Published online 31 July 2018

Nucleic Acids Research, 2018, Vol. 46, No. 16 ***8069–8078***
doi: 10.1093/nar/gky669

A strongly pairing fifth base: oligonucleotides with a *C*-nucleoside replacing thymidine

Tanja J. Walter and Clemens Richert*

Institute of Organic Chemistry, University of Stuttgart, 70569 Stuttgart, Germany

Received June 10, 2018; Revised July 06, 2018; Editorial Decision July 11, 2018; Accepted July 13, 2018

ABSTRACT

There are five canonical bases in DNA and RNA. Each base has its particular molecular recognition properties and base pairing strength. Thymine and uracil form only two hydrogen bonds when pairing with adenine, and duplexes rich in A:T base pairs are more labile than duplexes rich in C and G, making some sequences difficult to detect via hybridization in a genomic context. Here we report the synthesis of an ethynylmethylpyridone *C*-nucleoside, abbreviated 'W', that presents a similar recognition surface as thymidine in the major groove but pairs with A about as strongly as C pairs with G. A phosphoramidite building block was synthesized that allows for incorporation of W residues via automated synthesis in high yield. Melting point increases over duplexes containing T:A pairs of up to 17.5°C, or up to 5.8°C per residue were measured for oligonucleotides containing W. Further, the new base shows excellent fidelity, with a single mismatched G opposite W causing a melting point depression of up to 20.5°C. The strongly pairing replacement for thymidine is only slightly larger than its natural counterpart and performs well in different sequence contexts. It can be used to target weakly pairing A-rich sequences in biological studies.

INTRODUCTION

The most common secondary structure element found in nucleic acid complexes is the double helix. Helix formation is driven by base pairing (1). Base pairs are stabilized by hydrogen bonds and stacking of the aromatic heterocycles. The base pairing rules in duplexes depend on the nucleobases and the sugar/phosphate backbone (2), but in natural DNA, Watson–Crick pairing dominates in duplexes between complementary strands (3). Further, Watson–Crick pairing also underlies replication and the folding of strands into intricate 3D structures (4). The canonical bases found in DNA and RNA (adenine, cytosine, guanine, thymine and uracil) are well adapted to their roles in biology and also allow for rapid formation and dissociation of duplexes, as the charged backbone favors extended structures of well balanced thermal stability (5), but not ideal for simultaneous detection in sets of diverse sequences.

Among the canonical nucleosides, thymidine is unique, as it is only found in DNA. In the cell, this nucleoside is synthesized enzymatically from deoxyuridine monophosphate in reactions catalyzed by thymidylate synthase, an enzyme with an important role in cancer treatment. The methyl group of thymidine that distinguishes it from deoxyuridine is then used for molecular recognition by proteins. This includes enzymes critical for genome integrity and repair. When hydrolytic deamination of deoxycytidines leads to deoxyuridines, the most common mutation in DNA (6), the base pairing pattern changes from a base complementary to guanine to a base complementary to adenine. Unless quickly detected and repaired, this can cause mutations during replication. The missing methyl group at the 5-position of deoxyuridine residues originating from hydrolysis can also lead to a misincorporation of deoxyuridine (7). Thus, the methyl group is an important marker, facilitating recognition by repair enzymes, most notably uracil-DNA glycosylase, which removes uracil from DNA (8).

The base pairs between adenine and its complementary bases have long been known to be weaker than those between guanine and cytosine (9). As a result, duplex stabilities vary greatly between different sequences (10), complicating parallel detection of multiple DNA strands with high fidelity or sequence discrimination (11,12). This has led to a quest for oligonucleotides that give isostable duplexes, i.e. duplexes whose stability is independent of sequence, so that efficient hybridization can be achieved at a fixed, 'universal' hybridization temperature (13,14). At this universal temperature, good mismatch discrimination and thus high fidelity would be observed for each probe. The most obvious way to modulate affinity is through modification of the nucleobases. Among the modified nucleobases that bind their target bases with increased affinity are clamp-like base derivatives (15), expanded bases (16), and bases with alkynyl substituents. The latter can improve stacking in duplexes or triplexes (17–19), and are typically introduced via

*To whom correspondence should be addressed. Tel: +49 711 685 64311; Fax: +49 711 685 64321; Email: lehrstuhl-2@oc.uni-stuttgart.de

