Triplex-forming properties and enzymatic incorporation of a base-modified nucleotide capable of duplex DNA recognition at neutral pH

David A. Rusling [ID]*

School of Biological Sciences, University of Southampton, Southampton, Hampshire SO17 1BJ, UK

ABSTRACT

The sequence-specific recognition of duplex DNA by unmodified parallel triplex-forming oligonucleotides is restricted to low pH conditions due to a necessity for cytosine protonation in the third strand. This has severely restricted their use as gene-targeting agents, as well as for the detection and/or functionalisation of synthetic or genomic DNA. Here I report that the nucleobase 6-amino-5-nitropyridin-2-one (Z) finally overcomes this constraint by acting as an uncharged mimic of protonated cytosine. Synthetic TFOs containing the nucleobase enabled stable and selective triplex formation at oligopurine-oligopyrimidine sequences containing multiple isolated or contiguous GC base pairs at neutral pH and above. Moreover, I demonstrate a universal strategy for the enzymatic assembly of Z-containing TFOs using its commercially available deoxyribonucleotide triphosphate. These findings seek to improve not only the recognition properties of TFOs but also the cost and/or expertise associated with their chemical syntheses.

INTRODUCTION

Triplex-forming oligonucleotides (TFOs) bind within the major groove of oligopurine-oligopyrimidine duplex sequences through the formation of specific base triplets, generating a triplex structure (Figure 1A) (1–2). Experiments have shown that, at least in principle, TFOs can be exploited as gene-targeting agents for the transient control of gene activity in cell culture and *in vivo*, as well as for the permanent alteration of gene expression through directed mutagenesis (3,4). They have also been used *in vitro* to detect and/or isolate plasmid and genomic DNA, as well as to introduce functionality into artificial DNA nanostructures assembled by crossover strand exchange (5,6). Most of these studies have utilised purine-rich TFOs that bind in an antiparallel orientation relative to the central purine strand of their target duplex sequence (7). However, this motif is hampered by the tendency of purine-rich oligonucleotides to adopt other non-canonical structures, such as GA duplexes and G-quadruplexes, that compete with triplex formation. In addition, the triplets generated in antiparallel triplexes are not isomorphic and lead to structural distortions at the junctions between consecutive triplets, decreasing triplex stability (8). Greater success is likely to be achieved with pyrimidine-rich TFOs that bind in a parallel orientation and form isomorphic triplets with their target duplex (1,2). However, the efficacy of parallel triplexes is hampered by their poor binding affinity at neutral pH. These oligonucleotides require low pH conditions (pH < 6.0) necessary for imino protonation of the N3 position of cytosine (pK_a ~4.5) and formation of a second Hoogsteen hydrogen bond within the C$^+$-GC triplet (Figure 1B). [The notation X-RY used here refers to a triplet in which the TFO base X interacts with the duplex base pair RY, forming hydrogen bonds to base R.] Nevertheless, at low pH, parallel triplexes composed of protonated C$^+$-GC and T-AT triplets can be more stable than the underlying duplex, i.e. the affinity of the third strand for its duplex target is greater than the affinity of the W–C duplex strands for each other (9).

Much synthetic effort has been applied to developing cytosine analogues that alleviate the stringent pH requirement of parallel triplex formation but with limited success (10). Various nucleobases have been synthesised with a higher basicity than cytosine (e.g. 5-methylcytosine, 2-aminopyridine, etc.) (11–14). However, their reported pK_a values still fall short of what is required for binding at neutral pH. In addition, the use of charged bases, including cytosine, to target sequences composed of adjacent GC base pairs can be problematic, due to proton competition between contiguous bases located in the third strand (15,16). To address this, several uncharged nucleobases have been prepared that possess the appropriate hydrogen bonding groups without the need for protonation (e.g. pseudoisocytosine, 6-oxo-cytosine, pyrazine, etc.) (17–20). However,

*To whom correspondence should be addressed. Tel: +44 2380 594 373; Email: d.a.rusling@soton.ac.uk
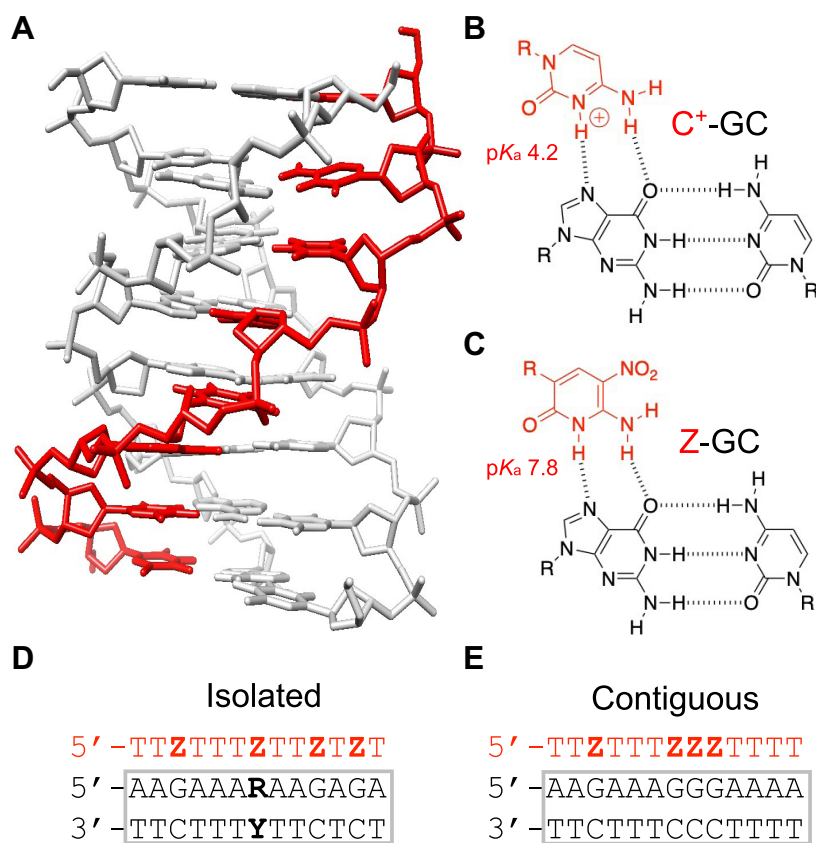
Figure 1. Triplex formation with the nucleobase 6-amino-5-nitropyridin-2-one (**Z**). (**A**) NMR structure of a parallel triplex formed by the binding of a third strand within the major groove of an oligopurine–oligopyrimidine duplex sequence (PDB code: 13DX). (**B**) Chemical structure of the C$^+$-GC triplet used in combination with T-AT under low pH conditions (<6.0). (**C**) Putative chemical structure of the Z-GC triplet used in combination with T-AT under neutral pH conditions; (**D, E**) Sequence of the two triplex motifs under study. Both triplexes contain four Z-GC triplets located at either isolated (**D**) or contiguous positions (**E**) within the structure. The triplex containing isolated modifications was also used for selectivity studies by positioning a single Z modification in the third strand opposite each base pair in turn, generating Z-RY triplets (highlighted RY in bold). TFOs and their bases are shown in red and the duplex sequences and base pairs in black.

the triplets formed by such bases are less stable at low pH than protonated C$^+$-GC, since the presence of a positive charge can partially alleviate the charge repulsion between the three negatively charged strands (15,16). In addition, structural alterations made to the pyrimidine ring system can led to tautomeric ambiguity, as well as a reduction in the chemical stability of the oligonucleotide, for example by epimerization (18,20). Other approaches have sought to increase triplex stability indirectly by attaching triplex-stabilising ligands and/or positive charged groups to the sugar and backbone of TFOs (10,21).

