3674–3684 *Nucleic Acids Research, 2001, Vol. 29, No. 17*

Synthesis and monitored selection of nucleotide surrogates for binding T:A base pairs in homopurine–homopyrimidine DNA triple helices

Andriy A. Mokhir[1,2], William H. Connors[1,2] and Clemens Richert[1,2,*]

[1]Department of Chemistry, University of Constance, Fach M 709, D-78457 Konstanz, Germany and
[2]Department of Chemistry, Tufts University, Medford, MA 02155, USA

Received April 12, 2001; Revised June 13, 2001; Accepted July 12, 2001

ABSTRACT

A total of 16 oligodeoxyribonucleotides of general sequence 5′-TCTTCTZTCTTTCT-3′, where Z denotes an *N*-acyl-*N*-(2-hydroxyethyl)glycine residue, were prepared via solid phase synthesis. The ability of these oligonucleotides to form triplexes with the duplex 5′-AGAAGATAGAAAGA-HEG-TCTTTCTATC-TTCT-3′, where HEG is a hexaethylene glycol linker, was tested. In these triplexes, an 'interrupting' T:A base pair faces the Z residue in the third strand. Among the acyl moieties of Z tested, an anthraquinone carboxylic acid residue linked via a glycinyl group gave the most stable triplex, whose UV melting point was 8.4°C higher than that of the triplex with 5′-TCTTCTGTCTTTCT-3′ as the third strand. The results from exploratory nuclease selection experiments suggest that a combinatorial search for strands capable of recognizing mixed sequences by triple helix formation is feasible.

INTRODUCTION

Nucleic acids play pivotal roles in cells and are thus important targets for therapeutics and diagnostics (1). For single-stranded nucleic acids, most notably mRNAs, ligands can readily be designed based on the base pairing rules of Watson and Crick. For folded RNA sequences, such as the components of the ribosomal machinery, a number of small molecule binders are available, including aminoglycosides (2), but no general design principle is known. For double-stranded genomic DNA, the situation is somewhat more complex. Ligands ranging from intercalators and reactive drugs (3) to minor groove binders, duplex-invading oligonucleotide derivatives (4,5), triple helix-forming oligonucleotides and sequence-specific proteins are known, but they cannot be adapted to binding any chosen sequence selectively. For minor groove binders, rules for binding DNA duplexes sequence-specifically are emerging (6). Rules for triple helix formation have been published (7), however, successful triplex formation, presumably the most general approach for binding long stretches of double-stranded DNA sequence-specifically, usually requires a homopurine–homopyrimidine target site, severely limiting the scope of this method. A number of approaches have been developed to overcome this limitation (reviewed in 8), but binding any mixed sequence with a third strand is still an elusive goal. Given the potential of gene targeting with triplex-forming strands (9–11), further work in this area is therefore warranted.

Binding DNA sequences that contain a pyrimidine interrupting a homopurine stretch is particularly challenging when the pyrimidine is a thymidine residue. While a number of modified residues have been developed that can be placed opposite a deoxycytidine residue with little or no loss in triplex stability (12–14), thymidines in the homopurine tract were found to be more difficult to bind. This is most probably due to the methyl group of thymine that protrudes into the major groove, thus generating an obstacle in the area where the third strand would otherwise be located (Fig. 1). Guanine as the nucleobase facing the thymine gives the most stable triplexes, if unmodified oligonucleotides are employed (15,16). Few modified residues rival its modest affinity (17), except for a PNA derivative, which still binds the less sterically demanding uracil derivative more tightly than its thymine counterpart (18).

Here we report the results from an exploratory study aimed at finding new residues for oligonucleotides that are capable of binding mixed sequence duplexes. A series of oligonucleotides containing a hydroxyethylglycinamide residue was prepared and their affinity for a target strand containing one interrupting T:A base pair was measured. This led to a derivative capable of binding with higher affinity than its counterpart with a natural deoxyguanosine residue facing the thymidine in the duplex. It is shown that strands with increased affinity for duplexes containing an interrupting base pair can be selected in mass spectrometrically monitored nuclease survival experiments.

MATERIALS AND METHODS

Chemicals employed as starting materials or for building block synthesis were the best commercially available grade from Acros (Geel, Belgium), Aldrich/Fluka/Sigma (Deisenhofen, Germany) or Advanced ChemTech (Louisville, KY) and were used without purification. DNA synthesis reagents were from Proligo (Hamburg, Germany). Phosphodiesterase II

*To whom correspondence should be addressed at: Department of Chemistry, University of Constance, Fach M709, D-78457 Konstanz, Germany. Tel: +49 7531 88 45 72; Fax: +49 7531 88 45 73; Email: clemens.richert@uni-kostanz.de

Dedicated to Prof. Gerhard Schröder, Karlsruhe

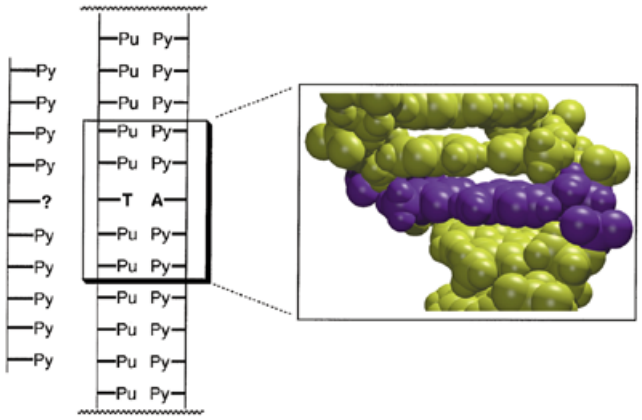


Figure 1. Forming a triplex with a target duplex containing an interrupting T:A base pair. A schematic representation of a homopurine (Pu)–homopyrimidine (Py) triplex containing the T:A base pair. The drawing on the right shows the B-form DNA duplex from the major groove with the interrupting base pair in purple. Note that the methyl group of T protrudes into the major groove.

(EC 3.1.16.1, type I-SA; from bovine spleen) was from Sigma. The 5′-monomethoxytritylamino-5′-deoxythymidine phosphoroamidite was synthesized as described (19,20). Yields of DNA hybrids were determined from the integration of the product peaks in the HPLC trace of the crude products. The integration was not corrected for the absorbance caused by the solvent front. MALDI-TOF spectra were recorded on a Bruker BIFLEX III spectrometer in negative, linear mode, unless otherwise noted. The matrix mixture for oligonucleotides was prepared from 2,4,6-trihydroxyacetophenone (THAP, 0.3 M in ethanol) and diammonium citrate (0.1 M in water) (2:1, v/v) unless otherwise noted. Calculated masses are average masses, m/z found are those for the pseudomolecular ions ($[M–H]^-$), detected as the maximum of the unresolved isotope pattern. The accuracy of mass determination with the external calibration used is approximately ±0.1%, i.e. ±2 Da at m/z 2000. HPLC was performed on a 250 × 4.6 mm Macherey–Nagel Nucleosil C4 column using gradients of CH_3CN (solvent B) in 0.1 M triethylammonium acetate buffer, pH 7. NMR spectra were recorded on Bruker AC 250 and JEOL 400 spectrometers. In the sequences given, T* denotes a 5′-amino-5′-deoxythymidine residue.

N-allyloxycarbonyl-*N*-(2-allyloxycarbonyloxyethyl)glycine *tert*-butyl ester (2)

N-(2-hydroxyethyl)glycine *tert*-butyl ester (21) (0.60 g, 3.43 mmol) was dissolved in THF (10 ml) and pyridine (0.61 ml, 7.55 mmol) was added. After stirring the solution for 10 min and cooling to 0°C, allyloxycarbonyl chloride (Alloc-Cl; 0.8 ml, 7.55 mmol) was added in portions over 30 min, while keeping the temperature below 5°C. The solution was allowed to warm to 22°C and was left stirring for 8 h. After complete conversion (TLC: silica, EtOAc/petroleum ether 55:45, R_f educt 0, R_f product 0.9, R_f mono-Alloc intermediate 0.7), the THF was evaporated, water (50 ml) was added, and the product was extracted with $CHCl_3$ (3 × 50 ml). The combined chloroform extracts were washed with saturated $NaHCO_3$ solution, HCl solution (1%), saturated NaCl solution and water, followed by drying over Na_2SO_4. The solvent was evaporated and the residue chromatographed on silica, eluting with EtOAc/petroleum ether (1:3) as an elution mixture. Yield 1.11 g (94%); ^{1}H NMR (250 MHz, $CDCl_3$) (2 rotamers) δ = 5.97–5.80 (m, 2H), 5.38–5.15 (m, 4H), 4.61–4.55 (m, 4H), 4.29 and 4.26 (pseudo-q, 4H, *J* = 4.8 Hz), 3.97 and 3.91 (2s, 2× 2H), 3.60 (2 tr, 4H, *J* = 4.5 Hz), 1.42 and 1.41 (2s, 2× 9H); FAB-MS m/z = 344 $[M+H]^+$, 288 $[M+H\text{-}H_2C{=}C(CH_3)_2]^+$.