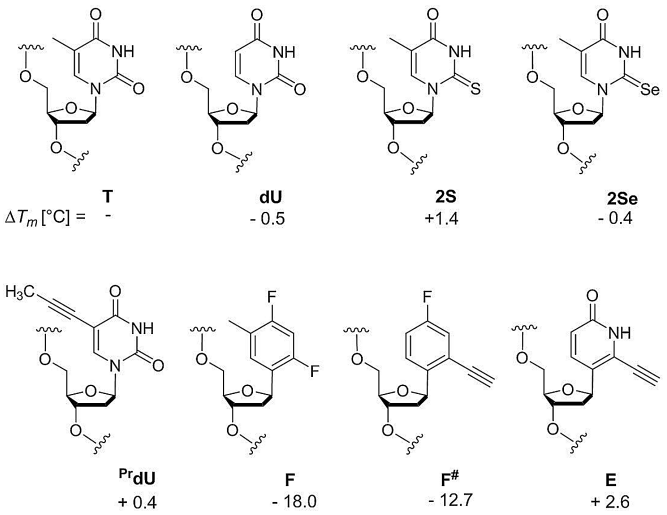


Figure 1. Literature-known thymidine analogues and their effect on duplex stability compared to thymidine.

palladium-catalyzed cross coupling, either on the nucleoside or the oligonucleotide level (20).

The pairing of T or U suffers from the fact that adenine displays only two hydrogen bonding sites at its Watson/Crick pairing edge (the amino group at position 6 and the nitrogen at position 1) and that both also pair well with guanine (21,22), making it difficult to suppress mispairing. This affects experiments, such as Southern or Nothern blotting, gene expression profiling with microarrays, or PCR. Given the potential rewards, it is not surprising that derivatives or analogs of thymidine have been synthesized that have the potential to pair more strongly with adenine than T. Figure 1 shows some such structures, together with the UV-melting point difference found for duplexes containing the T analogs, compared to the DNA duplexes featuring a natural T residue. The stabilizing effect of the methyl group at the 5-position of T can be gleaned from the comparison with deoxyuridine. Without the methyl group, the duplex melting point typically drops by approx. 0.5°C. In other structures, the carbonyl oxygen at the 2-position was replaced with sulphur (23) or selenium (24), leading to a melting point increase of approx. 1.4°C in the case of sulphur and a drop in melting point of approx. 0.4°C in the case of selenium. Likewise, when the methyl group was replaced with alkynyl substituents, such as in 5-propargyl-2′-deoxyuridine (25–28), modest increased in the melting point were found. A seemingly unlikely candidate for the replacement are *C*-nucleoside isosteres of T, such as 2,4-difluorotoluene (29). This base analog has a strongly destabilizing effect when placed opposite A, but when an ethynyl group replaces the fluorine facing C2 of A, affinity improves, as shown for $F^{\#}$, when compared with difluorotoluene F (30), even though in $F^{\#}$ the methyl group is missing. Reconstitution of the lactame group at positions 3 and 4 of thymine was achieved in the ethynylpyridone found in the nucleoside dubbed "E", significantly improving pairing strength (31).

For the *C*-nucleoside E, the ethynyl groups replacing the carbonyl oxygen is at the 2-position of thymidine to allow for interactions with the CH fragment at the 2-position of adenine and to strengthen the stacking interactions with neighboring bases in a duplex. But, E lacks the methyl group marker of T and is difficult to incorporate in oligonucleotides (31). Attempts to generate a phosphoramidite for automated DNA synthesis failed. Instead, the nucleoside had to be manually coupled to DNA strands previously phosphitylated on the solid support (31), resulting in modest yields, mostly due to side reactions caused by the poorly shielded lactame group. So, the absence of the methyl group at the 5-position caused both a problematic reactivity and made the base different from T for proteins binding in the major groove. The methyl group was also desirable to increase duplex stability through stacking interactions. Here, we report the synthesis of a *C*-nucleoside that overcomes these limitations. It features the methyl group at the 5-position of thymine, making it a full replacement for the fifth base. The new nucleobase, dubbed 'W', pairs more strongly than T and E and is readily incorporated in oligonucleotides by automated synthesis.