Another significant hurdle that has prevented adoption of the aforementioned analogues by the TFO community is a requirement for the synthesis of their bespoke phosphoramidites, several of which are not trivial. As far as I'm aware, these analogues are not routinely available without substantial cost, either as their phosphoramidites or within oligonucleotides. Further still, none of these analogues have been shown to be compatible with the enzymatic assembly of specific oligonucleotide sequences, which would greatly facilitate the range of TFOs that could be assembled in-house. It is clear that there is still very much a need to

develop analogues that are not only capable of extending triplex formation to neutral pH, but also reduce the cost and/or expertise associated with their incorporation into TFOs.

In this work, a commercially available nucleobase 6-amino-5-nitropyridin-2-one (**Z**) was investigated for its potential to overcome each of these limitations. **Z** was first developed by the Benner group for use as part of an artificially expanded genetic information system (AEGIS) (22,23). It is an uncharged C-glycoside mimic of protonated cytosine with an electron withdrawing nitro group attached to the 5-position. Since the pK_a value of its ring nitrogen is *ca.* 7.8, it exhibits the necessary hydrogen bonding pattern to recognise a GC base pair and form a Z-GC triplet at neutral pH (Figure 1C). Moreover it has been shown that, under appropriate experimental conditions, the deoxyribonucleotide triphosphate of the nucleobase can be incorporated into an oligonucleotide through primer extension by a polymerase (24–26). By exploiting these favourable properties this study demonstrates a universal strategy for the enzymatic assembly of modified TFOs capable of duplex DNA recognition at neutral pH.

MATERIALS AND METHODS

Oligonucleotides

The full sequence of the oligonucleotides used in this study are shown in Supplementary Figure S1. Unmodified and modified oligonucleotides containing the synthetic nucleobase 6-amino-5-nitropyridin-2-one (dZ) were purchased from Sigma Aldrich and Firebird Biomolecular Sciences, respectively. Whilst unmodified and modified deoxyribonucleotide triphosphates of 6-amino-5-nitropyridin-2-one (dZTP) were purchased from Promega and Firebird Biomolecular Sciences, respectively.

DNA fragment for DNase I footprinting

The 73-mer DNA fragment containing the embedded TFO target sequence used in footprinting experiments is shown in Supplementary Figure S1f. This was prepared by blunt-end cloning of the duplex shown in Supplementary Figure S1f into the SmaI site of pUC18. Incorporation of the intended duplex was confirmed by sequencing (MWG Eurofins). The recombinant plasmid was transformed into competent Escherichia coli TG2 cells and the plasmid isolated using a Qiagen Miniprep kit. The plasmid was subsequently digested by HindIII and SacI (New England Biolabs) and radiolabelled at the 3′-end of the HindIII site using exo-Klenow fragment (New England Biolabs) and [α-^{32}P]dATP (Perkin Elmer). The fragment was separated from the remainder of the plasmid on an 8% (w/v) non-denaturing polyacrylamide gel. After elution, the fragment was dissolved in 10 mM Tris–HCl pH 7.0 to give ∼10 c.p.s./μl as determined on a hand held Geiger counter (<10 nM DNA).

DNase I footprinting

TFO binding was determined by footprinting the 73-mer DNA fragment shown in Supplementary Figure S1f. This was performed by mixing 1.5 μl radiolabelled DNA with 3 μl TFO dissolved in either 50 mM sodium acetate containing 10 mM MgCl$_2$ at pH 5.0, or 40 mM tris acetate containing 10 mM MgCl$_2$ at pH 7.5. Final TFO concentrations varied between 0.03 and 30 μM and the complexes were left to equilibrate overnight at 20°C. DNase I digestion was carried out by adding 2 μl DNase I (typically 0.01 U/ml) dissolved in 20 mM NaCl containing 2 mM MgCl$_2$ and 2 mM MnCl$_2$. The reaction was stopped after 1 min by adding 4 μl of 80% formamide containing 10 mM EDTA, 10 mM NaOH and 0.1% (w/v) bromophenol blue. Products of digestion were separated on 12% (w/v) polyacrylamide gels containing 8 M urea. Samples were heated to 100°C for 3 min, before rapidly cooling on ice and loading onto the gel. Polyacrylamide gels (40 cm long and 0.3 mm thick) were run at 1500 V for ∼2 h and then fixed in 10% (v/v) acetic acid. These were transferred to Whatman 3MM paper and dried under vacuum at 86°C for 1 h. The dried gels were subjected to phosphorimaging using a Molecular Dynamics Typhoon PhosphorImager.

Fluorescence melting

Thermal melting profiles for triplexes generated with the duplexes shown in Supplementary Figure S1g, h were determined using SYBR green I (Thermofisher) and a Roche LightCycler. Oligonucleotides were prepared in 10 mM sodium cacodylate containing 10 mM MgCl$_2$ at the appropriate pH. Final concentration of the duplex was 1 μM and final TFO concentrations varied between 0.1 and 10 μM depending on the experiment. Samples were prepared in a total volume of 20 μl. Fluorescence emission from SYBR green I was recorded at 522 nm after excitation at 488 nm. Fluorescence profiles were obtained for both annealing and melting of the complexes between 30 and 95°C at a ramp rate of 0.2°C/min. (Although the slowest rate of continuous temperature change in the LightCycler is 0.1°C/s, slower melting profiles were obtained by increasing the temperature in 1°C steps, leaving the samples to equilibrate for a set amount of time.) Melting temperatures (T_m) were determined from the first derivatives of the profiles using the software provided with the machine and typically differed by <0.5°C between experiments.

Electrophoretic mobility shift assay (EMSA)

TFO binding was determined for the duplexes shown in Supplementary Figure S1g by subjecting the complexes to an EMSA. Oligonucleotides were prepared in 40 mM tris acetate containing 10 mM MgCl$_2$ at pH 7.5. Final concentration of the duplex was 1 μM and the TFO concentrations varied between 0.1 and 10 μM depending on the experiment. Samples were prepared in a total volume of 20 μl. The complexes were first heated to 95°C for 5 min, slowly cooled to room temperature and left to equilibrate for 16 h. Samples were separated on non-denaturing 15% (w/v) polyacrylamide gels run at 400 V for ∼2 h in 40 mM tris-acetate running buffer. Gels were subjected to post-staining with GelRed (Biotium) and visualised using a Gel Doc imaging system (27).

Enzymatic assembly of oligonucleotides

Oligonucleotides were assembled by primer extension of a template strand containing the appropriate W–C duplex complement to the TFO. Oligonucleotides were dissolved in Therminator® reaction buffer (New England Biolabs) at pH 8.8. Final concentration of the primer and template strands was 5 μM in a total volume of 20 μl. The strands were first heated to 95°C for 5 min, slowly cooled to room temperature, and then left at 4°C to equilibrate for 2 h. Deoxyribonucleotide triphosphates (e.g. dCTP/dTTP or dTTP/dZTP) were then added at a final concentration of 250 μM. The reaction mixture was pre-incubated at 72°C for 30 s before addition of 2 units of Therminator® polymerase (New England Biolabs). Samples were left >2 h to allow the extension reaction to reach completion. The phosphate-labelled template strand was degraded by adding 10 units of Lambda exonuclease (New England Biolabs) at 37°C for 2 h before its inactivation at 80°C for 10 min. For triplex binding experiments the assembled TFOs were used directly or prepared in batches, combined and concentrated by SpeedVac.

