Condensing the information in DNA with double-headed nucleotides†

Mick Hornum, Pawan K. Sharma, Charlotte Reslow-Jacobsen, Pawan Kumar, Michael Petersen and Poul Nielsen*

A normal duplex holds as many Watson–Crick base pairs as the number of nucleotides in its constituent strands. Here we establish that single nucleotides can be designed to functionally imitate dinucleotides without compromising binding affinity. This effectively allows sequence information to be more compact and concentrated to fewer phosphates.

Beside its biological functionality, DNA has emerged as a convenient scaffold for two- and three-dimensional objects in nanotechnology1,2 and information storage.3,4 Recently, we5–15 and others16–20 have examined the potential of introducing a second nucleobase onto the existing nucleotide scaffold as a means of increasing the molecular diversity of DNA. These so-called bis-headed or double-headed nucleotides enable various intra or extra-helical contacts depending on the location of the second nucleobase.21 Since such double-headed nucleotides contain two nucleobases within a single nucleotide unit, they can in principle convey twice the amount of information than their single-headed counterparts. Our previous work has shown8 that structures where the second nucleobase is attached to the 2'-position via a methylene linker, neatly positions both nucleobases in the duplex core in a coplanar stacking orientation. Hereby, both nucleobases in the double-headed nucleotide become available for Watson–Crick contacts.

We have previously developed a simple synthetic approach for introducing thymine or adenine onto the 2'-position of uridine via a spiroepoxide intermediate.13 This epoxide may be regioselectively opened by a nucleobase thus forming a double-headed arabinoc-configuration nucleoside. In continuation of this work, we here report the recognition potential of all four canonical nucleobases in this design. Specifically, the complete package of double-headed nucleotides, i.e. U\textsubscript{T}, U\textsubscript{C}, U\textsubscript{G} and U\textsubscript{A} (Fig. 1A), now finalizes the picture and ascertains the general propensity for both nucleobases to communicate efficiently with a cognate strand. Fig. 1B shows a snapshot from an MD simulation, and demonstrates how the double-headed design allows both nucleobases to interact with nucleobases in the opposite strand.
Whereas the syntheses of \( U_T \) and \( U_A \) are reported before,\(^{13} \) we herein report the successful syntheses of \( U_C \) and \( U_G \) (Fig. 1C). Starting from the protected 2'-spiroepoxy uridine intermediate, the \( U_C \) and \( U_G \) phosphoramidites for oligonucleotide synthesis were obtained in 40% and 31% overall yields, respectively, over four steps. The nucleophilic opening of the epoxides and the subsequent desilylation–tritylation–phosphitylation sequences all proceeded with complete regioselectivity and in good yields. For \( U_G \), the ring opening and the oligonucleotide synthesis were most readily achieved using a doubly protected guanine\(^{22} \) bearing \( O^6\)-allyl/\( N^3\)-isobutyl groups (Fig. 1C). The \( O^6\)-allyl group was easily removed by treating the final solid support-bound oligonucleotide with a mixture of \( \text{Pd}(\text{PPh}_3)_4 \) and \( \text{Et}_2\text{NH-HCO}_2 \) in \( \text{CH}_3\text{Cl}_2 \).\(^{22} \) For \( U_C \), \( N^1\)-benzoyl protected cytosine was successfully used. See ESI‡ for synthetic details.

To evaluate the dinucleotide behaviours of \( U_T \), \( U_C \), \( U_G \) and \( U_A \), they were individually placed centrally in a 13-mer sequence, \( 5'-\text{dGCTCACUCCTCCCA} \). Upon hybridization to appropriate complements in medium salt PBS pH 7.0 buffer, the melting temperatures (\( T_m \)) of the duplexes were determined by UV\(_{260}\) melting curve analysis, and compared to those of the corresponding native 13- and 14-mer duplexes (Fig. 2).

While the \( T_m \) of the 13-mer unmodified DNA duplex was 50 °C (Fig. 2A), the \( T_m \) values of the corresponding modified duplexes were 6.0–7.5 °C lower (Fig. 2B). This is expected considering the lack of base pairing possibility for the appended nucleobase, causing it instead to destabilize the duplex. More interestingly, the same 13-mer modified oligonucleotides were hybridized to natural 14-mer oligonucleotides, hereby allowing the additional nucleobases (\( X \)) to interact with the natural nucleobases (\( Y \)) placed directly across (Fig. 2D). Since this interaction gives rise to 14-mer duplexes, the \( T_m \) should be compared to the corresponding unmodified 14-mer duplexes (Fig. 2C). In Fig. 2, the \( T_m \) values of the green diagonals correspond to \( X-Y \) Watson–Crick matches. By comparing Fig. 2C and D, it is apparent that the native duplex stability is not compromised whenever the double-headed nucleotides take the roles as dinucleotides. In fact, both new monomers (\( U_C \) and \( U_G \)) display neutral effects on the duplex stability. Notably, when the additional C and G bases in \( U_C \) and \( U_G \) are positioned across mismatched bases, the \( T_m \) plummets by 14.0–16.0 °C for \( U_C \) and 10.0–11.5 °C for \( U_G \) (Fig. 2D), thus indicating a relative high power of mismatch discrimination compared to \( U_T \) and \( U_A \). In fact, the base pairing fidelity is essentially in line with the unmodified reference duplex, where mismatches in the same location are discriminated by 14.0–16.0 °C for 5'-UC and 8.0–10.5 °C for 5'-UG (Fig. 2C).