MATERIALS AND METHODS

General

Chemicals and solvents were purchased from *Acros Organics* (Geel, Belgium), *Carbolution* (Saarbrücken, Germany), *Carbosynth* (Compton, Berksire, Great Britain), *Sigma-Aldrich* (Deisenhofen, Germany) or *TCI* (Eschborn, Germany) and were used without further purification. If not otherwise noted, reactions were carried out under argon atmosphere. Thin layer chromatography was performed with *Merck Millipore* TLC Silica 60 F_{254} aluminum sheets with visualization by ultraviolet light and staining with phosphormolybdato cerium(IV) sulfate solution (25 g phosphormolybdic acid hydrate, 10 g cerium(IV) sulfate tetrahydrate, 60 mL conc. sulfuric acid, diluted with water to 1 l). Column chromatography was carried out with *Merck Millipore* silica 60 Dm (particle size: 0.040–0.063 mm). Glycal **8** was synthesized according to a literature protocol (31). Unmodified oligonucleotides were purchased from *Biomers* (Ulm, Germany) in lyophilized, HPLC-purified form, dissolved in water and used without further purification. Yields of oligonucleotides are based on the isolated amount of oligonucleotide, and the loading of the controlled pore glass support, as given by the supplier. Extinction coefficients for oligonucleotides were calculated through linear combination of the extinction coefficients of nucleotides. The ^{1}H-NMR (300 MHz), ^{31}P-NMR (121 MHz) and ^{13}C-NMR (75 MHz) spectra were measured on a Bruker Avance 300 spectrometer, and ^{1}H-NMR (500 MHz) and ^{13}C-NMR (125 MHz) spectra were measured on a Bruker Avance 500 MHz spectrometer in deuterated solvents. Chemical shifts are reported as δ values (ppm), relative to the solvent peak and coupling constants (*J*) are given in Hz. Multiplicities were designated as s, singlet; d, doublet; t, triplet; m, multiplet and br, for a broad peak. Infrared spectra were recorded on a Bruker Vector 22FT-IR Spectrometer, and selected peaks are reported in wave numbers (cm^{-1}). High resolution mass spectrometry was performed on all stable new compounds, using a Bruker microTOFQ spectrometer running *Bruker* Compass Data

Analysis software, version 4.2. MALDI-TOF mass spectra were measured in a linear negative mode on a *Bruker* microFlex mass spectrometer with a matrix consisting of a 2:1 (*v*/*v*) mixture of 2,4,6-trihydroxyacetophenone solution (THAP, 0.2 M in ethanol) and diammonium citrate (0.1 M in water). The *m*/*z* values are those of the maximum of the unresolved isotope envelope of the peak for the pseudomolecular ion $[M–H]^-$.

Synthesis of *C*-nucleoside 13

*3-Methyl-5-nitro-2-pivaloyloxypyridine (**2**).* 3-Methyl-5-nitro-2-pyridone (1.03 g, 6.68 mmol) was dissolved in dry acetonitrile (35 mL) and NEt_3 (1.4 mL, 10.02 mmol, 1.5 eq.) was added. Under ice bath cooling, pivaloyl chloride (1.2 mL, 10.02 mmol, 1.5 eq.) was added. The reaction mixture was stirred for 15 min at 0°C. After TLC showed full conversion, a saturated solution of $NaHCO_3$ (5 mL) and ethyl acetate (10 mL) were added. The organic layer was separated, and the aqueous phase was extracted twice with ethyl acetate (2 × 20 mL). The combined organic layers were dried over Na_2SO_4 and concentrated *in vacuo*. The resulting crude of **2** was used in the next step without further purification. An analytical sample was purified by chromatography, using silica (30 g), eluting with petroleum ether/ethyl acetate (5/1, v/v). TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.79$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 9.07 (d, *J* = 2.6 Hz, 1H), 8.39 (d, *J* = 2.3 Hz, 1H), 2.31 (s, 3H), 1.41 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 175.6, 161.1, 142.9, 142.2, 135.3, 127.1, 39.5, 27.0, 16.1; HRMS (ESI-TOF) *m*/*z* calcd. for $C_{11}H_{14}N_2O_4$ $[M+Na]^+$ 261.0846; found 261.0816.

*5-Amino-3-methyl-2-pivaloyloxypyridine (**3**).* In a pressure-stable reaction vessel, 3-methyl-5-nitro-2-pivaloyloxypyridine (**2**, 1.59 g, 6.68 mmol) was dissolved in methanol (30 mL) and palladium on charcoal (179 mg, 10% wt, dry) was added. Under a H_2 atmosphere (2.5 bar), the reaction mixture was shaken for 2 h at room temperature. The mixture was filtered over Celite and washed with ethyl acetate (20 mL). The filtrate was concentrated *in vacuo*. The resulting crude product (**3**) was used in the next step without further purification. An analytical sample was purified by chromatography, using silica (30 g) eluting with petroleum ether/ethyl acetate (3/1, v/v). TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.14$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.66 (d, *J* = 2.8 Hz, 1H), 6.90 (d, *J* = 2.7, Hz, 1H), 3.59 (br s, 2H), 2.08 (s, 3H), 1.37 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 176.9, 149.5, 141.5, 132.4, 126.7, 125.4, 39.1, 27.2, 15.9; HRMS (ESI-TOF) *m*/*z* calcd. for $C_{11}H_{16}N_2O_2$ $[M+Na]^+$ 231.1104; found 231.1109.