RESULTS

Experiments with synthetic TFOs

Since the triplex-forming properties of Z have not been established its use within synthetic TFOs assembled through phosphoramidite chemistry was first investigated. These were characterised alongside equivalent unmodified oligonucleotides bearing cytosines at the same positions. Initial studies examined the targeting of an oligopurine-oligopyrimidine duplex sequence 13 base pairs (bp) in length containing four isolated GC base pairs using the TFO sequence shown in Figure 1D. The targeting of such a sequence was previously shown to be strongly pH dependent (28,29).

TFO binding and selectivity were examined by the enzymatic protection assay DNase I footprinting (30). Various concentrations of each oligonucleotide were incubated with a 73-mer duplex fragment containing the 13-bp triplex target sequence that was ^{32}P-labelled at the 3′-end of pyrimidine-containing strand (Supplementary Figure S1f). The complexes were then digested by DNase I and the labelled fragments separated by denaturing polyacrylamide gel electrophoresis. DNase I is a double-strand specific endonuclease that generates single strand nicks in the phosphodiester backbone by cleaving the O3′P bond. Cleavage in the presence of TFO bound to its target sequence resulted in missing bands or a 'footprint' on account of the TFO occluding the action of the enzyme at these positions. This technique was therefore used to reveal regions that had undergone triplex formation.

The interaction of the TFOs at pH 5.0 which facilitates cytosine protonation in the control oligonucleotide were characterised first. Representative cleavage patterns for the fragment in the absence and presence of the oligonucleotides are shown in Figure 2A. In both cases, clear footprints are evident for the interaction of each TFO that, as expected, disappear as the concentration of the oligonucleotide is decreased (lanes 2–8). By comparison with a Maxim-Gilbert marker (lane 1) it can be seen that the footprints are located at the intended target sequences (boxed regions). In each case, the binding site is slightly overestimated due to the large size of the DNase I enzyme, but the cleavage pattern remains unaltered for the rest of the fragment. The concentrations of TFOs used in each gel are identical and visual comparison between gels reveals that the footprint produced by the Z-containing TFO persists to a lower concentration than the unmodified oligonucleotide. This was surprising, since at low pH it was expected that protonated cytosine would bind more strongly to GC than Z, which lacks a positive charge on its ring system. Experiments were also undertaken to examine whether the Z-containing TFO could bind in the absence of magnesium (Supplementary Figure S2). Although the affinity of the TFO is reduced, a clear footprint is seen at the oligonucleotide binding site that persists to the low micromolar range.

The binding of the TFOs at pH 7.5 where cytosines in the control oligonucleotide will not be protonated was examined next. Representative cleavage patterns for the fragment in the absence and presence of the control and modified TFOs are shown in Figure 2B. This time, footprints

are only observed for the Z-containing TFO and no interaction of the unmodified TFO can be seen, even at oligonucleotide concentrations as high as 30 μM. Moreover, as with those obtained at pH 5.0, the footprints persist well into the nanomolar range, and there is no apparent loss of binding affinity as the pH is increased from 5.0 to 7.5. This clearly demonstrates that oligonucleotides containing Z enable very stable triplex formation under neutral pH conditions.

Since these experiments were undertaken on relatively short DNA fragments, experiments were also conducted to determine whether the Z-containing TFO was capable of binding to a unique site located within a DNA plasmid; a more physiologically relevant target for the TFO (Supplementary Figure S3). These experiments were based on a simple restriction endonuclease (REase) protections assay; the TFO target sequence was positioned adjacent to a restriction site for the REase so that upon binding of the TFO the underlying plasmid DNA would be protected from digestion by the enzyme. These experiments revealed that the TFO was not only capable of binding and protecting the plasmid from digestion at pH 7.8 but also that this could be achieved at the relatively high temperature of 65°C (i.e. the optimal cleavage temperature for the REase).

To examine the pH dependence of the nucleobase over a wider pH window the thermal stabilities of the triplexes were determined by fluorescence melting studies using SYBR green I and a Roche LightCycler® (31). SYBR green I fluoresces upon binding to duplex and triplex DNA and the subsequent melting of the complexes leads to a decrease in fluorescence signal recorded at 522 nm (32). The melting temperature (T_m) of the complexes was determined from the mid-point of the melting profiles using first derivatives. For these experiments the TFO target site was embedded within a 31-mer duplex and the oligonucleotides used in these experiments are shown in Supplementary Figure S1g. Complexes were first annealed and then melted at 0.2 °C/minute and no hysteresis was observed between fluorescence profiles. This confirmed that, as previously reported, the Z nucleobase does not undergo significant epimerization at high temperatures (22,23).

The melting profiles of the complexes at various pH values are shown in Figure 3A and the T_m values calculated from these in Table 1. In the absence of TFO the duplex melts with a single transition yielding T_m values of between 73 and 75°C depending on the pH value tested. These are close to that determined under similar ionic conditions using T_m prediction software (www.idtdna.com/calc/analyzer/) and demonstrated that as expected, duplex stability is only slightly affected by a change in pH. The same experiments were then repeated in the presence of the C-containing oligonucleotide and the melting profiles are shown in the graph on the left of the figure. At pH 5.0 the melting curve (red curve) for the complex is shifted to the right of that obtained for the duplex alone (grey curve), yielding an increase in T_m value of *ca.* 1.2°C. This demonstrates that the triplex formed by the association of the oligonucleotide was of a similar stability to that of the underlying duplex, which is three times longer than the TFO target sequence. In contrast, as the pH is increased to 6.0 the stability of the triplex decreases and the melting profile
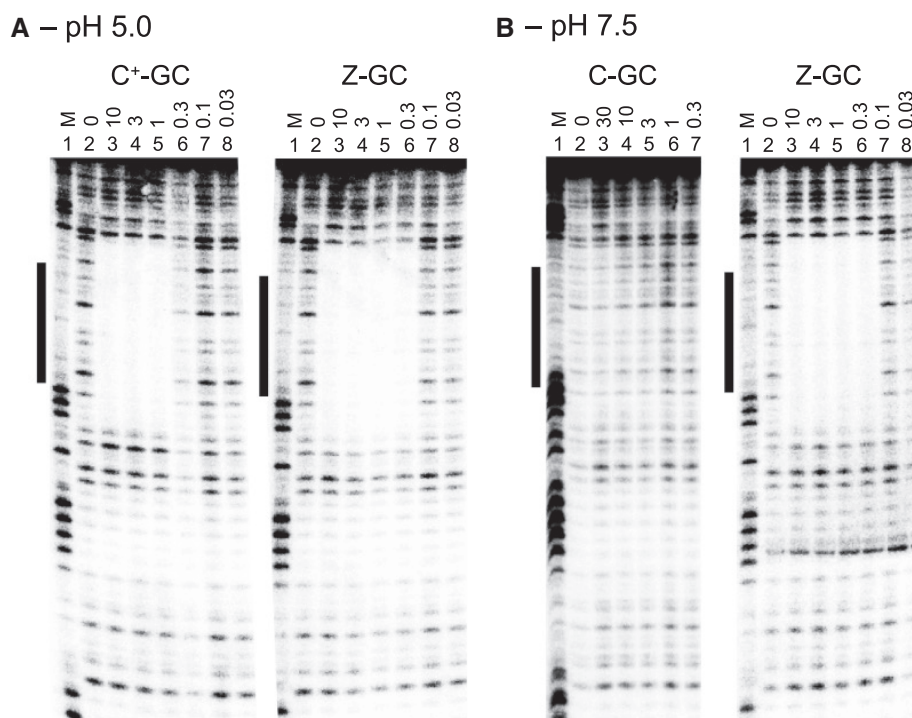
Figure 2. Binding and selectivity. DNase I cleavage patterns for the 73-mer duplex fragment containing the embedded TFO target sequence in the absence and presence of TFOs designed to generate C-GC and Z-GC triplets. Experiments were performed in sodium acetate buffer containing magnesium at pH 5.0 (**A**) or tris acetate buffer containing magnesium at pH 7.5 (**B**). TFOs were incubated with the fragment overnight at 4 °C at a final concentration of 10, 3, 1, 0.3, 0.1 and 0.03 μM as indicated before digestion by the enzyme. Samples were separated on a denaturing 12% polyacrylamide gel and subjected to phosphorimaging. The pyrimidine-containing strand of the fragment was labelled with ^{32}P at its 3′-end. The TFO target sequence is shown by the boxes adjacent to each gel and was determined by comparison with bands in a Maxim-Gilbert marker lane (labelled 'M').