N-allyloxycarbonyl-*N*-(2-hydroxyethyl)glycine *tert*-butyl ester (3)

Doubly Alloc-protected ester **2** (1.11 g, 3.23 mmol) was dissolved in DMF (20 ml) and the solution cooled to 0°C. Aqueous NaOH (2 N, 20 ml) was added in portions over 15 min and allowed to warm to 22°C. After overnight stirring, the solution was neutralized with 10% HCl, diluted with water (60 ml) and extracted with EtOAc (3 × 100 ml). The combined organic phases were dried over Na_2SO_4 and the solvent was evaporated *in vacuo*. The oily residue was chromatographed on silica using EtOAc/petroleum ether (1:1). Yield 0.95 g (95%); TLC (EtOAc/petroleum ether 55:45) R_f 0.7; ^{1}H NMR (250 MHz, $CDCl_3$) (2 rotamers) δ = 5.95–5.75 (m, 2H), 5.30–5.12 (m, 4H), 5.09 (br s, 2H), 4.59–4.50 (m, 4H), 4.12 (pseudo-q, 4H, *J* = 5.2 Hz), 3.90 and 3.86 (2 s, 2× 2H), 3.51 (pseudo-q, 4H, *J* = 5.4 Hz), 1.42 and 1.41 (2 s, 2× 9H).

N-allyloxycarbonyl-*N*-(2-(4,4′-dimethoxytriphenylmethoxy)ethyl)glycine methyl ester (4)

N-allyloxy-protected *tert*-butyl ester **3** (0.95 g, 3.07 mmol) was dissolved in a mixture of trifluoroacetic acid and CH_2Cl_2 (95:5, 20 ml) and left stirring for 2 h. The solvents were evaporated at reduced pressure and the residue co-evaporated with toluene (3 × 50 ml), followed by drying at 0.1 Torr overnight. Analysis of an analytical sample showed complete deprotection. ^{1}H NMR (250 MHz, DMSO-d_6) (2 rotamers) δ = 6.54 (br s, 2H), 5.99–5.82 (m, 2H), 5.33–5.12 (m, 4H), 4.52–4.43 (m, 4H), 3.98 (pseudo-q, 4H, *J* = 5.0 Hz), 3.66 and 3.59 (2s, 2× 2H), 3.30 (pseudo-q, 4H, *J* = 4.9 Hz). The remaining material was dissolved in methanol (100 ml) and treated with sulfuric acid (94%, 5 ml, 94 mmol) in portions over 15 min under vigorous stirring. After 2 h, the solution was slowly neutralized with sodium bicarbonate and filtered. The methanol was removed *in vacuo* and the remaining oily material partitioned between ethyl acetate and water. The organic phases were combined, dried over Na_2SO_4, and concentrated *in vacuo* (TLC: petroleum ether/ethyl acetate 55:45; R_f educt 0.05, R_f product 0.6). The residue was dried at 0.1 Torr for 16 h and then mixed with DMAP (0.1 g, 0.8 mmol), and 4,4′-dimethoxytriphenylmethylchloride (1.25 g, 3.7 mmol), followed by drying at 0.1 Torr for 1 h. The mixture was dissolved in pyridine (25 ml) under argon and left stirring for 24 h. Methanol (5 ml) was added, and the solution was stirred for additional 30 min. The solvents were evaporated, the oily residue co-evaporated with toluene (3 × 50 ml), and chromatographed on silica using CH_2Cl_2/petroleum ether (1:3). Yield 0.4 g (25%); TLC (CH_2Cl_2/TEA 98:2) R_f 0.6; ^{1}H NMR (250 MHz, acetone-d_6) (2 rotamers) δ = 7.50–7.18 (m, 18H), 6.89 (pseudo-d, 8H, *J* = 8.5 Hz), 5.99–5.83 (m, 2H), 5.26 (dd, 2H, *J* = 17.5 and 5.1 Hz), 5.15 (d, 2H, *J* = 10.5 Hz), 4.55 (d, 4H, *J* = 5.2 Hz), 4.19 and 4.13 (2s, 2× 2H), 3.77 and 3.65 (2 s, 6H), 3.55 (t, 4H, *J* = 5.9 Hz), 3.28 (pseudo-q, 4H, *J* = 5.5 Hz); ^{13}C NMR (101 MHz, acetone-d_6) (2 rotamers, not all signals

split) δ = 170.1, 170.0, 158.8, 155.8, 155.3, 145.4, 136.1, 133.3, 130.0, 128.0, 127.8, 126.7, 116.5, 116.1, 113.1, 86.3, 65.8, 65.6, 62.2, 62.0, 54.7, 51.4, 51.3, 49.5, 48.6, 48.0; MALDI-TOF MS (6-aza-2-thiothymine matrix, positive mode) m/z = 558 $[M+K]^+$ (18%), 542 $[M+Na]^+$ (23%).

N-allyloxycarbonyl-*N*-(2-(4,4′-dimethoxytriphenylmethoxy)ethyl)glycine triethylammonium salt (5)

DMT-protected methyl ester **4** (0.4 g, 0.77 mmol) was dissolved in methanol (30 ml) and aqueous NaOH (5 N, 15 ml) was added. The resulting slurry was stirred for 10 h, followed by neutralization with the triethylammonium form of Dowex cation exchange resin. The solution was filtered, the solvents evaporated and the residue vacuum dried (0.1 Torr) for 24 h. Yield 0.29 g (61 %). ^{1}H NMR (250 MHz, acetone-d_6) (2 rotamers) δ = 7.50–7.10 (3 m, 18H), 6.86 (dd, 8H, J = 2.5 and 8.5 Hz), 5.98–5.78 (m, 2H), 5.34–5.07 (m, 4H), 4.96 (br s, 2H), 4.52 (d, 4H, J = 5.5 Hz), 3.91 and 3.90 (2 s, 4H), 3.75 and 3.74 (2 s, 2× 3H), 3.50 (m, 4H), 3.20 (m, 4H), 2.90 (q, 12 H, *J* = 6.2 Hz), 1.18 (t, 18H, *J* = 6.2 Hz); ^{13}C NMR (62.5 MHz, acetone-d_6) (2 rotamers, not all signals split) δ = 175.0, 174.5, 159.4, 157.1, 156.8, 146.3, 146.2, 137.1, 137.0, 134.7, 134.5, 130.8, 128.84, 128.80, 128.5, 127.3, 117.0, 116.5, 113.8, 86.7, 66.3, 65.8, 62.8, 62.6, 55.4, 52.2, 49.3, 48.6, 45.8, 9.6; MALDI-TOF MS (6-aza-2-thiothymine matrix, positive mode) m/z = 303 $[DMT]^+$ (100%).

Synthesis of oligonucleotides 12a–p

Assembly of DNA portion of oligonucleotides. The 3′-DNA portion of the oligodeoxyribonucleotides was synthesized on an ABI 381 DNA synthesizer according to the manufacturer's recommendations for 1.0 µM scale syntheses. For the preparation of the CPG-bound form of the 5′-amino-5′-deoxyoligonucleotide **9**, the phosphoroamidite of 5′-monomethoxytritylamino-5′-deoxythymidine (19,20) was employed, using standard coupling conditions.

*Compound **10** (5′-DMT-OCH$_2$CH$_2$N(Alloc)CH$_2$CO-T*C^{Bz}TT-TC^{Bz}T-CPG).* A sample of the triethylammonium salt of linker building block **5** (150 mg, 242 µmol), HBTU (84 mg, 218 µmol) and HOBT (33 mg, 242 µmol) were dried at 0.1 Torr for 1 h and then dissolved in DMF (0.4 ml), followed by addition of DIEA (93 µl, 533 µmol). The coupling mixture was added to a polypropylene reaction vessel containing T*C^{Bz}TTTC^{Bz}T-CPG (**9**; 90 mg, 3 µmol loading of the 3′-terminal nucleoside), and the mixture vortexed vigorously. The reaction was allowed to proceed for 2 h with vortexing every 5 min. The supernatant was carefully aspired with a syringe bearing a small gauge needle, the residue washed with DMF (2 × 1 ml) and methanol (2 × 1 ml) and then dried at 0.1 Torr for 4 h. An analytical sample of the CPG (10–15 beads) was treated with 27% aqueous ammonia solution at room temperature for 8 h. After evaporation of the ammonia, the solution was analyzed by MALDI-TOF MS. The coupling efficiency was found to be >70 % by integration of the MALDI peaks of product and uncoupled DNA heptamer.

*(5′-DMT-OCH$_2$CH$_2$NHCH$_2$CO-T*C^{Bz}TTTC^{Bz}T-CPG).* A solution of Pd(PPh$_3$)$_4$ (2.5 mg, 2.1 µmol) and PPh$_3$ (0.5 mg, 1.9 µmol) in CH$_2$Cl$_2$ (0.2 ml) was purged with argon and added to a slurry of diethylammonium bicarbonate (2.5 mg, 15 µmol) in CH$_2$Cl$_2$ (0.2 ml) that had been saturated with argon. After vortexing, a burgundy-colored, clear solution was obtained that was added to the CPG-bearing protected oligomer **10** (DMT-OCH$_2$CH$_2$N(Alloc)CH$_2$-CO-T*C^{Bz}TTTC^{Bz}T-CPG; 10 mg) in a polypropylene reaction vessel. The reaction vessel was covered with aluminum foil, and the mixture was vortexed every 5 min. After 1 h, the supernatant was aspired, and the CPG was thoroughly washed with DMF (1.5 ml), a solution of triethylamine in DMF (1%, 1.5 ml), a solution of ethyl dithiocarbamate in DMF (0.5%, 1.5 ml), DMF (2 × 1.5 ml) and methanol (2 × 1.5 ml), followed by drying the CPG at 0.1 Torr. An analytical sample, deprotected with 27% aqueous ammonia solution at room temperature for 8 h and analyzed by MALDI, showed close to 100% removal of the Alloc group.