*3-Amino-2-bromo-5-methyl-6-pivaloyloxypyridine (**4**).* Amine **3** (1.39 g, 6.68 mmol) was dissolved in acetonitrile (20 mL), and the solution was cooled to 0°C. A solution of *N*-bromosuccinimide (1.19 g, 6.68 mmol, 1 eq.) in acetonitrile (10 mL) was added dropwise within 10 min, while stirring at 0°C. After TLC showed full conversion, water (15 mL) was added, and the mixture was extracted twice with ethyl acetate (2 × 20 mL). The combined organic layers were washed with brine (10 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The resulting crude product was purified by chromatography, using silica (30 g) and a gradient of 10–30% ethyl acetate in petroleum ether, yielding **4** as a dark brown solid 1.42 g (4.94 mmol, 74% over three steps). TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.66$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 6.93 (s, 1H), 2.04 (s, 3H), 1.36 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 176.6, 147.2, 140.1, 126.4, 125.6, 122.6, 39.1, 27.2, 15.5; HRMS (ESI-TOF) m/z calcd. for $C_{11}H_{15}BrN_2O_2$ $[M+Na]^+$ 309.0209; found 309.0220.

*2-Bromo-5-methyl-4-pivaloyloxypyridin-3-yl diazonium tetrafluoroborate (**5**).* An aliquot of $BF_3{\cdot}OEt_2$ (1 mL, 7.94 mmol, 1.5 eq.) was added to a solution of amine **4** (1.52 g, 5.29 mmol) in dry THF (5 mL) at −10°C. The reaction mixture was stirred for 2 min, then *tert*-butyl nitrite (800 μL, 6.35 mmol, 1.2 eq.) was added. After 10 min, the precipitated solid was filtered off and washed three times with cold diethyl ether (3 × 10 mL). Diazonium salt **5** was obtained as a light yellow solid, 1.93 g (4.97 mmol, 94%). The product is unstable and was not studied by mass spectrometry for this reason. ^{1}H-NMR (500 MHz, CD_3CN): δ = 7.99 (s, 1H), 2.16 (s, 3H), 1.16 (s, 9H); ^{13}C-NMR (125 MHz, CD_3CN): δ = 179.0, 163.9, 143.6, 134.3, 128.0, 37.7, 26.2, 14.9; IR: $\tilde{\nu}$ = 3088, 2973, 2285, 2218 (R-N_2^+), 1698, 1416, 1039 cm^{-1}.

*2-Bromo-3-iodo-5-methyl-6-pivaloyloxypyridine (**6**).* To a solution of diazonium salt **5** (812 mg, 2.09 mmol) in dry acetonitrile (20 mL), potassium iodide (417 mg, 2.51 mmol, 1.2 eq.) was added in one portion. The reaction mixture was stirred at room temperature for 2 h. After TLC showed full conversion, water (10 mL) was added. The mixture was extracted twice with ethyl acetate (2 × 20 mL). The combined organic layers were dried over Na_2SO_4 and then concentrated *in vacuo*. The resulting crude product was purified by chromatography, using silica (30 g) and a gradient of 20–50% ethyl acetate in petroleum ether, yielding **6** as an orange solid (637 mg, 1.60 mmol, 76%). TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.80$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.99 (s, 1H), 2.12 (s, 3H), 1.41 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 175.6, 155.9, 151.6, 142.7, 126.5, 95.9, 39.3, 27.1, 15.1; HRMS (ESI-TOF) *m*/*z* calcd. for $C_{11}H_{13}BrINO_2$ $[M+Na]^+$ 419.9067; found 419.9073.