for the complex exhibits two transitions (pink curve); one at high temperature that can be attributed to the melting of the underlying duplex (with the same T_m as the duplex alone), and one ca. 20°C lower, due to the dissociation of the TFO. As expected, melting profiles for the complexes at pH 7.0, 8.0 and 9.0 showed only duplex melting and no interaction of the unmodified TFO, demonstrating the pH dependent nature of the C$^+$-GC triplet.

The same experiments were repeated in the presence of the Z-containing oligonucleotide and the melting profiles are shown in the graph on the right of the figure. This time at pH 5.0, the melting curve for the complex (red curve) is shifted much more to the right of that obtained for the duplex alone (grey curve), yielding an increase in T_m values of *ca.* 7.6°C. This is a remarkable increase in stability compared to the unmodified control oligonucleotide and demonstrates that the Z-GC triplet is again more stable than C$^+$-GC at pH 5.0. Moreover, repeating the experiments at pH 6.0, 7.0, 8.0 and 9.0 revealed a similar increase in T_m for the melting of the triplexes with only a minor drop in stability as the pH is increased. The melting profiles of the complexes in the presence of varying concentrations of the third strand was also examined, and as expected, the T_m increases as the concentration of the third strand increases (Supplementary Figure S4; Table S1).

It is well established that triplexes generated with contiguous C$^+$-GC triplets are less stable than those generated with isolated C$^+$-GC triplets on account of proton competition between adjacent cytosines (11). The same melt-

ing experiments were therefore repeated with a duplex sequence containing three adjacent GC base pairs using the TFO sequence shown in Figure 1E and the 31-mer duplex shown in Supplementary Figure S1h. The melting profiles of the complexes at various pH values are shown in Figure 3B and the T_m values calculated from these in Table 1. Experiments undertaken with the C-containing oligonucleotide are shown in the graph on the left of the figure. It is again clear that the unmodified TFO is capable of triplex formation at pH 5.0 and 6.0, but not at pH 7.0, 8.0 and 9.0. As expected, the stability of these complexes is between 10 and 20°C lower than that achieved for the targeting of the sequence that contains the same number of GC base pairs but at isolated positions. Experiments undertaken with the Z-containing oligonucleotide are shown in the graph on the right of the figure. In contrast to C, these reveal only a minor drop in stability of the complexes compared to those generated with the targeting of isolated GC base pairs; T_m values from these complexes are about 2°C lower than those observed with the sequence containing distributed cytosines. It is clear that triplex formation by oligonucleotides containing Z are less dependent on the arrangements of GC base pairs in the target duplex sequence.

Lastly, the selectivity of Z for each base pair in turn was assessed by performing an electrophoretic mobility shift assay (EMSA) on the complexes shown in Figure 1D at pH 7.5 (Figure 4). In these experiments, the TFO containing isolated substitutions was targeted against four duplexes,
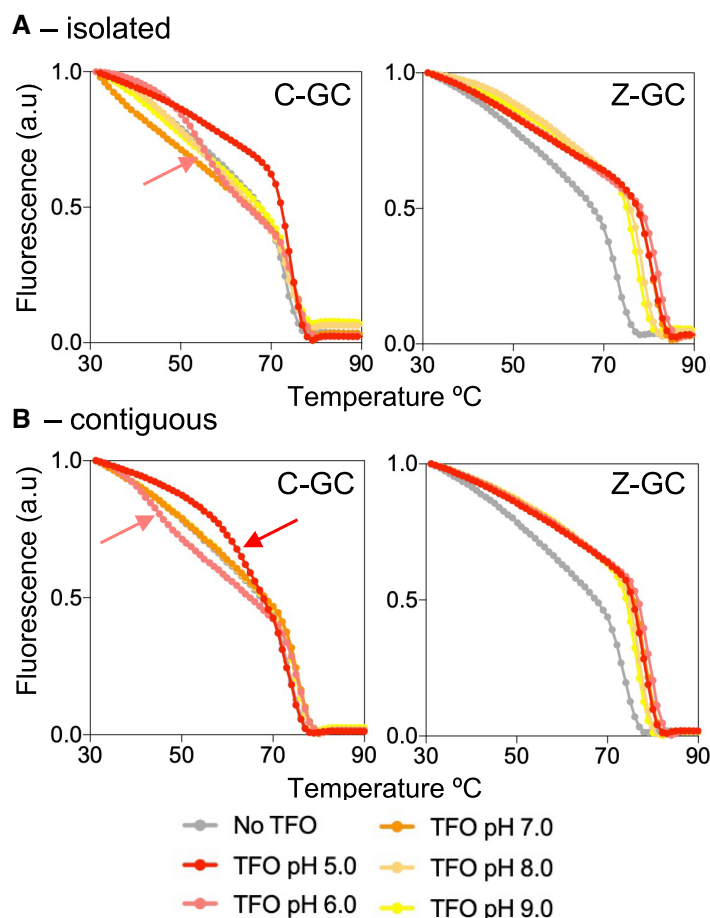
A – isolated

B – contiguous

Figure 3. pH dependence and sequence influence. Fluorescence melting profiles for the triplexes containing isolated (**A**) or contiguous (**B**) C-GC or Z-GC triplets. The 13-mer target sequence was located centrally within a 31-mer duplex. Oligonucleotides were prepared in sodium cacodylate buffer containing magnesium at the pH values indicated. The final concentrations of the duplexes and TFOs were 1 and 10 μM, respectively. The complexes were melted at a rate of 0.2°C/min in the presence of SYBR green and the fluorescence signal recorded at 522 nm after excitation at 488 nm. The arrows indicate TFO dissociation in melting profiles where the TFO dissociated before melting of the underlying duplex.