*(5′-DMT-OCH$_2$CH$_2$N(CO-CH$_2$-NH-Fmoc)CH$_2$CO-T*C^{Bz}TT-TC^{Bz}T-CPG).* A sample of the secondary amine-bearing CPG (90 mg), prepared as described in the preceding paragraph, and a mixture of Fmoc-Gly-OH (90 mg, 300 µmol), HBTU (102 mg, 270 µmol) and HOBT (42 mg, 300 µmol) were dried separately under vacuum (0.1 Torr) for 30 min. The mixture of the acid, HBTU and HOBT was dissolved in DMF (1.5 ml) and DIEA (120 µl, 660 µmol) was added. After vortexing, the slightly darkened solution was added to the reaction vessel containing the CPG. The reaction mixture was vortexed every 5 min. After 1 h, the supernatant was aspired, the residue was washed with DMF (2 × 1 ml) and methanol (2 × 1 ml) and dried at 0.1 Torr. An analytical sample of the CPG was treated with 27% aqueous ammonia solution at room temperature for 8 h and, after evaporating the ammonia, the solution was analyzed by MALDI-TOF MS to ensure successful coupling.

Removal of Fmoc groups. The Fmoc-bearing, solid phase-bound oligonucleotide (*5′-DMT-OCH$_2$CH$_2$N(CO-CH$_2$-NH-Fmoc)CH$_2$CO-T*C^{Bz}TTTC^{Bz}T-CPG*) (90 mg) was treated with a solution of piperidine in DMF (1.5 ml, 1:4 v/v) for 30 min. At the end of the deprotection reaction, the solution was aspired, the CPG was washed with DMF (2 × 1 ml) and methanol (2 × 1 ml) and dried at 0.1 Torr.

Coupling of carboxylic acid building blocks. The glycine-bearing, solid phase-bound oligonucleotide (*5′-DMT-OCH$_2$CH$_2$N(CO-CH$_2$-NH$_2$)CH$_2$CO-T*C^{Bz}TTTC^{Bz}T-CPG* or its diglycine derivative; 1 mg) was coupled to a mixture of the carboxylic acid (25 mg, 100 µmol), HBTU (34 mg, 90 µmol), HOBT (14 mg, 100 µmol) and DIEA (40 µl, 220 µmol) in DMF (0.5 ml) as described above for the preparation of the intermediate with one glycine residue. For the thiadiazole derivative producing **12n**, an analytical sample was deprotected and analyzed. MALDI-TOF MS for $C_{103}H_{125}SN_{22}O_{52}P_6$ $[M–H]^-$ calculated 2726.0, found 2726.2.

Acetylation of amino groups with acetic anhydride. Acylation of free amino groups to give derivatives **12a–c** was performed under the conditions employed for capping during DNA synthesis, i.e. by exposing the CPG to a mixture of the solutions 'Cap A' (acetic anhydride/2,6-lutidine/THF 1:1:8) and 'Cap B' (DMAP/THF 6.5% w/v).

Assembly of the 5′-DNA portion of the oligonucleotide and deprotection. Samples of the CPG bearing the acylated acyclic

residue (1 mg each) were transferred to polypropylene reaction columns for ABI DNA synthesizers (100 mg capacity), dried at 0.1 Torr and employed for DNA synthesis using a standard protocol for 0.2 μM synthesis, except for the first two elongation steps, for which coupling times of 15 min were used. After assembly of the 5′-portion of the oligonucleotides, the terminal DMT group was removed on the synthesizer. Trityl release indicated a total yield of 92% over the six extension cycles. The samples were deprotected with 27% aqueous ammonia solution at room temperature overnight, followed by removal of excess ammonia with a stream of air directed onto the solution, lyophilization and HPLC purification.

Compound **12a**: yield 42%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 31 min. MALDI-TOF MS for $C_{132}H_{175}N_{32}O_{87}P_{12}$ $[M–H]^-$ calculated 3971.7, found 3973.2. Compound **12b**: yield 51%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 31 min. MALDI-TOF MS for $C_{134}H_{178}N_{33}O_{88}P_{12}$ $[M–H]^-$ calculated 4028.7, found 4029.3. Compound **12c**: yield 12%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 31 min. MALDI-TOF MS for $C_{136}H_{181}N_{34}O_{89}P_{12}$ $[M–H]^-$ calculated 4085.8, found 4083.8. Compound **12d**: yield 42%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 34 min. MALDI-TOF MS for $C_{142}H_{181}N_{34}O_{88}P_{12}$ $[M–H]^-$ calculated 4141.8, found 4143.8. Compound **12e**: yield 47%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 33 min. MALDI-TOF MS for $C_{144}H_{184}N_{35}O_{89}P_{12}$ $[M–H]^-$ calculated 4198.8, found 4199.7. Compound **12f**: yield 20%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 32 min. MALDI-TOF MS for $C_{140}H_{180}N_{35}O_{88}P_{12}$ $[M–H]^-$ calculated 4130.8, found 4131.3. Compound **12g**: yield 31%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 32 min. MALDI-TOF MS for $C_{142}H_{183}N_{36}O_{89}P_{12}$ $[M–H]^-$ calculated 4187.8, found 4186.6. Compound **12h**: yield 32%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 34 min. MALDI-TOF MS for $C_{144}H_{186}N_{35}O_{89}P_{12}$ $[M–H]^-$ calculated 4200.8, found 4199.2. Compound **12i**: yield 14%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 35 min. MALDI-TOF MS for $C_{146}H_{189}N_{36}O_{90}P_{12}$ $[M–H]^-$ calculated 4257.8, found 4259.4. Compound **12j**: yield 18%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95 % B, Rt = 34 min. MALDI-TOF MS for $C_{143}H_{185}N_{34}O_{88}P_{12}$ $[M–H]^-$ calculated 4157.8, found 4160.6. Compound **12k**: yield 51%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 34 min. MALDI-TOF MS for $C_{145}H_{188}N_{35}O_{89}P_{12}$ $[M–H]^-$ calculated 4214.8, found 4214.3. Compound **12l**: yield 31%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 35 min. MALDI-TOF MS for $C_{147}H_{182}N_{33}O_{90}P_{12}$ $[M–H]^-$ calculated 4220.8, found 4221.1. Compound **12m**: yield 35%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 35 min. MALDI-TOF MS for $C_{149}H_{185}N_{34}O_{91}P_{12}$ $[M–H]^-$ calculated 4277.8, found 4278.0. Compound **12n**: yield 29%; HPLC 0% B for 5 min, in 30 min to 20% B, in 10 min to 95% B, Rt = 30 min. MALDI-TOF MS for $C_{136}H_{179}N_{36}SO_{88}P_{12}$ $[M–H]^-$ calculated 4213.9, found 4214.4. Compound **12o**: yield 11%; HPLC 0% B for 5 min, in 30 min to 30% B, in 10 min to 95% B, Rt = 33 min. MALDI-TOF MS for $C_{152}H_{182}N_{33}O_{91}Br_4P_{12}$ $[M–H]^-$ calculated 4612.4, found 4613.3. Compound **12p**: yield 31%; HPLC 0% B for 5 min, in 30 min to 30% B, in 10 min to 95% B, Rt = 33 min. MALDI-TOF MS for $C_{154}H_{185}N_{34}O_{92}Br_4P_{12}$ $[M–H]^-$ calculated 4669.4, found 4669.1. Compound **13** (AGAAGATAGAAAAGA-HEG-TCTTTCTATCTTCT): yield 7%; HPLC 0% B for 5 min, in 30 min to 30% B, Rt = 25 min. MALDI-TOF MS for $C_{288}H_{371}N_{102}O_{175}P_{28}$ $[M–H]^-$ calculated 8929.0, found 8930.3. Compound **16** (AGAAGAAAGAAAGA-HEG-TCTTTCTTTCTTCT): yield 25%; HPLC 0% B for 5 min, in 30 min to 20% B, Rt = 29 min. MALDI-TOF MS for $C_{288}H_{371}N_{102}O_{175}P_{28}$ $[M–H]^-$ calculated 8929.0, found 8927.4. Compound **17** (AGAAGACAGAAAGA-HEG-TCTTTCT-GTCTTCT): yield 31%; HPLC 0% B for 5 min, in 30 min to 20% B, Rt = 29 min. MALDI-TOF MS for $C_{287}H_{370}N_{103}O_{175}P_{28}$ $[M–H]^-$ calculated 8929.9, found 8933.3. Compound **18** (AGAAGAGAGAAAGA-HEG-TCTTTCTCTCTTCT): yield 15%; HPLC 0% B for 5 min, in 30 min to 20% B, Rt = 29 min. MALDI-TOF MS for $C_{287}H_{370}N_{103}O_{175}P_{28}$ $[M–H]^-$ calculated 8929.9, found 8936.6.