2-Bromo-5-methyl-6-pivaloyloxy-3-[3′-O-(tert-butyldimethylsilyl)-2′-deoxy-2′,3′-didehydro-β-D-*ribofuranos-1′-yl]-pyridine (**8**).* Samples of $Pd(OAc)_2$ (124 mg, 0.55 mmol, 0.2 eq) and $P(PhF_5)_3$ (590 mg, 1.11 mmol, 0.4 eq) were dissolved in dry acetonitrile (10 mL) and stirred at room temperature for 30 min. The solution was then added to a stirred mixture of 2-bromo-3-iodo-5-methyl-6-pivaloyloxypyridine (**6**, 1.54 g, 3.88 mmol, 1.4 eq), Ag_2CO_3 (764 mg, 2.77 mmol, 1 eq) and glycal **7** (31) (638 mg, 2.77 mmol, 1 eq) in dry acetonitrile (10 mL). The reaction mixture was stirred at room temperature for 5 h. After TLC showed full conversion, the reaction mixture was filtered through celite, eluting with ethyl acetate (20 mL), and the solvent was then removed *in*

vacuo. The resulting crude product (**8**) was used in the subsequent step without further purification. An analytical sample was purified by chromatography, using silica and a gradient of 0–1% methanol in dichloromethane, yielding **8** as an off-white solid. TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.86$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.87 (s, 1H), 6.03 (dd, $J = 4.1$, $J = 1.3$, 1H), 4.9 (s, 1H), 4.69 (br t, $J = 2.3$, 1H), 3.85 (ddd $J = 21.0$, $J = 12.3$, $J = 2.9$, 2H), 2.2 (s, 3H), 1.41 (s, 9H), 0.97 (s, 9H), 0.27 (s, 3H), 0.25 (s, 3H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 175.9, 155.4, 151.6, 141.2, 137.2, 136.0, 125.3, 100.5, 83.6, 82.2, 63.1, 39.2, 27.1, 25.5, 15.6, 4.9, 5.0; ESI-MS *m/z* calcd. for $C_{22}H_{34}BrNO_5Si$ $[M+Na]^+$ 522.13; found 522.13.

2-Bromo-5-methyl-6-pivaloyloxy-3-(2′,3′-didehydro-2′,3′-dideoxy-3′-oxo-β-D-*ribofuranos-1′-yl)pyridine (**9**).* In a polypropylene tube, crude **8** (1.39 g, 2.77 mmol) was dissolved in THF (30 mL) and 3 HF·NEt_3 (900 μL, 5.54 mmol, 2 eq) was added. After 30 min, TLC showed full conversion. To quench remaining HF, methoxytrimethylsilane (2 mL) was added, and the mixture was stirred for an additional 30 min. The mixture was then filtered through celite, eluting with ethyl acetate (20 mL). The filtrate was concentrated *in vacuo*. The resulting crude **9** was used in the next step without purification. An analytical sample was purified by chromatography, using silica and a gradient of methanol (0–2%) in dichloromethane, yielding **9** as a off-white solid. TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.50$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.71 (s, 1H), 5.18 (dd, $J = 10.7$, $J = 5.9$, 1H), 3.88 (t, $J = 3.6$, 1H), 3.79 (m, 2H), 2.96 (dd, $J = 17.8$, $J = 6.3$, 1H), 2.05 (dd, $J = 18.3$, $J = 10.3$, 1H), 1.96 (s, 3H),1.18 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 212.7, 175.9, 155.6, 139.6, 135.8, 135.1, 125.7, 82.3, 75.1, 61.5, 44.1, 39.3, 27.1, 15.5; HRMS (ESI-TOF) m/z calcd. for $C_{16}H_{20}BrNO_5$ $[M+Na]^+$ 408.0417; found 408.0423.

2-Bromo-5-methyl-6-pivaloyloxy-3′-(2′-deoxy-β-D-*ribofuranos-1′-yl)pyridine (**10**).* Crude **9** (1.07 g, 2.77 mmol) was dissolved in a mixture of acetonitrile and acetic acid (1/1, v/v, 40 mL) and, under ice bath cooling, $NaBH(OAc)_3$ (880 mg, 4.15 mmol, 1.5 eq.) was added in one portion. After 15 min, TLC showed full conversion. The reaction was quenched with MeOH (10 mL), and the solvents were removed *in vacuo*. The resulting crude was chromatographed on silica (40 g), using a gradient of 1–5% methanol in dichloromethane, to yield 734 mg of title compound **10** (1.89 mmol, 68% over three steps) as an off-white solid. TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.14$; ^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.66 (s, 1H), 5.25 (dd, J = 9.8, J = 5.7, 1 H), 4.34 (m, 1H), 3.96 (m, 1H), 3.77 (m, 2H), 2.5, (ddd, *J*=15.1, *J*=11.6, *J*=4.5, 1H), 2.1 (s, 3H), 1.79 (ddd, *J*=13.6, *J*=5.9, *J*=2.5, 1H), 1.36 (s, 9H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 176.1, 155.1, 139.4, 136.9, 134.9, 125.2, 87.1, 73.4, 63.3, 42.6, 39.3, 27.1, 15.5; HRMS (ESI-TOF) *m/z* calcd. for $C_{16}H_{22}BrNO_5$ $[M+Na]^+$ 410.0575; found 410.0574.