Table 1. pH dependence of synthetic TFOs. T_m values (°C) calculated at different pH values for the triplexes containing either C^+-GC or Z-GC triplets at isolated or contiguous positions. Values in parenthesis indicate TFO dissociation in melting profiles where the TFO dissociates before melting of the underlying duplex. Representative fluorescence melting profiles are shown in Figure 3

		pH 5.0	pH 6.0	pH 7.0	pH 8.0	pH 9.0
Isolated	Duplex	73.0	75.1	75.3	73.6	73.2
	C^+-GC	74.2	75.3 (54.9)	75.2	74.2	74.4
	Z-GC	80.6	81.9	80.8	78.4	77.4
Contiguous	Duplex	73.5	75.3	75.1	74.7	74.1
	C^+-GC	73.1 (63.5)	75.4 (43.8)	75.1	75.0	74.6
	Z-GC	78.1	79.2	78.5	77.0	76.5

Figure 4. Selectivity. Electrophoretic mobility shift assay for triplexes generated with C or Z positioned opposite each base pair in turn at a single position. The 13-mer target sequence was located centrally within a 31-mer duplex. Oligonucleotides were prepared in tris-acetate buffer containing magnesium at pH 7.5. The final concentration of the duplexes was 1 μM, whilst the final concentration of the TFO was varied between 0.1 and 1 μM as indicated. The complexes were separated on a 15% non-denaturing polyacrylamide gel in tris-acetate running buffer. Gels were subjected to post-staining with GelRed. Arrows indicate either double-stranded (ds) or triple-stranded (ts) complexes.

each designed to position a single Z substitution in the third strand opposite a different base pair in the duplex, generating Z-GC, Z-CG, Z-AT or Z-TA triplets. Analysis of the gel reveals that, as expected, the C-containing control oligonucleotide does not bind to any of the duplexes at pH 7.5. In contrast, the Z-containing oligonucleotide interacts with only its intended GC-containing duplex; no interaction with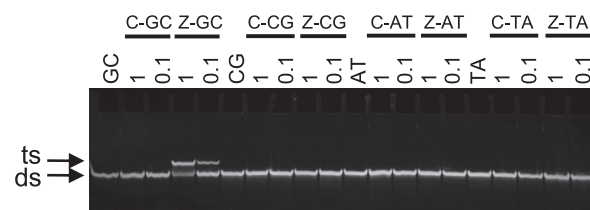 the duplexes containing the remaining base pair combinations was observed. This highlights that the increased affinity of Z for GC does not come at a cost of its selectivity, and that single triplet mismatches with Z-containing oligonucleotides are destabilising.
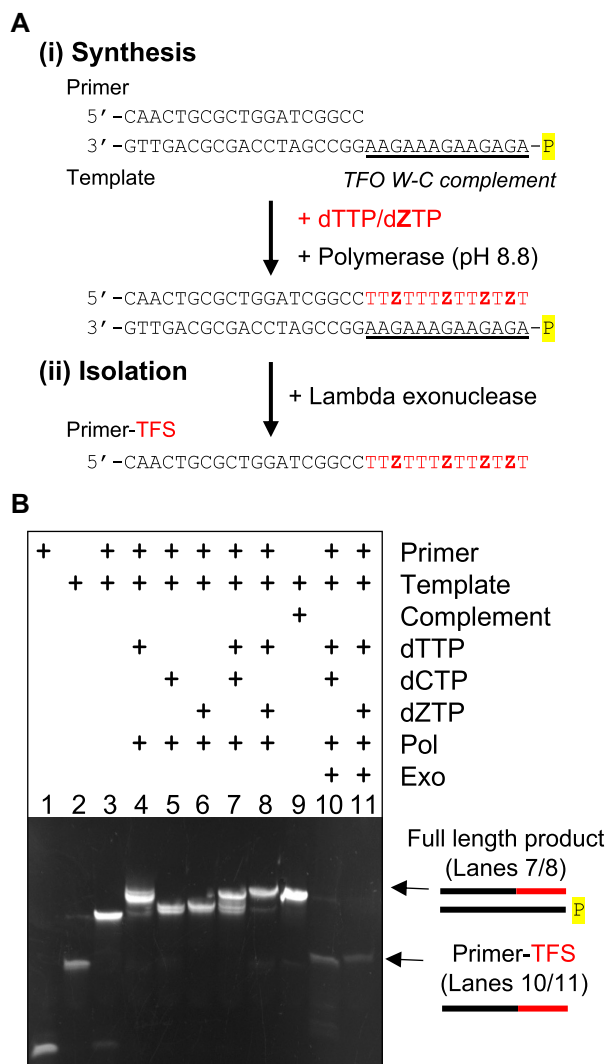
Figure 5. Template-directed assembly of Z-TFOs. (**A**) Strategy for the assembly of oligonucleotides containing a triplex-forming sequence (TFS; shown in red) containing Z. (i) Primer extension of a template that contains the W–C complement to the TFO (underlined) allows the incorporation of Z by the formation of a GZ base pair at high pH (and the absence of dCTP). (ii) The modified oligonucleotide can then be isolated by selective degradation of the phosphate-labelled template (in yellow) by the action of lambda exonuclease. (**B**) Electrophoretic mobility shift assay showing the products of the extension and digestion of such reactions. The composition of each sample is shown above each lane of the gel. Final concentration of the strands and dNTPS was 5 and 100 μM, respectively. Extension reactions were performed for 2 h at 72°C with 2 units of Therminator Pol, whilst digestion reactions were performed for 2 h with 10 units of Lambda exonuclease at 37°C. The complexes were then separated on a 15% non-denaturing polyacrylamide gel and subjected to post-staining with GelRed.

Experiments with enzymatically generated TFOs

Having established the favourable triplex-forming properties of the Z nucleobase I next investigated the feasibility of generating an oligonucleotide enzymatically with a triplex-forming sequence (TFS) containing the nucleobase (e.g. Figure 5A). The strategy exploits the previous finding that Z forms a stable base pair with guanine at pH values above the pK_a (pH 7.8) of its ring proton (25,26). This allows the incorporation of the deoxynucleotide triphosphate of Z (dZTP) opposite template G in the absence of dCTP by primer extension using a specialised polymerase (e.g. Therminator®). This was exploited to generate an oligonucleotide containing a TFS composed of Z and T by designing a template strand that contains the W–C complement to the TFS sequence at the 5′-end of the template, i.e. adjacent to the primer binding site (e.g. Figure 5A(i)). After extension the template strand was removed using Lambda exonuclease, which digests DNA with a 5′-phosphate (the primer with the Z-containing extension was not degraded as it contains a 5′-OH, e.g. Figure 5A(ii)). The result is an oligonucleotide composed of a primer tail attached to the assembled TFS. Such a strategy can be adapted to generate oligonucleotides capable of targeting different oligopurine-oligopyrimidine sequences through simple adjustment to the 5′-end of the template sequence.

Before undertaking the enzymatic studies the base pairing properties of Z with G at different pH values was analysed by comparing the interactions of C- and Z-containing oligonucleotides with a duplex complement that positions these bases opposite G (Supplementary Figure S1j). Fluorescence melting profiles of the duplexes are shown in Supplementary Figure S5 and the T_m values calculated from these in Supplementary Table S2. At pH 5.0, where Z is protonated, the T_m value for the duplex containing ZG was, as expected, >14°C less stable than the complex containing CG. But as the pH is increased, and Z becomes unprotonated, the relative difference in stability of the two duplexes decreases. At pH 9.0, the T_m value for the duplex is only 4°C less stable than the duplex containing CG and equates to a decrease in stability of around 1.5°C per Z modification. This confirmed the pH dependent nature of the interaction of Z with G and that primer extension would be possible at pH 8.8, i.e. the pH of the polymerase buffer (25,26).

Experiments were then undertaken to test this approach using a template strand that contained the W–C complement to the isolated TFO sequence shown in Figure 1D. The primer and template strand were annealed in Therminator® reaction buffer before addition of the dNTPs and the Therminator® polymerase. The samples were left for 2 h at 72°C to allow primer extension to reach completion. The products of such a reaction are shown in the gel in Figure 5B alongside appropriate control samples. Samples in lanes 1–3 show the mobilities of the primer, template and primer-template complexes in the absence of dNTPs, whilst the samples in lanes 4–8 show the extension reactions generated in the presence of various combinations of dNTPs. As expected, the addition of either dCTP or dZTP alone did not generate any observable extension products, evidenced by the samples having the same mobility as the primer-template control. Whilst, addition of dTTP alone generated some extended products, presumably due to its incorporation opposite the two As at the start of the extension sequence, and further mismatching with Gs. In contrast, combinations of either dCTP and dTTP, or dZTP and dTTP (lanes 7 and 8) show full length products and the mobility of these complexes through the gel is identical to a

control sample shown in lane 9 that contained the template annealed to its full length W–C complement, i.e. the anticipated product. The feasibility of using Lambda exonuclease to remove the phosphate-labelled template strand was then examined by incubating the extension products with the enzyme for 2 h. The samples in lanes 10 and 11 reveal that the enzyme is capable of acting on duplexes containing ZG mismatched base pairs and that the reaction goes to completion.