UV melting experiments

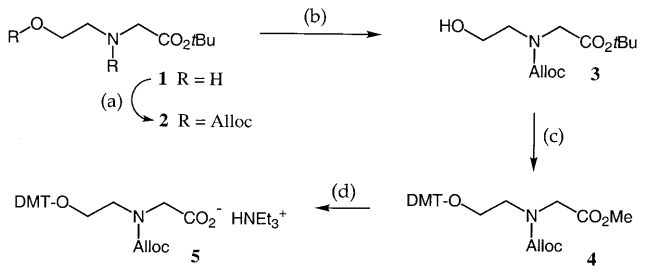
Melting curves were recorded on a Perkin-Elmer Lambda 10 spectrometer at 260 nm in semimicrocuvettes with a sample volume of 0.5 ml at heating or cooling rates of 1°C/min with 1.5 μM strand concentration in 10 mM phosphate buffer pH 6 and 1 M NaCl, except for the mismatch discrimination experiments where 0.3 μM strand concentrations were used. Annealing involved heating to 90°C followed by cooling to 5°C at a rate of 2°C/min. Melting points are averages of the extrema of the first derivative of the 91-point smoothed curves from two heating and two cooling curves.

Nuclease selections

A solution (2.0 μl) containing 5′-AGAAGATAGAAAGA-HEG-TCTTTCTATCTTCT-3′ (13), **12a**, **12b**, **12d**, **12f**, **12h**, **12j** and **12o** (100 pmol/strand) in an aqueous buffer of $(NH_4)_2SO_4$ (250 mM) pH 6 was heated to 95°C for 5 min and then cooled to 22°C at a rate of 2°C/min. A solution (1 μl) of bovine spleen phosphodiesterase (EC 3.1.16.1, phosphodiesterase II, Type I-SA, 4×10^{-4} U/μl) was added, followed by rapid mixing. Immediately afterwards, a sample of the solution (0.5 μl) was drawn for the control data point at $t = 0$ min. The reaction was allowed to proceed with sampling (0.5 μl) at stated intervals. Samples drawn were thoroughly mixed with a premixed MALDI matrix solution of diammonium citrate (1 μl, 0.1 M in water) and 6-aza-2-thiothymine (2 μl, saturated solution in ethanol). The matrix solution also contained a solution of DNA heptadecamer $(T)_{17}$ (0.7 μl, 17 pmol/μl) as an internal standard. Of the resulting mixture, a fraction (1.5 μl) was spotted on a stainless steel MALDI target and vacuum dried. MALDI spectra of the resulting semicrystalline matrix deposit were acquired in negative, linear mode at 20 kV total extraction voltage with delayed extraction. Four or five spectra from 200–300 laser shots of 50–60 μJ/shot were acquired per sample. Spectra in which the internal standard yielded less then 1000 ion counts were discarded. Data analysis was performed with the computer routines DAS and Automaton (22).

RESULTS

The residue to face the interrupting T:A base pair was designed to be flexible and suitable for combinatorial syntheses that employ the large pool of commercially available carboxylic acids. Among the modified residues reported thus far for



Scheme 1. (a) Alloc-Cl, pyridine, THF; (b) NaOH, H_2O, DMF; (c) (i) TFA, CH_2Cl_2, (ii) MeOH, H_2SO_4, (iii) DMT-Cl, pyridine, DMAP; (d) (i) NaOH, H_2O, MeOH, (ii) Dowex $HNEt_3^+$.

binding to interrupting purines, the least destabilizing abasic residue is propanediol phosphate (23), with the same number of atoms in the backbone as the ribose phosphates of natural residues. A residue with an identical number of atoms in the backbone chain was used in the present study. To reduce electrostatic repulsion with phosphates of the target strands, one of the two linkages to the remainder of the sequence was an amide. Using a phosphodiester at the other terminus of the linker and the amino group at a position found in PNA residues gave a hydroxyethylglycine backbone. A glycine residue was appended to the internal amine and an acyl residue was linked to this to interact with the T:A base pair. The acyclic backbone of the linker was believed to more easily avoid a steric clash with the methyl group of the thymidine residue in the target duplex than bulkier and more rigid cyclic linkers. It was known that a PNA derivative containing a nucleobase analog on a four atom linker to the backbone allows formation of thymine-containing triple helices (18).

Besides residues with a single glycine moiety linking the aromatic acyl residue and the hydroxyethylglycine backbone, side chains with two glycine residues between the linker and the acyl moiety were prepared. This was expected to allow for triplexes, in which the appended aromatic molecule reaches across the major groove and interacts with the adenine of the interrupting base pair. Glycine-linked aromatic moieties have previously been proposed as ligands for interrupting purines (24).

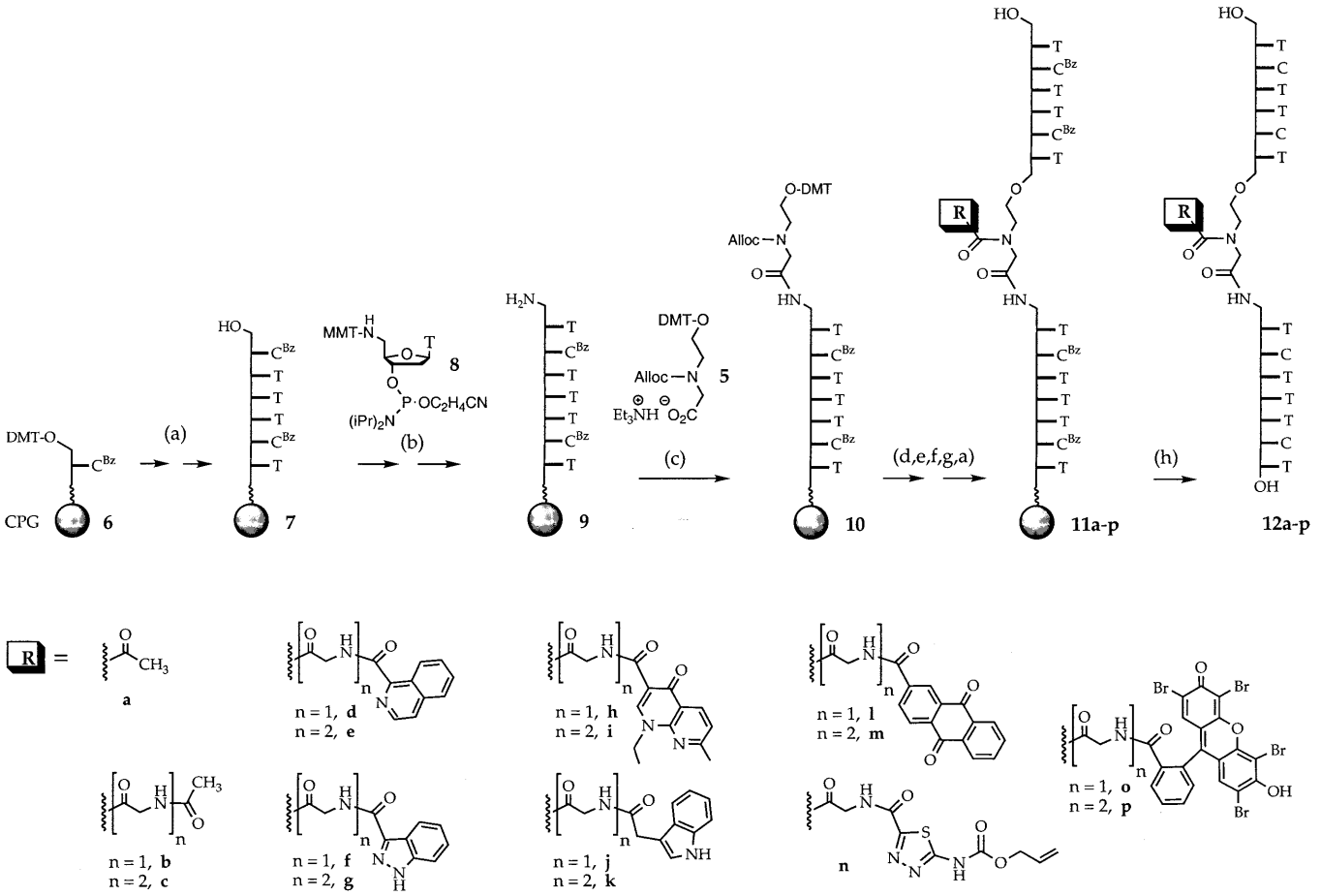
Elaboration of the oligonucleotides to be employed as third strands started from hydroxyethylglycine *t*-butyl ester (**1**) (21), whose doubly allyloxycarbonyl protected derivative **2** was hydrolyzed to **3** (Scheme 1). The *t*-butyl ester (**3**) was converted to a methyl ester, and the free hydroxyl group was DMT-protected to give **4**. With both the hydroxyl group and the amine suitably protected, the methyl ester was hydrolyzed to give the triethylammonium salt of acid **5**. An alternative, shorter synthesis produced Alloc-OCH_2CH_2N(Fmoc)CH_2CO_2H as a building block. However, this compound proved unsatisfactory, since Alloc deprotection could not be readily induced once this linker was coupled to the 3′-terminus of the DNA (results not shown). Control deprotection reactions in solution proceeded uneventfully with this alternative linker.