3-Methyl-2-pivaloyloxy-5-(2′-deoxy-β-D-*ribofuranos-1′-yl)-6-(tri*iso*propylsilylethynyl)-pyridine (**11**).* In a pressure-stable reaction vessel, nucleoside **10** (567 mg, 1.46 mmol) was dissolved in dry DMF (3 mL). Then, $Pd(PPh_3)Cl_2$ (102 mg, 0.15 mmol, 0.1 eq), CuI (56 mg, 0.29 mmol, 0.2 eq), NEt_3 (400 μL, 2.92 mmol, 2 eq), and triisopropylsilyacetylene (820 μL, 3.65 mmol, 2.5 eq) were added. The reaction vessel was sealed, and the reaction mixture was stirred at 80°C for 2 h. After TLC showed full conversion, the black suspension was filtered through celite, eluting with ethyl acetate (20 mL). The filtrate was concentrated *in vacuo*. The crude product was chromatogred on silica with a gradient of 1–5% methanol in dichloromethane, yielding 690 mg (1.38 mmol, 95%) of the desired nucleoside (**11**) as a light brown glass. TLC (petroleum ether/ethyl acetate, 1/1, v/v): $R_f = 0.23$; ^{1}H-NMR (300 MHz, $CDCl_3$): δ = 7.65 (s, 1H), 5.55 (dd, $J = 9.9$, $J = 5.9$, 1H), 4.35-4.31 (m, 1H), 3.96-3.71 (m, 2H), 2.45 (ddd, $J = 12.9$, $J = 5.8$, $J = 2.3$, 1H), 2.10 (s, 3H), 1.83 (ddd, $J = 9.4$, $J = 5.3$, $J = 3.7$, 1H), 1.59 (br s, 1H), 1.32 (s, 9H), 1.08 (s, 21H); ^{13}C-NMR (125 MHz, $CDCl_3$): δ = 136.5, 124.5, 123.5, 102.6, 96.3, 87.0, 84.9, 76.0, 73.6, 63.2, 42.6, 39.2, 38.9, 27.2, 18.6, 11.1; HRMS (ESI-TOF) *m/z* calcd. for $C_{27}H_{43}NO_5Si$ $[M+Na]^+$ 512.2803; found 512.2802.

3-Methyl-2-pivaloyloxy-5-[2′-deoxy-5′-O-*(dimethoxytrityl)*-β-D-*ribofuranos-1′-yl]-6-(tri*iso*propylsilylethynyl)-pyridine (**12**).* Nucleoside **11** (230 mg, 0.46 mmol) was coevaporated twice from pyridine and dissolved in dry pyridine (5 mL). Then, DMAP (3 mg, 0.02 mmol, 0.05 eq.) was added, and the reaction mixture was stirred for 30 min. Subsequently, DMT-Cl that had been coevaporated twice from dry pyridine (204 mg, 0.60 mmol, 1.3 eq.), dissolved in pyridine (2 mL), was added. The reaction mixture was stirred for 18 h at room temperature. After TLC showed full conversion, volatiles were removed *in vacuo* and the crude was purified by chromatography, using silica (20 g, deactivated with dichloromethane containing 0.5% NEt_3 prior to use) and a gradient of methanol (0–2%) in dichloromethane to afford the desired nucleoside (**12**) as a colorless solid (283 mg, 0.36 mmol, 77%). TLC (dichloromethane/methanol, 99/1, v/v): $R_f = 0.74$; ^{1}H-NMR (500 MHz, CD_3CN): δ = 7.97 (s, 1H), 7.51–7.24 (m, 9H), 6.90 (d, $J = 8.6$, 4H), 5.55 (dd, $J = 9.9$, $J = 5.9$, 1H), 4.36–4.32 (m, 1H), 4.04-4.01 (m, 1H), 3.79 (s, 6H), 3.32–3.25 (m, 2H), 2.46 (ddd, $J = 12.9$, $J = 5.8$, $J = 2.3$, 1H), 2.03 (s, 3H), 1.95 (ddd, $J = 9.4$, $J = 5.3$, $J = 3.7$, 1H), 1.39 (s, 9H), 1.18 (s, 21H); ^{13}C-NMR (125 MHz, CD_3CN): δ = 175.8, 158.2, 155.1, 144.6, 140.1, 137.6,135.6, 135.5, 135.2, 129.5, 129.4, 127.6, 127.3, 126.3, 126.2, 112.6, 102.2, 95.5, 85.9, 85.6, 75.9, 72.7, 63.7, 54.4, 63.7, 54.4, 42.3, 38.4, 25.8, 25.8, 17.6, 17.3, 14.8, 10.5, 10.4, 0.5, 0.3, 0.2, 0.1; HRMS (ESI-TOF) *m/z* calcd. for $C_{48}H_{61}NO_7SiNa$ $[M+Na]^+$ 814.4110; found 814.4117.