Since the aim of this strategy is to use the enzymatically-generated oligonucleotide for duplex recognition at physiological pH the interaction of the assembled oligonucleotide with a duplex containing its intended oligopurine-oligopyrimidine target sequence was characterised. The oligonucleotide was generated as a batch of 10 individual samples, pooled and concentrated by SpeedVac. To demonstrate the simplicity of this approach the sample was not subjected to further purification. The pH dependent binding of the oligonucleotide, and a C-containing control, were first determined by examining the thermal stability of the complexes by fluorescence melting studies. These experiments were undertaken with a 13-mer duplex composed of the oligopurine-oligopyrimidine sequence and the oligonucleotides are shown in Supplementary Figure S1i. The melting profiles of the complexes at various pH values are shown in Figure 6A and the T_m values calculated from these in Table 2. In the absence of TFO the duplex melts with a single transition yielding T_m values of between 50 and 53°C depending on the pH tested. In the presence of the C-containing TFO the melting curve for the complex is shifted to the right of that obtained for the duplex alone, but only for the experiment undertaken at pH 5.0, yielding an increase in T_m value of *ca.* 3°C. As expected, no shift is observed for the complexes generated at pH 7.0 or 9.0 due to a lack of cytosine protonation in the third strand. In contrast, in the presence of the enzymatically generated Z-containing TFO the melting curves for the complexes generated at pH 5.0, 7.0 and 9.0 are all shifted to the right, resulting in an increase in T_m of between 10 and 15°C. This is almost identical to T_m data obtained for the interaction of the short synthetic Z-containing TFO with the same duplex and confirms the correct assembly of the TFS; misincorporation of the wrong nucleotides in the TFO during primer extension would have decreased triplex stability (Supplementary Table S3). This demonstrates that, for the first time, an enzymatically generated TFO containing an unnatural nucleobase is capable of duplex recognition at neutral pH.

It is possible that the presence of the primer sequence attached to the TFO prevents binding of the oligonucleotide to an oligopurine-oligopyrimidine sequence embedded within a longer duplex. This was investigated by performing an EMSA using the same 31-mer duplex used in the earlier studies and the gel is shown in Figure 6B. Analysis of the gel reveals that, as expected, the C-containing control oligonucleotide did not interact with duplexes at pH 7.5, even at a concentration of 10 uM. Whilst the Z-containing TFO binds in a concentration dependent fashion with a band-shift evident that persists into the nanomolar range. It is clear that the presence of the primer-tail does not prevent the association of the TFS with its intended duplex target sequence.

Lastly, to demonstrate the universality of this approach two further oligonucleotides were prepared. The first oligonucleotide contained a TFS composed of contiguous Z modifications to allow the targeting of a duplex sequence containing adjacent GC base pairs (Supplementary Figure S1d). Whilst, the second oligonucleotide contained a TFS with two additional contiguous Z modifications at the 3′-end of the oligonucleotide to allow the targeting of a longer 15-mer duplex target sequence (Supplementary Figure S1e). This is important, as longer TFO sequences would be required for the targeting of unique sequences within the human genome (>15 bases in length). As previously reported dZTP can be incorporated opposite multiple adjacent template Gs by Therminator®️ polymerase (22). Indeed, the gels shown in Supplementary Figure S6 reveals the successful assembly and isolation of the required oligonucleotides. Moreover, the enzymatically-generated oligonucleotides were shown to bind in a concentration dependent fashion to their intended duplex sequences (Supplementary Figure S7). These studies demonstrate that primer extension with dZTP can be exploited to assemble TFOs capable of targeting different oligopurine-oligopyrimidine sequences through simple alterations to template design.

DISCUSSION

Strategies that overcome the pH dependence of triplex formation have been in development for over 30 years but with limited success (10). Here I have shown that the AEGIS nucleobase 6-amino-5-nitropyridin-2-one (Z) finally overcomes this constraint by acting as an uncharged mimic of protonated cytosine. Using multiple experimental approaches it was shown that TFOs containing the nucleobase enable stable and selective triplex formation at oligopurine-oligopyrimidine sequences containing multiple isolated or contiguous GC base pairs at pH values as high as 9.0. This is three pK_a units higher than what can be achieved using unmodified oligonucleotides containing cytosine.

A pertinent question to ask is how does Z compare to other nucleobase analogues developed to overcome the pH requirement of triplex formation (e.g. 2-aminopyridine, 6-oxo-cytosine, pseudoisocytosine, pyrazine, etc.)? Since previous reports have characterised these analogues within oligonucleotides of different length and base composition it is hard to compare differences in affinity directly. Moreover, several of these analogues have only been investigated as their 2′-*O*-methyl derivatives, a modification known to enhance triplex stability by promoting N-type sugar puckers in the third strand (11–13,17–20). Nevertheless, it has been shown that triplexes generated with these analogues are much less stable than their underlying duplex sequence at neutral pH. This is in stark contrast to what is observed in the melting experiments shown here where, under similar pH and buffer conditions, the triplexes are substantially more stable than the duplex containing their target sequences, even when the duplex is three times longer than the TFO target site. In fact, the thermal stability of these complexes might be underestimated in this study, as the stability of triplex DNA has been shown to be dependent on the stability of the underlying duplex (9). This could be examined by determining melting profiles for duplexes that are longer
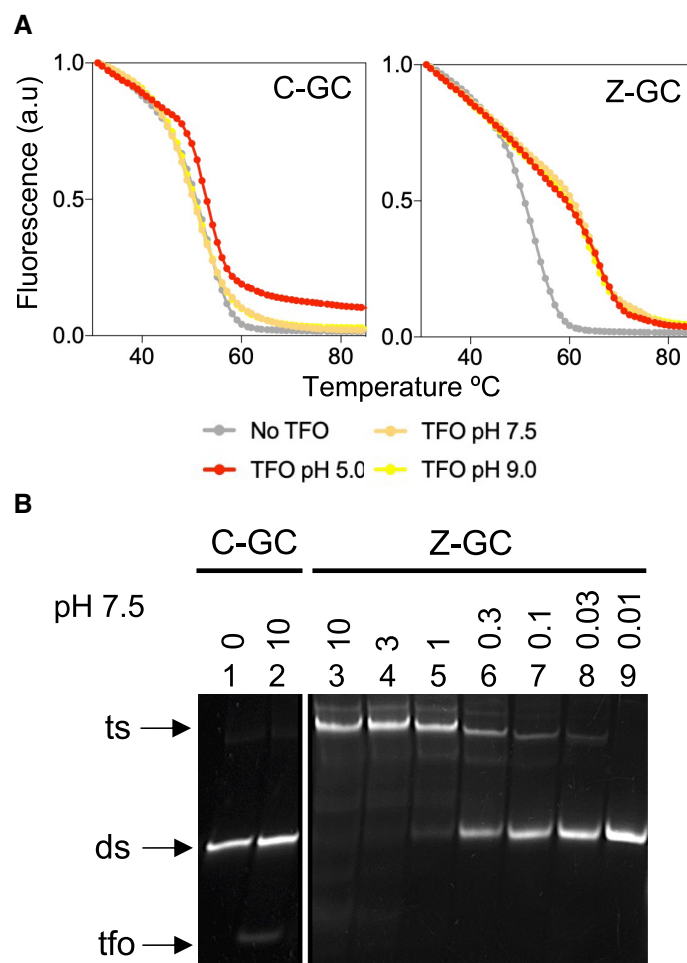
Figure 6. Triplex formation at neutral pH with an enzymatically assembled TFO containing isolated substitutions. (**A**) Fluorescence melting profiles for the triplexes containing isolated C-GC or Z-GC triplets. The target sequence was located centrally within a 13-mer duplex. Samples were prepared in sodium cacodylate buffer containing magnesium at the indicated pH. The final concentration of the duplexes and enzymatically generated TFO was 1 μM. The complexes were annealed and then melted at a rate of 0.2 °C/min in the presence of SYBR green and the fluorescence signal recorded at 522 nm after excitation at 488 nm. (**B**) Electrophoretic mobility shift assay for the triplexes containing isolated C-GC and Z-GC triplets. The target sequence was located centrally within a 31-mer duplex. Oligonucleotides were prepared in tris-acetate buffer containing magnesium at pH 7.5. The final concentration of the duplex was 1 μM and the final concentrations of the TFO were varied between 0.01 μM and 10 μM as indicated. The complexes were separated on a 15% non-denaturing polyacrylamide gel in tris-acetate running buffer. Gels were subjected to post-staining with GelRed. Arrows indicate the tfo (tfo), double-stranded (ds) or triple-stranded (ts) complexes.