With building block **5** in hand, oligonucleotides with the acyclic residue were elaborated as shown in Scheme 2. Starting from controlled pore glass with a 3′-terminated deoxycytidine residue (**6**), phosphoramidite synthesis provided **7**, whose elongation with **8** (19,20) produced N-terminted **9**. Coupling of **5** to **9** under peptide coupling conditions furnished **10**. Two approaches to assembling full-length oligonucleotide **11** were tested. The first involved immediate continuation of DNA synthesis, followed by Alloc deprotection and coupling of carboxylic acids to the secondary amine of the linker. The second approach involved immediate Alloc deprotection and elaboration of the acyl side chain, followed by DNA synthesis. The first of the two approaches proved unsuccessful, largely because Alloc deprotection could not be induced in high yields, once the 5′-portion of the DNA had been appended. Probably, the 'brush' of 5′-terminal DNA strands protruding from above the linker prevented the palladium catalyst access to the linker in the solid support bound strands. When the deprotection of the secondary amine was performed first, smooth conversions were observed for both Alloc deprotection and the subsequent coupling/Fmoc deprotection reactions generating the side chains of the linker. DNA syntheses on the side chain bearing hybrid gave fully assembled **11a–p** in high yield, as monitored by trityl release and MALDI analysis of crude products. Fully deprotected oligonucleotides **12a–p** were obtained by treatment with ammonium hydroxide at room temperature, followed by single pass HPLC purification.

Among the carboxylic acids employed in the syntheses of **12a–p** was acetic acid, (compounds **12a–c**) whose terminal acetyl group was not expected to bind tightly to the DNA. The seven other acids employed contained at least one hydrogen bonding acceptor site to allow for hydrogen bonding to NH6 of the adenosine residue, and aromatic moieties capable of stacking with the neighboring nucleobases in the third strand. Six of them were commercially available, including eosin (derivatives **12o** and **12p**) that was included as a residue that cannot intercalate. The thiadiazole building block employed for preparing **12n** was synthesized in two steps from ethyl 5-amino-1,3,4-thiadiazol-2-yl-acetate (25). All carboxylic acid building blocks gave high coupling yields with the activation mixture chosen (HBTU, HOBT and Hünig's base).

To test the ability of the modified strands to engage in triple helix formation with a homopurine duplex containing one interrupting T:A base pair, the intramolecular target duplex 5′-AGAAGATAGAAAGA-HEG-TCTTTCTATCTTCT-3′ (**13**) was prepared. In this sequence 'HEG' denotes a hexaethyleneglycol linker, or '18-spacer', that links the two strands of the duplex (26,27). Covalent linking was expected to shift the melting point of the duplex to a temperature high enough for both triplex–duplex and duplex–single strand transitions to be observed separately in UV melting experiments. This was indeed found to be the case for all triplexes studied.

The melting points of the triplexes formed between **12a–p** and **13** are compiled in Table 1. The anthraquinone-bearing sequence **12l** gave the highest melting point (47.8°C), and eosin-bearing **12o** gave the lowest melting point (<15°C). While for the anthraquinone, the greatest triplex stabilization was observed with a single glycine residue, the nalidixic acid residue and the isochinoline carboxylic acid residues were more stabilizing with two glycine residues (**12i** and **12e**, respectively). All derivatives with terminal aromatic acid residues gave more stable triplexes than **12a** and **12b**, with the exception of eosin derivatives **12o** and **12p** and diglycinyl-linked indole **12k**. The wide range of melting points observed and the low number of compounds destabilizing the triplexes suggested that (i) the appended residues were interacting with



Scheme 2. (a) DNA synthesis via phosphoramidite protocol; (b) coupling cycle with the phosphoramidite building block of 5′-amino-5′-deoxythymidine (**8**); (c) **5**, HOBT, HBTU, DIEA, DMF; (d) $[Pd(PPh_3)_4]$, PPh_3, $[H_2NEt_2]^+[HCO_3]^-$, CH_2Cl_2; (e) Fmoc-Gly-OH, HOBT, HBTU, DIEA, DMF; (f) piperidine, DMF; (e) and (f) repeated for derivatives **11c**, **e**, **g**, **i**, **k**, **m**, **p**; (g) **R**-CO_2H, HBTU, HOBT, DIEA, DMF; (h) NH_4OH.

the duplex and (ii) that few of the appended residues disfavored binding, e.g. by inducing aggregation or intramolecular folding of the third strand.

The extent to which anthraquinone derivative **12l** was stabilizing the triplex became more apparent when a control triplex was studied in which a deoxyguanosine residue was facing the interrupting base pair. The melting point of this triplex (**14**:**13**, Table 1) was 39.4°C, i.e. 8.4°C lower than that of **12l**:**13**. As expected (15,16), a thymidine residue facing the T:A base pair (triplex **15**:**13**) gave a considerably less stable triplex, confirming earlier results that the G:T:A triplet is the most stable among those formed with unmodified DNA and an interrupting T:A base pair. Only a triplex without any interrupting base pair (**16**:**15**) gave a higher melting point than that of **12l** and **13** (last entry, Table 1). Representative melting curves are shown in Figure 2.

Next, the selectivity of the anthraquinone-glycine-bearing residue for interrupting T:A base pairs was studied. For this, all four possible base pairs were presented to the acyclic residue via intramolecular duplexes **13**, **16**, **17** and **18** (Table 2). The melting points showed that while the modified third strand did differentiate between pyrimidine:purine base pairs and purine:pyrimidine base pairs, with a higher affinity for the former, the melting points of the triplexes in which the **12l** was facing interrupting T:A and C:G base pairs was virtually the same.

In order to test the feasibility of combinatorial searches for more selective derivatives of the acyclic residue, exploratory nuclease survival selection experiments were performed. These were monitored via MALDI-TOF mass spectrometry (SMOSE) (28) under conditions known to allow quantitative detection of oligonucleotides (29). The principle underlying the nuclease selections is shown schematically in Figure 3. A library of modified oligonucleotides is mixed with one equivalent of target duplex, leading to a hybridization equilibrium in which the oligonucleotide with the higher affinity for the target is in triplex form to a greater extent. Treating the hybridization mixture with a single strand-specific nuclease leads to a more rapid degradation of the more exposed, lower affinity strand, and a selective survival of the higher affinity strand. Nuclease survival experiments have previously been performed to identify oligonucleotides with 5′-terminal modifications with high affinity for complementary single strands (forming duplexes) (22,30) but not with duplex targets and/or internally modified oligonucleotides.

First, a control experiment was performed, in which a mixture of the unmodified DNA strands 5′-TCTTCTT-TCTTTCT-3′ (**15**) and 5′-TCTTCTGTCTTTCT-3′ (**14**) was

Table 1. UV melting temperatures of triplexes between intramolecular duplexes and a third strand at 1.5 μM strand concentration, 1 M NaCl and 10 mM phosphate buffer pH 6

Duplex	Third strand	Melting point ± SD (°C)[a]	ΔT_m (°C)[b]
(**13**)[c]	**12a**	24.9 ± 2.0	–14.5
	12b	23.7 ± 1.8	–15.7
	12d	25.4 ± 1.1	–14.0
	12e	31.5 ± 1.0	–7.9
	12f	26.3 ± 1.8	–13.1
	12h	33.4 ± 1.5	–6.0
	12i	36.4 ± 0.6	–3.0
	12j	27.9 ± 0.8	–11.5
	12k	22.2 ± 1.5	–17.2
	12l	47.8 ± 2.0	+8.4
	12m	38.4 ± 1.2	–1.0
	12n	30.0 ± 0.6	–9.4
	12o	<15	> –24.4
	12p	13.9 ± 0.9	–25.5
	TCTTCT**G**TCTTTCT (**14**)[c]	39.4 ± 0.9	–
	TCTTCT**T**TCTTTCT (**15**)[c]	27.1 ± 0.8	–12.3
(**16**)[c]	TCTTCT**T**TCTTTCT (**15**)[c]	52.8[d]	+13.4

[a]Standard deviation.
[b]Melting point difference to triplex **13**:**14**.
[c](**13**), AGAAGA**T**AGAAAGA-HEG-TCTTTCT**A**TCTTCT; (**16**), AGAA-GAA**A**GAAAGA-HEG-TCTTTCT**T**TCTTCT. The nucleobases constituting or facing the inverted purine are in bold, solely to facilitate recognition within the sequence.
[d]Obtained from a single melting curve, no SD.