3-Methyl-2-pivaloyloxy-5-[2′-deoxy-3′-O-*(2-cyanoethyl-*N,N-*diisopropylamino)-phosphino-5′*-O-*(dimethoxytrityl)*-β-D-*ribofuranos-1′-yl]-6-(tri*iso*propylsilylethynyl)-pyridine (**13**).* Nucleoside **12** (705 mg, 0.89 mmol) was co-evaporated twice from acetonirile, and then dissolved in dry acetonitrile (10 mL). Diisopropylammonium tetrazolide (26) (107 mg, 0.62 mmol, 0.7 eq.) was added, and the reaction mixture was stirred for 30 min. Then,

2-cyanoethyl-*N*,*N*-diisopropylaminochlorophosphite (340 μL, 1.07 mmol, 1.2 eq.) was added. The reaction mixture was stirred for 4 h at room temperature. After TLC showed full conversion, a saturated aqueous solution of $NaHCO_3$ (10 mL) was added. The aqueous phase was extracted twice with methyl *tert*-butyl ether (2 × 20 mL). The combined organic layers were dried over Na_2SO_4, concentrated *in vacuo*, and the resulting crude product was purified by chromatography, using silica (previously deactivated with dichloromethane containing 1% NEt_3) and methyl *tert*-butyl ether containing NEt_3 (1%) to afford the desired nucleoside **13** as a colorless solid (778 mg, 0.78 mmol, 88%). The phosphoramidite was stored in dry form under argon at −10°C. The phosphoramidite is a labile mixture of diastereomers that was not characterized in detail. TLC (petroleum ether/ethyl acetate, 5/1, v/v, 0.5% Net_3): $R_f = 0.52$; ^{31}P-NMR (121 MHz, CD_3CN): δ = 147.6, 146.8.

DNA synthesis and deprotection

The synthesis of oligonucleotide strands **14**–**19** was performed via automated solid-phase DNA synthesis on an Expedite 8909 synthesizer on a 1 μmol scale using commercial phosphoramidites or phosphoramidite **13**. The coupling time for **13** was extended to 400 s. At the end of the synthesis, the DMT protecting groups at the 5′-position were removed on the support. The controlled pore glass was then placed in a polypropylene vessel and treated with saturated aqueous ammonia (25%, 1 mL) for 18 h at 55°C. Subsequently, the solution was aspirated, and the support was washed with water (2 × 100 μL). Then, excess ammonia was removed from the combined solution with a gentle stream of nitrogen, and the pH was found to be ∼7. The samples were lyophilized and treated with a TBAF solution (1 M in THF, 500 μL), adding some water (5–20 μL), depending on the length of the oligonucleotide, until it was fully dissolved. After 5 h, triethylammonium acetate buffer (1 M, 500 μL) and water (500 μL) were added, and the THF was removed with a gentle stream of nitrogen.

HPLC purification

Reversed-phase HPLC was carried out on a HPLC system with a type *L-6200 Intelligent Pump* and a photodiode detector *L-3000* from *Hitachi* (Tokyo, Japan). For this, lyophilized oligonucleotides were dissolved in water (1 mL), the solution was filtered (pore size: 0.45 μm) and subjected to reversed-phase HPLC, using a Nucleosil C18 column (250 × 4.6 mm, *Macherey-Nagel*, Düren, Germany) with a gradient of acetonitrile in triethylammonium acetate buffer (0.1 M, pH 7) and detection at 260 nm. Two different gradient protocols were used for oligonucleotides **14**–**16** (1–15% CH_3CN in 43 min) and oligonucleotides **17**–**19** (1–25% CH_3CN in 50 min).