than the ones used in this study. In addition, the triplets generated by the aforementioned analogues are of comparable, or lower stability, than protonated C^+-GC at low pH. This is particularly evident for analogues that lack a positive charge on their ring system (e.g. 6-oxo-cytosine, pseudoisocytosine). This again contrasts to what is observed with footprinting and melting experiments, where the uncharged Z-GC triplet is more stable than C^+-GC at pH 5.0.

The reason for this increased stability is most likely due to the electron withdrawing nitro group at the 5-position of the nucleobase, elegantly introduced by the Benner laboratory to reduce epimerization (22). Its presence might improve the strength of Hoogsten hydrogen bonding and/or lead to favourable stacking interactions with neighbouring triplets. Indeed, a crystal structure for a duplex containing multiple ZP base pairs exhibit altered stacking interactions, where P is the AEGIS nucleobase specifically designed to partner with Z (33). Intriguingly, it was shown that this

Table 2. pH dependence of enzymatically-generated TFOs. T_m values (°C) calculated at different pH values for the triplexes containing either C^+-GC or Z-GC triplets at isolated positions. Representative fluorescence melting profiles are shown in Figure 6A

	pH 5.0	pH 7.5	pH 9.0
Duplex	50.6	52.7	51.4
C^+-GC	53.4	51.1	51.1
Z-GC	66.1	64.7	64.0

promoted a more A-like than B-like duplex conformation. This would be favourable for triplexes which prefer to adopt more A-like helices to accommodate the third strand. In an attempt to examine this point further the triplexes were characterised by circular dichroism. Triplex DNA can be characterised by a negative peak at 210 nm due to the A-like nature of the triplex (34). However, there was no obvious

difference spectra obtained for the unmodified and modified triplexes (Supplementary Figure S8). Further experiments will be required to determine the structural nature of the enhanced stability of the Z-GC triplet. For example, it would be useful to compare Z-containing triplexes with those containing 1-deaza cytosine, which is a mimic of Z without the nitro group. To my knowledge only the nucleoside of the base has ever been synthesised (35).

Experiments also shown here also revealed that a Z-containing TFO was capable of targeting a unique site located within plasmid DNA at a physiologically relevant pH; TFO binding was shown to protect the underlying DNA from cleavage by a REase by binding adjacent its recognition sequence. For the future biological applications of these TFOs it will be important to demonstrate that they have activity in a cellular setting. Studies have already been undertaken using purine-rich TFOs and such systems could be used to determine whether Z-containing pyrimidine-rich TFOs offer improved efficacy, for example by comparing equivalent TFOs by transcription-inhibition assays of luciferase expression (3,4).

Another reason that the Z nucleobase is likely to be widely adopted by the TFO community comes from a synthetic standpoint. Firstly, unlike previously developed analogues the phosphoramidite of the nucleobase is available from a commercial supplier and can be incorporated into oligonucleotides by standard phosphoramidite chemistry. Secondly, oligonucleotides containing the nucleobase can be purchased directly. And lastly, and perhaps most usefully, oligonucleotides containing the nucleobase base can be assembled by primer extension using the deoxynucleotide triphosphate of the nucleobase (24,25).

The assembly approach reported here relies on a simple extension strategy using Therminator® DNA polymerase. Since the polymerase lacks 3′-5′ proof-reading activity it is possible that extension products generated by the enzyme contain misincorporated nucleotides positioned opposite template Gs instead of the intended Z nucleotides. However, the Z-G base pair is thermodynamically more stable than any other mismatched base pair and Z will be preferentially incorporated by the enzyme (26,36). Indeed, previous studies using sequences similar to the ones used in this study showed that no full length products were generated in the absence of dZTP (25). Moreover, comparison of T_m data between the synthetic and template-made TFOs in this study revealed little difference in binding of the oligonucleotides; misincorporation of the wrong nucleotide in the TFO would have substantially reduced TFO affinity.

The assembly approach used here also relied on the use of Lambda exonuclease to remove the template strand after primer extension. This generated an oligonucleotide composed of the primer and the extended triplex-forming sequence. Experiments demonstrated that the addition of the primer tail did not prevent the association of the oligonucleotide with a target sequence embedded within a longer duplex sequence. However, the presence of the tail is likely to decrease oligonucleotide affinity and experiments that examine the influence of tail length are in progress. Moreover, since the sequence of the primer can be easily adjusted, the presence of a single-stranded tail might also allow the targeting of an aptameric sequence to a precise location within synthetic or genomic DNA by triplex formation. Conversely, in some circumstances the primer-tail might not be required, and it would be more useful to isolate the TFS from the primer oligonucleotide. This could be achieved by replacing the Lambda exonuclease digestion with an appropriate nicking and release strategy that allows the TFO to dissociate from the remainder of the primer-template. Alternatively, extension of an RNA primer would allow the removal of the primer by RNase H or alkali heat treatment. Indeed it is possible to extend RNA primers generating full length Z-containing products (Supplementary Figure S9). These findings will not only improve the recognition properties of TFOs but also reduce the cost and/or expertise associated with their chemical syntheses. This is particularly encouraging at a time when the modulation of gene expression by oligonucleotides is coming to fruition.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Moser,H.E. and Dervan,P.B. (1987) Sequence-specific cleavage of double-helical DNA by triple helix formation. *Science*, **238**, 645–650.
2. Le Doan,T., Perrouault,L., Praseuth,D., Habhoub,N., Decout,J.L., Thuong,N.T., Lhomme,J. and Hélène,C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-(alpha)-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic Acids Res.*, **15**, 7749–7760.
3. Seidman,M.M. and Glazer,P.M. (2003) The potential for gene repair by triplex formation. *J. Clin. Invest.*, **112**, 487–494.
4. Rusling,D.A., Broughton-Head,V.J., Brown,T. and Fox,K.R. (2008) Towards the targeted modulation of gene expression by modified triplex-forming oligonucleotides. *Curr. Chem. Biol.*, **2**, 1–10.
5. Potaman,V.N. (2003) Applications of triple-stranded nucleic acid structures to DNA purification, detection and analysis. *Expert Rev. Mol. Diagn.*, **3**, 481–496.
6. Chandrasekaran,A.R. and Rusling,D.A. (2018) Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res.*, **46**, 1021–1037.
7. Beal,P.A. and Dervan,P.B. (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science*, **251**, 1360–1363.
8. Rahakrishnan,I. and Patel,D.J. (1993) Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure*, **1**, 135–152.
9. Rusling,D.A., Rachwal,P.A., Brown,T. and Fox,K.R. (2009) The stability of triplex DNA is affected by the stability of the underlying duplex. *Biophys. Chem.*, **145**, 105–110.
10. Fox,K.R., Brown,T. and Rusling,D.A. (2018) DNA recognition by parallel triplex formation. In: Waring,M.J. (ed). *DNA-Targeting Molecules as Therapeutic Agents, RSC Publishing*. Cambridge, pp 1–32.