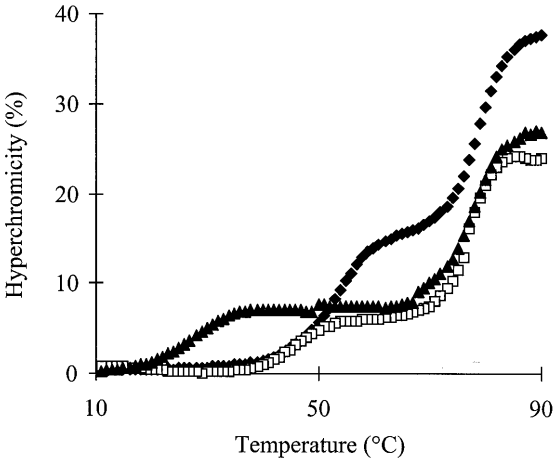


Figure 2. UV melting curves of triplexes between intramolecular duplex 5′-AGAA-GATAGAAAGA-HEG-TCTTTCTATCTTCT-3′ (**13**) and **12j** (triangles) or **12l** (squares) and between 5′-AGAAGAAAGAAAGA-HEG-TCTTTCTTTCTTCT-3′ (**16**) and 5′-TCTTCTTTCTTTCT-3′ (**15**) (diamonds) at 1.5 μM strand concentration, 1 M NaCl and 10 mM phosphate buffer, pH 6.

mixed with one equivalent of 5′-AGAAGAAAGAAAGA-HEG-TCTTTCTTTCTTCT-3′ (**16**), annealed and treated with bovine spleen phosphodiesterase (Type I-SA, EC 3.1.16.1). Monitoring the degradation with MALDI-TOF mass spectrometry revealed a delayed degradation of homopyrimidine strand **15**, whereas the relative signal intensity for **14**, with its G:A:T mismatch, decreased rapidly (Fig. 4A). When a library of modified strands **12a**, **12b**, **12d**, **12f**, **12h**, **12j** and **12o** was subjected to nuclease attack in the presence of duplex **13**, compound **12h** survived the longest (Fig. 4B). This was expected for an affinity-based selection, since **12h** has the highest triplex melting point among the compounds tested (Table 1). The selectivity with which the nuclease selection identified the highest affinity triplex-forming oligonucleotide could also clearly be seen in the degradation kinetics (Fig. 5).

The protection from nuclease attack for the individual compounds was quantified via 'protection factors'. As detailed in an earlier publication (22), these are obtained by integrating the area under the point-to-point fits of the degradation kinetics and calculating the ratio of integrals from the compound of interest and the control compound (in this case acetylated strand **12a**), thus allowing for a quantitative comparison of compounds employed in different selections. Protection factors for this and a subsequent selection involving strands **12a**, **12c**, **12g**, **12e**, **12k**, **12i** and **12m** are compiled in Table 3. A selection with **12l** and **12a**, **b**, **d**, **f**, **h**, **j**, **o** as competitors showed only peaks for **12h** and **12l** at the end of the assay, with that for **12l** being more prominent, giving a PF value of 3.6 for **12l**. Plotting melting points against protection factors revealed a discernible (r^2 = 0.79), though not strict correlation (see Supplementary Material, Fig. S27). Given that one technique is an equilibrium measurement, whereas the other a kinetic readout of a binding equilibrium, the correlation was considered satisfactory. Additional control experiments indicated that the selection assay is quite rugged. Neither switching to a different enzyme (nuclease S1, EC 3.1.30.1) nor employing shorter oligonucleotides (TCTTCTTTCT, TCTTCTGTCT, TCTTC-TATCT, TCTTCTCTCT and the intramolecular duplex AGAAGAAAGA-HEG-TCTTTCTTCT as the target) lowered the selectivity of the assay (Fig. S26).

DISCUSSION

The results obtained in the present study indicate that it is possible to employ PNA-like linker residues in oligonucleotide analogs designed to bind homopurine target duplexes with an interrupting base pair. Though the acyclic residue by itself does lower the stability of triplexes, with a melting point decrease of 14.5°C for **12a**:**13** compared to **14**:**13** (Table 1), acyl residues appended to it can more than compensate for this decrease in stability. One could have expected that the flexible linker would decrease triplex stability to a greater extent, since PNA–DNA hybrids are not generally found to have target strand affinities that are the linear sum of those of their subunits (31). In particular, the residues linking PNA and DNA portions have proven difficult to design and less prone to obey Watson–Crick base pairing rules as stringently as their neighbors (32,33). Earlier work on modified oligonucleotides for triplex formation had shown, however, that appending an intercalator internally can be more beneficial for target affinity than

Table 2. UV melting temperatures of triplexes between intramolecular duplexes containing different base pairs at position 7 and compound **12l** as the third strand at 0.3 μM strand concentration, 1 M NaCl and 10 mM phosphate buffer pH 6

Duplex	Melting point (°C) ± SD[a]	ΔT_m (°C)[b]
AGAAGA**T**AGAAAGA-HEG-TCTTTCT**A**TCTTCT (**13**)[c]	44.0 ± 0.6	–
AGAAGA**A**AGAAAGA-HEG-TCTTTCT**T**TCTTCT (**16**)[c]	39.0 ± 0.5	–5.0
AGAAGA**C**AGAAAGA-HEG-TCTTTCT**G**TCTTCT (**17**)[c]	44.1 ± 0.9	+0.1
AGAAGA**G**AGAAAGA-HEG-TCTTTCT**C**TCTTCT (**18**)[c]	36.6 ± 1.1	–7.4

[a]Standard deviation.
[b]Melting point difference to triplex **13**:**12l**.
[c]The nucleobases constituting or facing the base pair varied are in boldface, solely to facilitate recognizing them in the sequence.

appending it to the terminus, in agreement with the high target affinity found for **12l** (34).

Though anthraquinone-bearing **12l** does bind to the duplex target with greater affinity than the best among the unmodified oligonucleotides tested (G-containing **14**), it does show only moderate selectivity in its interactions with duplexes. Therefore, it is useful for bridging gaps in homopurine sequences caused by interrupting base pairs, but not for binding to a T:A base pair at that site selectively over a C:G base pair. While the purine/pyrimidine selectivity may be due to interactions between the nucleotides and the linker portion of the PNA-like residue, the lack of T/C selectivity may be due to intercalation of the aromatic portion. A naphthalene diimide intercalator tethered to a deoxycytidine residue in the third strand of a duplex has previously been reported to bind to dC:dG base pairs selectively (12). In the present molecular system, it is not clear whether both neighboring nucleobases can accomplish stacking with the intercalator. If the anthraquinone appended to the linker is indeed intercalating, one should expect a disruption of base stacking in the third strand, since no nucleobase would be located at the interrupting T:A base pair itself, and a kink in the helix that would compensate for this is not very

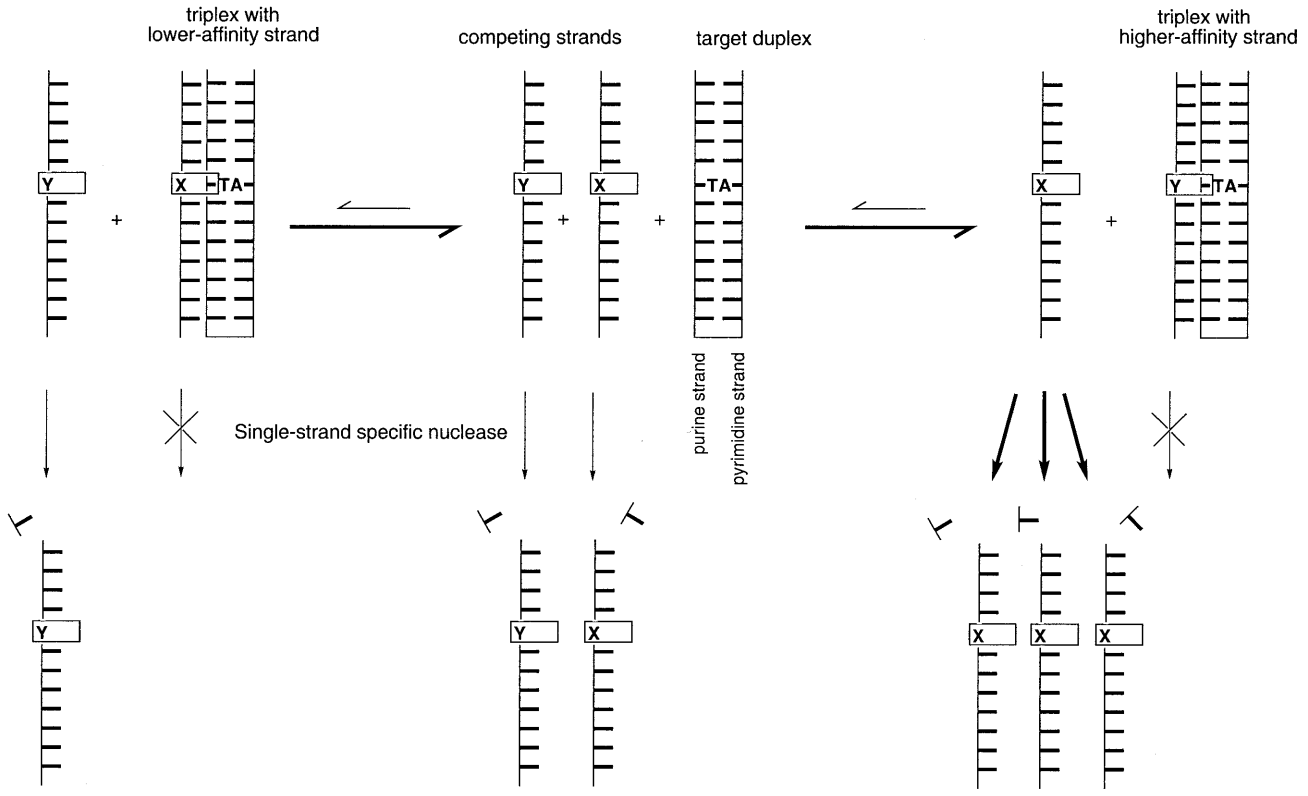


Figure 3. Cartoon of the principle underlying the nuclease selection of a strand with increased affinity for a target duplex. Two modified strands are competing for one equivalent of duplex. The modified strand with the higher affinity for the duplex is in the triplex state to the greatest extent, thus being shielded from nuclease attack, and persisting longer than its competitor. Selections were performed with libraries of up to seven modified strands per assay.