Having validated that the double-headed nucleotides \( U_T \), \( U_C \), \( U_G \) and \( U_A \) are functional dinucleotides, we next introduced the double-headed nucleotides in so-called (+1)-zipper arrangements in the centre of an 11-mer DNA duplex. This enables the formation of a twelfth Watson–Crick base pair (Fig. 3B). Remarkably, these 11-mer duplexes that contain 12 Watson–Crick base pairs show \( T_m \) values 5.5–9.0 °C higher than the regular, genuine 12-mer duplexes (compare Fig. 3A and B). This data strongly indicates that favourable T-A and G-C base pairs are formed in the centre of the duplex, and that these additional base pairs stabilize the duplex more than regular base pairs. This net stabilizing effect might be attributed to the reduced torsional freedom and electrostatic repulsion of the phosphodiester backbone accompanying the decreased number of phosphates. Specifically, the (+1) zipper constructs that carries \( U_T \) and \( U_A \) give rise to increases in the \( T_m \) of 6.0–6.5 °C,\(^{13} \) whereas the corresponding (+1) zipper constructs carrying \( U_C \) and \( U_G \) give rise to increases of 5.5–9.0 °C, compared to the regular 12-mer duplexes. It is not clear why there is a 3.5 °C incongruence between the 5'-U\(_C\)A3'-UC and 5'-U\(_G\)A3'-UA motifs, since nearest-neighbour effects are expected to be almost similar for the two designs. Nevertheless, the present results show that all four double-headed nucleotides – \( U_T \), \( U_C \), \( U_G \) and \( U_A \) – can be used to form very stable duplexes that contain an additional base pair.

The fidelity of recognition of the twelfth base pair is evaluated by examining the off-diagonal values in Fig. 3B. As seen, Watson–Crick pairing rules are obeyed, and the zipper constructs in general display good mismatch discrimination. Only the G-T mismatch base pair is poorly discriminated with a drop in the \( T_m \) of only 2.5–3.0 °C compared to duplexes with AT matched base pairs. A similar stable G-T mismatch was also observed when \( U_T \),

---

**Fig. 2** Structures of DNA duplexes and their hybridization data (\( T_m \), values*). (A) 13-Mer reference duplex. (B) Same as A except that \( U \) is replaced with one of \( U_T \), \( U_C \), \( U_G \) and \( U_A \). (C) Reference duplex containing a new central base pair \( X-Y \), where \( X \) and \( Y \) are combinations of the natural nucleobases. (D) Same as C except that the central 5'-UX dinucleotide is replaced with one of \( U_T \), \( U_C \), \( U_G \) and \( U_A \). \( T_m \) values (*in °C) were obtained from the melting curves (\( A_{260} \) vs. temperature) recorded in 2.5 mM \( \text{Na}_2\text{HPO}_4 \), 5.0 mM \( \text{NaH}_2\text{PO}_4 \), 100 mM \( \text{NaCl} \), 0.1 mM \( \text{EDTA} \), pH 7.0 using 1.0 μM of each strand.

---

9718 | Chem. Commun., 2017, 53, 9717–9720 This journal is © The Royal Society of Chemistry 2017
is positioned across a 3'-UG dinucleotide (Fig. 2D). Nevertheless, the G-T mismatched zipper structure is still 6.5–9.5 °C less stable than the corresponding G-C matched zipper. Notably, both T-T and A-A mismatches are efficiently discriminated by ~10 °C compared to the A-T pair, whereas the C-C and G-G mismatches are discriminated by 7.5–11.0 °C and 14.0–17.5 °C, respectively, compared to the G-C pair. The C-C pair may form a hemiprotonated23 (‘‘i-motif type’’) C-C base pair and thus be partially stabilized.

