, the duplexes were studied by circular dichroism (CD) spectroscopy. For duplexes containing either a single or two opposite modifications (Fig. 2 and 3), a general consistency with the B-type dsDNA was observed (Fig. S17 and S18, ESI†). For the fully modified DhNA duplexes, more severe deviation from the spectra of dsDNA, dsRNA and DNA·RNA hybrids was observed (Fig. S19, ESI†) more so for D11 than for D10 (Fig. 6). The corresponding single strands were also measured (Fig. 6) and the three DhNA strands all indicate some degree of pre-organization for duplex formation. Hence the increased duplex stability of DhNA might result from a combination of reduced electrostatic repulsion and pre-organization upon hybridization due to stacking of the DhNA monomers. Future structural investigations might enlighten this further.

In conclusion, we have presented three novel DhNA monomers including the first guanosine-based analogues, and we have established that they behave as condensed dinucleotides similar to previous monomers. We now have in hand 9 of the 16 possible monomers (Fig. 2B) representing all bases in both 1′ and 2′-positions. This made us able to construct for the first time duplexes that consist entirely of DhNA monomers. We have shown that a 5-mer DhNA duplex can hold 10 Watson–Crick base pairs and that it is far more thermally stable (+27.0 °C) than its native DNA equivalent. Very stable fully modified DhNA duplexes can be obtained despite the type (I or II, Fig. 1). Evidently, the concept of DhNAs for the purpose of condensing the molecular information of natural nucleic acids to shorter oligomers and a reduced number of charges has been proved viable. We are convinced that DhNA represents a new XNA with potential applications for information storage in nanobiotechnology or in therapeutics e.g. nucleic acid aptamers. Moreover, its distinctive design and functionality demonstrate the versatility of nucleic acid inspired polymers.

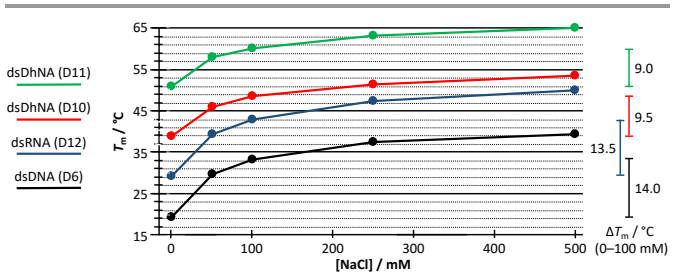


Fig. 5 Melting temperatures (T_m, °C) of duplexes D6 and D10–D12 at varying (0–500 mM) NaCl concentrations (referring to Fig. 4 and Fig. S15, ESI†).

This work was supported by a grant from the Independent Research Fund Denmark – Natural Sciences (DFF – 7014-00052). There are no conflicts to declare.

Notes and references

1. L. K. McKenzie, R. El-Khoury, J. D. Thorpe, M. J. Damha and M. Hollenstein, *Chem. Soc. Rev.*, 2021, **50**, 5126-5164.
2. K. Morihiro, Y. Kasahara and S. Obika, *Mol. BioSyst.*, 2017, **13**, 235-245.
3. M. Vybornyi, Y. Vyborna and R. Häner, *Chem. Soc. Rev.*, 2019, **48**, 4347-4360.
4. V. B. Pinheiro and P. Holliger, *Trends Biotechnol.*, 2014, **32**, 321-328.
5. P. K. Sharma, P. Kumar and P. Nielsen, *Aust. J. Chem.*, 2016, **69**, 1094-1101.
6. P. Kielkowski, R. Pohl and M. Hocek, *J. Org. Chem.*, 2011, **76**, 3457-3462.
7. M. K. Gupta, B. R. Madhanagopal and K. N. Ganesh, *J. Org. Chem.*, 2021, **86**, 414-428.
8. U. K. Mishra, Y. S. Sanghvi, M. Egli and N. G. Ramesh, *J. Org. Chem.*, 2021, **86**, 367-378.
9. P. Kumar, P. K. Sharma and P. Nielsen, *J. Org. Chem.*, 2014, **79**, 11534-11540.
10. M. Hornum, P. K. Sharma, C. Reslow-Jacobsen, P. Kumar, M. Petersen and P. Nielsen, *Chem. Commun.*, 2017, **53**, 9717-9720.
11. K. Beck, C. Reslow-Jacobsen, M. Hornum, C. Henriksen and P. Nielsen, *Bioorg. Med. Chem. Lett.*, 2019, **29**, 740-743.
12. K. M. Beck, M. B. Krogh, M. Hornum, P. T. Ludford, Y. Tor and P. Nielsen, *Org. Biomol. Chem.*, 2020, **18**, 7213-7223.
13. M. Hornum, J. Stendevad, P. K. Sharma, P. Kumar, R. B. Nielsen, M. Petersen and P. Nielsen, *Chem. Eur. J.*, 2019, **25**, 7387-7395.
14. J. C. Chaput and P. Herdewijn, *Angew. Chem. Int. Ed.*, 2019, **58**, 11570-11572.
15. Y. Hayakawa, M. Hirose and R. Noyori, *J. Org. Chem.*, 1993, **58**, 5551-5555.
16. K. M. Beck, L. Ruder, T. S. Nicolai, R. L. Pham, N. A. Risgaard, M. Hornum and P. Nielsen, *Eur. J. Org. Chem.*, 2021, 1949-1957.