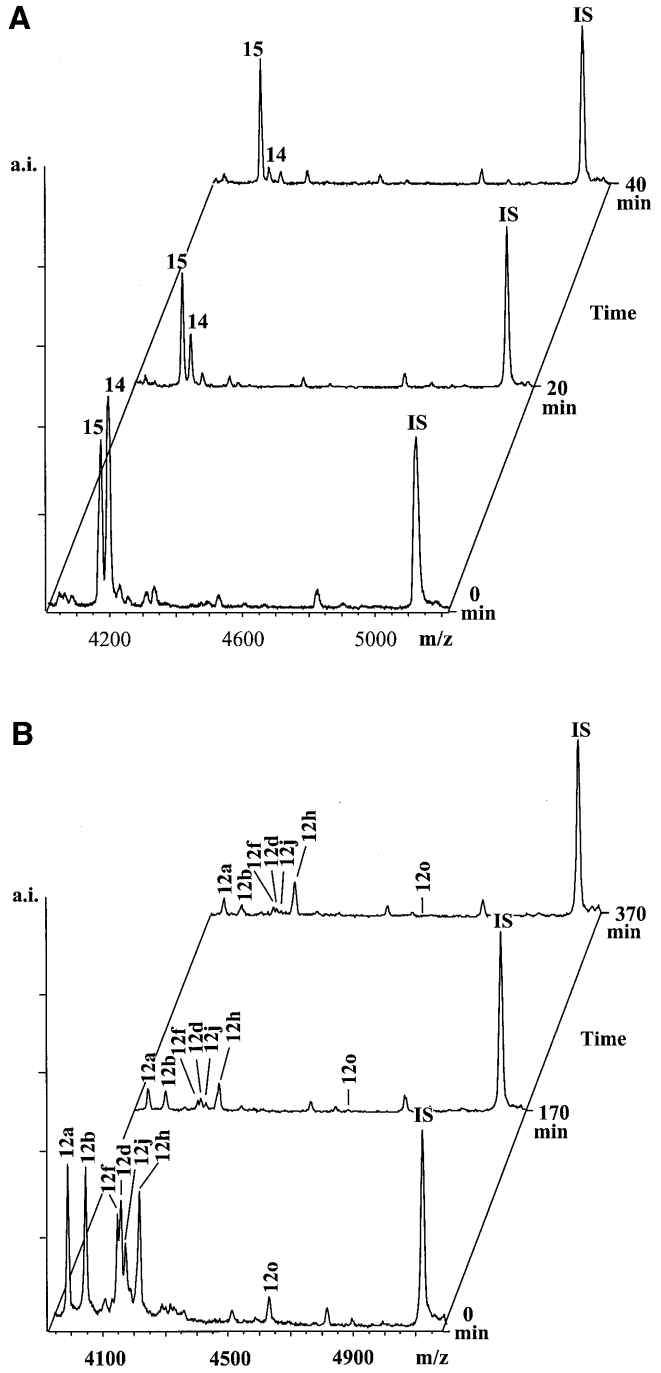


Figure 4. (**A**) Representative MALDI-TOF mass spectra from a monitored nuclease selection experiment involving intramolecular target duplex 5′-AGAAGAAA-GAAAGA-HEG-TCTTTCTTTCTTCT-3′ (**16**) and third strands 5′-TCTTCT-GTCTTTCT-3′ (**14**) and 5′-TCTTCTTTCTTTCT-3′ (**15**). The pseudomolecular ion signal for target strand **16** is outside the mass region of interest and is not shown. IS is the internal standard, DNA strand $(T)_{17}$, used for quantification. Assay conditions: 4×10^{-4} U of bovine spleen phosphodiesterase, 100 pmol per strand in 250 mM $(NH_4)_2SO_4$ pH 6, 22°C. Note the selective survival of **15**, which is fully complementary to **16** according to the Hoogsteen base pairing rules. (**B**) Representative MALDI-TOF mass spectra from a monitored nuclease selection assay involving target duplex 5′-AGAAGATAGAAAGA-HEG-TCTTTCTATCTTCT-3′ (**13**) and modified third strands **12a**, **12b**, **12d**, **12f**, **12h**, **12j** and **12o**. IS marks the peak of the internal standard $(T)_{17}$. Same experimental conditions as in (A).

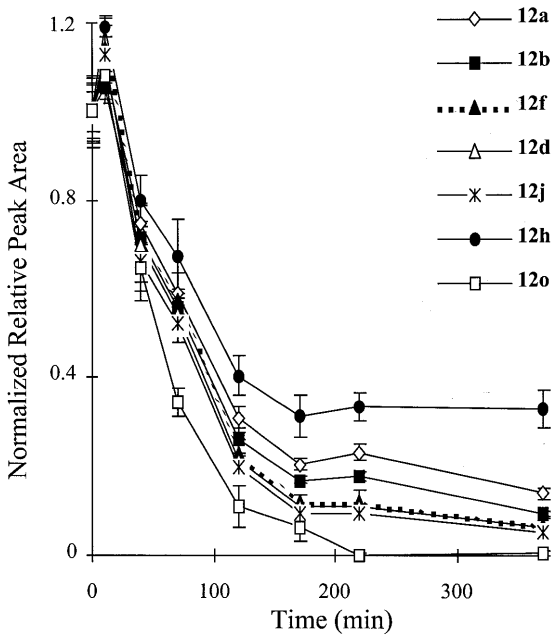


Figure 5. Kinetics of the disappearance of full-length oligonucleotides from the nuclease selection involving target duplex **13** and modified third strands **12a**, **12b**, **12d**, **12f**, **12h**, **12j** and **12o**. Selected MALDI spectra of the underlying assay are shown in Figure 4B. Data points are the average ± one standard deviation of the relative peak intensities (analyte/internal standard), normalized to $t_0 = 1$, obtained from ≥4 mass spectra.

Table 3. Protection factors obtained from MALDI monitored nuclease selection assays involving target duplex **13** and libraries of third strands

Library	Third strand	Nuclease protection factor
Library A	**12a**	1
	12b	0.9
	12d	0.6
	12f	0.6
	12h	1.3
	12j	0.6
	12o	0.5
Library B	**12a**	1
	12c	0.9
	12g	1.0
	12e	1.4
	12k	1.0
	12i	1.9
	12p	1.9

Selections were induced by adding 4×10^{-4} U of bovine spleen phosphodiesterase (EC 3.1.16.1) to a solution containing 100 pmol of each strand.

likely. The reduced hyperchromicity associated with the triplex–duplex transition of the modified strands compared to control triplex **15**:**16** (Fig. 2) confirms the hypothesis of disrupted stacking, and suggests an approach for optimizing the current linker residue.

Given that no truly T:A-specific residue was found, it is worthwhile considering the potential of combinatorial syntheses combined with nuclease selections. Since covalently linking the two strands of the target duplex not only helped to obtain well resolved triplex–duplex transitions in the UV melting experiments (*vide supra*), but also prevented peak overlap between target and probe strands in the MALDI spectra obtained from the nuclease selection, the current system seems appropriate for monitored selections. As also observed in our earlier work on nuclease selections (35), longer DNA sequences do not interfere with the MALDI detection of shorter probes. One could expect better selectivity when using longer target duplexes, since the termini of a bound third strand should be less accessible to an exonuclease if the duplex offers the protecting effect of overhangs at the termini of triple helical region. Together with the fact that triple helix formation is known to be highly specific, in that mismatch discrimination (as measured in decreased UV melting points) is greater than for duplexes (36), nuclease selections performed under optimized conditions may identify improved T:A-binding residues with high selectivity.

In the current study, the melting point differences between the derivatives employed are larger than those observed for comparable systems forming duplexes (37), and the protection factors also span a considerable range. Many of the folded nucleic acid structures that are key components of the transcriptional machinery, i.e. interesting drug targets, offer grooves on their surfaces. For these, nuclease selection experiments similar to those reported here may, therefore, identify binders where UV melting experiments become unreliable because of the large background caused by the unfolding target. Even if no systematic approach to finding new oligonucleotide binders was to emerge from the current work, the extension of our study to strands binding duplexes with more than one interrupting purine in their sequence emerges as a clear goal for the near future.