In addition, the ability of these modified 11-mer oligo-nucleotides were examined for their binding affinity towards natural 12-mer oligonucleotides (Fig. 3C). In accordance with the results from the 14-mer design reported in Fig. 2, the duplex stabilities were not compromised by the replacement of the central 5'-UX dinucleotides with any of the double-headed nucleotides. In this case, the duplex stability was in fact stabilized by 2.0 °C when 5'-UC was replaced with UC (compare Fig. 3C with Fig. 3A). Minor destabilizations were seen with UA, UC and UC (drops in the Tm value of −2.5 °C, −1.5 °C and −1.0 °C, respectively, relative to 5'-UT, 5'-UA and 5'-UG).

In continuation of these results with dsDNA, we examined the suitability of the double-headed nucleotides in the iso-sequential DNA:RNA. The measured Tm values are shown in Fig. 4. By comparing the green diagonal in Fig. 4A with Fig. 4B, it is clear that the DNA:RNA duplex stabilities are significantly reduced whenever the UT, UC, UA and UA monomers are introduced in place of the corresponding regular dinucleotides. The largest destabilization was observed with UT (ΔTm = −10.5 °C) and the smallest destabilization with UT (ΔTm = −3.5 °C). The significant destabilizations are in stark contrast to the dsDNA case. In addition, the double-headed nucleotides exhibit poor mismatch discrimination with as low as 0.5 °C in the case of UA (Fig. 4B). These results manifests that the dsDNA duplex is the perfect structure for the double-headed nucleotides (Fig. 2 and 3), whereas the geometry of DNA:RNA duplexes do not easily accommodate the compressed dinucleotide structure. This markedly inequivalence between dsDNA and DNA:RNA most likely arises from a combination of two factors; (1) the tilted base pairs of the A-type helix prohibit sufficient stretching of the backbone to house the double-headed structure, and (2) the organization of the double-headed nucleotide in the 2'-endo conformation is essential in order to enable Watson–Crick contacts for both nucleobases. Accordingly, the double-headed nucleotides can also be used as a tool for obtaining preferential recognition of DNA as compared to RNA.

In conclusion, the present results clearly indicate that both nucleobases of all four double-headed nucleotides, UT, UC, UA, and UA, efficiently communicate with the nucleobases of an opposite DNA strand. Importantly, all contacts obey Watson–Crick base

---

**Fig. 3** Structures of DNA duplexes and their hybridization data (Tm values). (A) 12-mer reference duplex, where X and Y are combinations of the natural nucleobases. (B) Same as A except that 5'-UX and 5'-UY are replaced with combinations of UT, UC, UA and UA. (C) Same as A except that the central 5'-UX dinucleotide is replaced with one of UT, UC, UA and UA. Tm values (in °C) were obtained from the melting curves (A260 vs. temperature) recorded in 25 mM Na2HPO4, 5.0 mM NaH2PO4, 100 mM NaCl, 0.1 mM EDTA, pH 7.0 using 1.0 µM of each strand.

---

**Fig. 4** Structures of DNA:RNA duplexes and their hybridization data (Tm values). (A) 12-mer reference DNA:RNA duplexes, where X and Y are combinations of the natural nucleobases. (B) Same as A except that the central 5'-UX dinucleotide is replaced with one of UT, UC, UA and UA. Tm values (in °C) were obtained from the melting curves (A260 vs. temperature) recorded in 25 mM Na2HPO4, 5.0 mM NaH2PO4, 100 mM NaCl, 0.1 mM EDTA, pH 7.0 using 1.0 µM of each strand.
pairing rules, and mismatches opposite to U_c and U_g are very well-discriminated in dsDNA. These results demonstrate that the malleable backbone of the dsDNA duplex easily adapts to the situation where two modified 11-mer duplexes hybridize to form 12 base pairs. These matched zipper contacts lead to markedly thermostabilization (up to +9.0 °C) compared to the regular duplexes; i.e. tighter binding is achieved when the same number of base pairs in the ladder is concentrated onto fewer phosphates. The stability and specificity of this additional base pair indicates that the nucleic acid backbone fully adapts to the extension. We expect that all double-headed nucleotides based on other nucleosides than uridine can be accommodated in this way. In this manner, the duplex carries a larger number of Watson–Crick base pairs per phosphate unit, and information can therefore be delivered using a shorter sequence. As such, this approach could lead to a new paradigm in nucleic acid communication, and in the use of DNA for high density storing of information.

Work is now in progress to explore the scope of the monomers, e.g. how many of these artificial motifs can be accommodated by the duplex, and how well are non-canonical and chemically modified nucleobases accepted in double-headed nucleotides. Also, current activities are directed towards synthesizing other nucleobase combinations and investigating the eligibility of the double-headed nucleotides in e.g. polymerase reactions.

Conflicts of interest

There are no conflicts of interest to declare.

Notes and references