11. Lee,J.S., Woodsworth,M.L., Latimer,L.J. and Morgan,A.R. (1984) Poly(pyrimidine).poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic. Acids. Res.*, **12**, 6603–6614.

12. Bates,P.J., Laughton,C.A., Jenkins,T.C., Capaldi,D.C., Roselt,P.D., Reese,C.B. and Neidle,S. (1996) Efficient triple helix formation by oligodeoxyribonucleotides containing alpha- or beta-2-amino-5-(2-deoxy-D-ribofuranosyl) pyridine residues. *Nucleic Acids Res.*, **24**, 4176–4184.

13. Hildbrand,S. and Leumann,C. (1996) Enhancing DNA triple helix stability at neutral pH by the use of oligonucleotides containing a more basic deoxycytidine analog. *Angew. Chem. Int. Ed.*, **35**, 1968–1970.

14. Rusling,D.A., Le Strat,L., Powers,V.E.C., Broughton-Head,V.J., Booth,J., Lack,O., Brown,T. and Fox,K.R. (2005) Combining nucleoside analogues to achieve recognition of oligopurine tracts by triplex-forming oligonucleotides at physiological pH. *FEBS Lett.*, **579**, 6616–6620.

15. Roberts,R.W. and Crothers,D.M. (1996) Prediction of the stability of DNA triplexes. *Proc. Natl. Acad. Sci.*, **93**, 4320–4325.

16. James,P.J., Brown,T. and Fox,K.R. (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different portions of C+-GC and T-AT triplets. *Nucleic. Acids. Res.*, **31**, 5598–5606.

17. Ono,A., Ts'o,P.O.P. and Kan,L.S. (1991) Triplex formation of oligonucleotides containing 2′-O-methylpseudoisocytidine in substitution for 2'-deoxycytidine *J. Am. Chem. Soc.*, **113**, 4032–4033.

18. Hartono,Y.D., Pabon-Martinez,Y.V., Uyar,A., Wengel,J.K., Lundin,K.E., Zain,R., Smith,C.I.E., Nilsson,L. and Villa,A. (2017) Role of pseudoisocytosine tautomerization in triplex-forming oligonucleotides: in silico and in vitro studies. *ACS Omega*, **2**, 2165–2177.

19. Xiang,G., Soussou,W. and Mclaughlin,L.W. (1994) A new pyrimidine nucleoside (m^{5ox}C) for the pH-independent recognition of G-C base pairs by oligonucleotide-directed triplex formation. *J. Am. Chem. Soc.*, **116**, 11155–11156.

20. von Krosigk,U. and Benner,S.A. (1995) pH-independent triple helix formation by an oligonucleotide containing pyrazine donor-donor-acceptor base. *J. Am. Chem. Soc.*, **117**, 5361–5362.

21. Walsh,S., El-Sagheer,A.H. and Brown,T. (2018) Fluorogenic thiazole orange TOTFO probes stabilise stabilise parallel DNA triplexes at pH 7 and above. *Chem. Sci.*, **9**, 7681–7687.

22. Hutter,D. and Benner,S.A. (2003) Expanding the genetic alphabet: non-epimerizing nucleoside with the *py*DDA hydrogen-bonding pattern. *J. Org. Chem.*, **68**, 9839–9842.

23. Yang,Z., Hutter,D., Sheng,P., Sismour,A.M. and Benner,S.A. (2006) Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res.*, **34**, 6095–6101.

24. Yang,Z., Chen,F., Alvarado,J.B. and Benner,S.A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.*, **133**, 15105–15112.

25. Yang,Z., Durante,M., Glushakova,L.G., Sharma,N., Leal,N.A., Bradley,K.M., Chen,F. and Benner,S.A. (2013) Conversion strategy using an expanded genetic alphabet to assay nucleic acids. *Anal. Chem.*, **85**, 4705–4712.

26. Reichenbach,L.F., Sobri,A.A., Zaccai,N.R., Agnew,C., Burton,N., Eperon,L.P, de Ornellas,S., Eperon,I.C., Brady,R.L. and Burley,G.A. (2016) Structural basis of the mispairing of an artificially expanded genetic information system. *Chem.*, **1**, 946–958.

27. Scaria,P.V. and Shafer,R.H. (1991) Binding of ethidium bromide to a DNA triple helix. *J. Biol. Chem.*, **266**, 5417–5423

28. Rusling,D.A., Nandhakumar,I.S., Brown,T. and Fox,K.R. (2012) Triplex-directed recognition of a DNA nanostructure assembled by crossover strand exchange. *ACS Nano*, **6**, 3604–3613.

29. Rusling,D.A., Chandrasekaran,A.R., Ohayon,Y.P., Brown,T., Fox,K.R, Sha,R., Mao,C. and Seeman,N.C. (2014) Functionalizing designer DNA crystals with a triple-helical veneer. *Angew. Chem. Int. Ed.*, **53**, 3979–3982.

30. Hampsire,A.J., Rusling,D.A., Broughton-Head,V.J. and Fox,K.R. (2007) Footprinting: a method for determining the selectivity, affinity and kinetics of DNA-binding ligands. *Methods*, **42**, 128–140.

31. Darby,R.A.J, Sollogoub,M., McKeen,C., Brown,L., Risitano,A., Brown,N., Barton,C., Brown,T. and Fox,K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecuar beacons and a LightCycler. *Nucleic Acids Res.*, **30**, e39.

32. Trantakis,I.A., Fakis,M., Tragoulias,S.S., Christopoulos,T.K., Persephonis,P., Giannetas,V. and Ioannou,P. (2010) Ultrafast fluorescence dynamics of Sybr GreenI/DNA complexes. *Chem. Phys. Lett.*, **485**, 187–190.

33. Georgiadis,M.M., Singh,I., Kellett,W.F., Hoshika,S., Benner,S.A. and Richards,N.G.J. (2015) Structural basis for a six nucleotide genetic alphabet. *J. Am. Chem. Soc.*, **137**, 6947–6955.

34. Gray,D.M., Hung,S.H. and Johnson,K.H. (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol.*, **245**, 19–34.

35. Sollogoub,M., Fox,K.R., Powers,V.E.C. and Brown,T. (2002) First synthesis of 1-deazacytidine, the C-nucleoside analogue of cytidine. *Tetts. Letts.*, **43**, 3121–3123.

36. Wang,X., Hoshika,S., Peterson,R.J., Kim,M-J., Benner,S.A. and Kahn,J.D. (2017) Biophysics of artificially expanded genetic information systems. thermodynamics of dna duplexes containing matches and mismatches involving 2-Amino-3-nitropyridin-6-one (Z) and Imidazo[1, 2-a]-1, 3, 5-triazin-4(8H)one (P). *ACS Synth. Biol.*, **6**, 782–792.