SUPPLEMENTARY MATERIAL

MALDI spectra of modified oligonucleotides, NMR spectra of small molecules and representative spectra from a nuclease assay with a decamer duplex and four strands competing for triplex formation are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank C. N. Tetzlaff for help with computer issues. This work was supported in part by the NIH (grant GM54783), the Deutsche Forschungsgemeinschaft (grant no. RI 1063/1), and the University of Constance. The NMR facility at the Chemistry Department of Tufts University was supported by N.S.F. grant CHE-9723772 (to C.R. and M. d'Alarcao).

REFERENCES

1. Blackburn,G.M. and Gait,M.J. (eds) (1996) *Nucleic Acids in Chemistry and Biology*. Oxford University Press, Oxford, UK.
2. Hill,W.E., Dahlberg,A.A., Garett,R.A., Moore,P.B., Schlessinger,D. and Warner,J.R. (eds) (1990) *The Ribosome, Structure, Function and Evolution*. American Society for Microbiology, Washington, DC.
3. Hecht,S.M. (ed.) (*1996*) *Bioorganic Chemistry – Nucleic Acids*. Oxford University Press, New York.
4. Lohse,J., Dahl,O. and Nielsen,P.E. (1999) Double duplex invasion by peptide nucleic acid: a general principle for sequence-specific targeting of double-stranded DNA. *Proc. Natl Acad. Sci. USA*, **96**, 11804–11808.
5. Ishihara,T. and Corey,D.R. (1999) Rules for strand invasion by chemically modified oligonucleotides. *J. Am. Chem. Soc.*, **121**, 2012–2020.
6. Dervan,P.B. and Bürli,R.W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.*, **3**, 688–693.
7. Thoung,N.T. and Hélène,C. (1993) Sequence-specific recognition and modification of double-helical DNA by oligonucleotides. *Angew. Chem. Int. Ed. Engl.*, **32**, 666–692.
8. Gowers,D.M. and Fox,K.R. (1999) Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.*, **27**, 1569–1577.
9. Giovannangeli,C. and Hélène,C. (2000) Triplex technology takes off. *Nat. Biotechnol.*, **18**, 1245–1246.
10. Barre,F.X., Ait-Si-Ali,S., Giovannangeli,C., Luis,R., Robin,P., Pritchard,L.L., Hélène,C. and Harel-Bellan,A. (2000) Unambiguous demonstration of triple-helix-directed gene modification. *Proc. Natl Acad. Sci. USA*, **97**, 3084–3088.
11. Matteucci,M., Lin,K.-Y., Huang,T., Wagner,R., Sternbach,D.D., Mehrotra,M. and Besterman,J.M. (1997) Sequence-specific targeting of duplex DNA using a camptothecin-triple helix forming oligonucleotide conjugate and topoisomerase I. *J. Am. Chem. Soc.*, **119**, 6939–6940.
12. Gianolio,D.A. and McLaughlin,L.W. (1999) Selective recognition of a dC-dG base pair by oligonucleotide-directed triplex formation using a dC residue tethering an intercalator. *J. Am. Chem. Soc.*, **121**, 6334–6335.
13. Stilz,H.U. and Dervan,P.B. (1993) Specific recognition of CG base pairs by 2-deoxynebularine within the purine·purine·pyrimidine triple-helix motif. *Biochemistry*, **31**, 2177–2185.
14. Huang,C.-Y., Bi,G. and Miller,P.S. (1996) Triplex formation by oligonucleotides containing novel deoxycytidine derivatives. *Nucleic Acids Res.*, **24**, 2606–2613.
15. Griffin,L.C. and Dervan,P.B. (1989) Recognition of thymine adenine base pairs by guanine in a pyrimidine triple helix motif. *Science*, **245**, 967–971.
16. Kiessling,L.L., Griffin,L.C. and Dervan,P.B. (1992) Flanking sequence effects within the pyrimidine triple-helix motif characterized by affinity cleaving. *Biochemistry*, **31**, 2829–2834.
17. Durland,R.H., Rao,T.S., Bodepudi,V., Seth,D.M., Jayaraman,K. and Revankar,G.R. (1995) Azole substituted oligonucleotides promote antiparallel triplex formation at non-homopurine duplex targets. *Nucleic Acids Res.*, **23**, 647–653.
18. Eldrup,A.B., Dahl,O. and Nielsen,P.E. (1997) A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J. Am. Chem. Soc.*, **119**, 11116–11117.
19. Tetzlaff,C.N., Schwope,I., Bleczinski,C.F., Steinberg,J.A. and Richert,C. (1998) A convenient synthesis of 5′-amino-5′-deoxythymidine and preparation of peptide-DNA hybrids. *Tetrahedron Lett.*, **39**, 4215–4218.
20. Bannwarth,W. (1988) Solid phase synthesis of oligodeoxynucleotides containing phosphoramidate internucleotide linkages and their specific chemical cleavage. *Helv. Chim. Acta*, **71**, 1517–1527.
21. Breipohl,G., Will,D.W., Peyman,A. and Uhlmann,E. (1997) Novel synthetic routes to PNA monomers and PNA-DNA linker molecules. *Tetrahedron*, **53**, 14671–14686.
22. Altman,R.K., Schwope,I., Sarracino,D.A., Tetzlaff,C.N., Bleczinski,C.F. and Richert,C. (1999) Selection of modified oligonucleotides with increased target affinity via MALDI-monitored nuclease survival assays. *J. Combin. Chem.*, **1**, 493–508.
23. Mayfield,C. and Miller,D. (1994) Effect of abasic linker substitution on triplex formation, Sp1 binding, and specificity in an oligonucleotide targeted to the human Ha-ras promoter. *Nucleic Acids Res.*, **22**, 1909–1916.
24. Sasaki,S., Nakashima,S., Nagatsugi,F., Tanaka,Y., Hisatome,M. and Maeda,M. (1995) Design of a novel artificial nucleobase for the selective formation of a triple-complex with a cytosine-guanine base pair. *Tetrahedron Lett.*, **52**, 9521–9524.
25. Kiseleva,V.V., Gakh,A.A. and Fainzilberg,A.A. (1990) Synthesis of C-azolylacetic acid esters via ethyl carbethoxyacetimidate. *Bull. Acad. Sci. USSR Div. Chem. Sci.*, **39**, 1888–1895.
26. Durand,M., Chevrie,K., Chassignol,M., Thuong,N.T. and Maurizot,J.C. (1990) Circular dichroism studies of an oligodeoxyribonucleotide containing a hairpin loop made of a hexaethylene glycol chain: conformation and stability. *Nucleic Acids Res.*, **18**, 6353–6359.
27. Altmann,S., Labhardt,A.M., Bur,D., Lehmann,C., Bannwarth,W., Billeter,M., Wüthrich,K. and Leupin,W. (1995) NMR studies of DNA

duplexes singly cross-linked by different synthetic linkers. *Nucleic Acids Res.*, **23**, 4827–4835.
28. Berlin,K., Jain,R.K., Tetzlaff,C., Steinbeck,C. and Richert,C. (1997) Spectrometrically monitored selection experiments: quantitative laser desorption mass spectrometry of small chemical libraries. *Chem. Biol.*, **4**, 63–77.
29. Sarracino,D. and Richert,C. (1996) Quantitative MALDI-TOF spectrometry of oligonucleotides and a nuclease assay. *Bioorg. Med. Chem. Lett.*, **6**, 2543–2548.
30. Mokhir,A.A. and Richert,C. (2000) Synthesis and monitored selection of 5′-nucleobase-capped oligodeoxyribonucleotides. *Nucleic Acids Res.*, **28**, 4254–4265.
31. Van der Laan,A.C., Havenaar,P., Oosting,R.S., Kuyl-Yeheskiely,E., Uhlmann,E. and van Boom,J.H. (1998) Optimization of the binding properties of PNA-(5′)-DNA chimera. *Bioorg. Med. Chem. Lett.*, **8**, 663–668.
32. Uhlmann,E., Will,D.W., Breipohl,G., Langner,D. and Ryte,A. (1996) Synthesis and properties of PNA/DNA chimeras. *Angew. Chem. Int. Ed. Engl.*, **35**, 2632–2635.
33. Uhlmann,E., Peyman,A., Breipohl,G. and Will,D.W. (1998) PNA: synthetic polyamide nucleic acids with unusual binding properties. *Angew. Chem. Int Ed. Engl.*, **37**, 2797–2823.
34. Francois,J.C. and Hélène,C. (1999) Recognition of hairpin-containing single-stranded DNA by oligonucleotides containing internal acridine derivatives. *Bioconjug. Chem.*, **10**, 439–446.
35. Stoerker,J., Mayo,J., Tetzlaff,C.N., Sarracino,D.A., Schwope,I. and Richert,C. (2000) Rapid genotyping via MALDI-monitored nuclease selection from probe libraries. *Nat. Biotechnol.*, **18**, 1213–1216.
36. Kool,E.T. (1998) Recognition of DNA, RNA, and proteins by circular oligonucleotides. *Acc. Chem. Res.*, **31**, 502–510.
37. Berlin,K., Jain,R.K., Simon,M.D. and Richert,C. (1998) A porphyrin embedded in DNA. *J. Org. Chem.*, **63**, 1527–1535.