Analytical data for modified oligonucleotides

5′-CWGCAG-3′ (14). HPLC: gradient of 1–15% CH_3CN in 43 min, product detected at t_r = 32.5 min; yield: 50 nmol, 5%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{61}H_{75}N_{22}O_{33}P_5$ $[M\text{-}H]^-$ 1798, found 1798. **5′-CCWCCTT-3′ (15)**. HPLC: gradient of CH_3CN, 1–15% in 43 min, product detected at t_r = 32.5 min; yield: 390 nmol, 39%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{69}H_{89}N_{17}O_{42}P_6$ $[M–H]^-$ 2013, found 2012. **5′-TGGWWGAC-3′ (16)** HPLC: gradient of CH_3CN, 1–15% in 43 min, product detected at t_r = 37.0 min; yield: 60 nmol, 6%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{85}H_{103}N_{27}O_{46}P_7$ $[M–H]^-$ 2454, found 2455. **5′-WWGTCAWWG-3′ (17)** HPLC: gradient of CH_3CN, 1–25% in 50 min, product detected at t_r = 37.5 min; yield: 50 nmol, 5%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{100}H_{117}N_{22}O_{52}P_8$ $[M\text{-}H]^-$ 2705, found 2705. **5′-GTCWWWWWWGC-3′ (18)** HPLC: gradient of CH_3CN, 1–25% in 50 min, product detected at t_r = 38.0 min; yield: 20 nmol, 2%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{126}H_{152}N_{24}O_{65}P_{10}$ $[M–H]^-$ 3350, found 3348. **5′-CTTTTCWTTCTT-3′ (19)** HPLC: gradient of CH_3CN 1–25% in 50 min, product detected at t_r = 38.0 min; yield: 330 nmol, 33%; MALDI-TOF-MS: *m*/*z*: calcd. for $C_{120}H_{157}N_{26}O_{78}P_{11}$ $[M\text{-}H]^-$ 3550, found 3548.

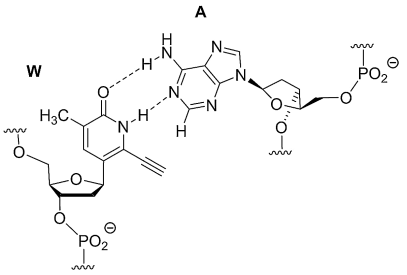
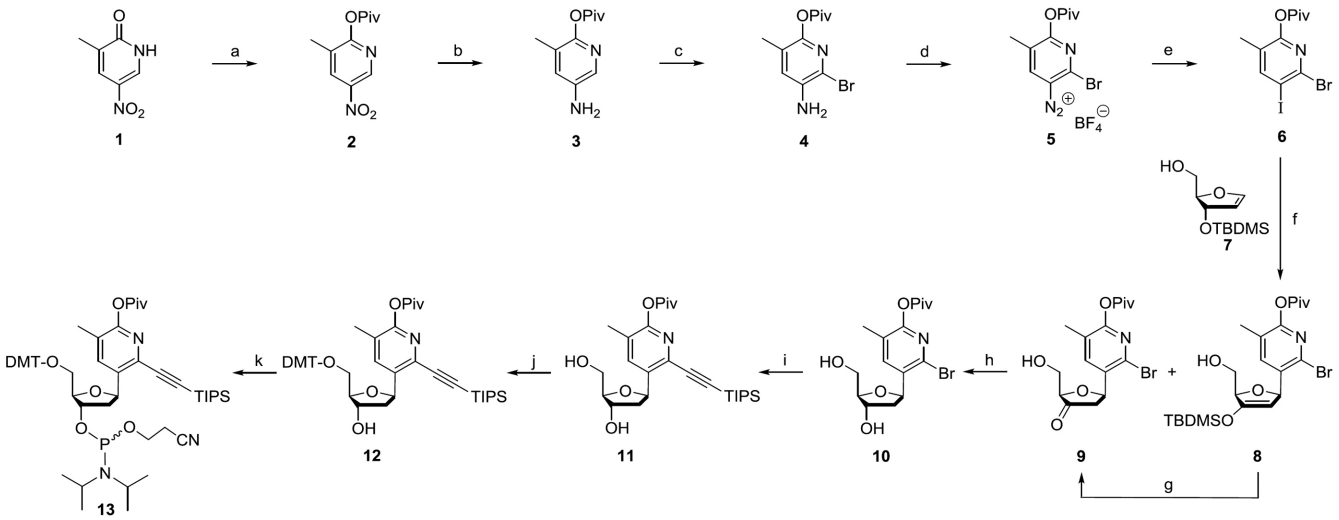


Figure 2. Structure of the W residue in a DNA strand and its base pair with A in a duplex. The hydrogen at position 2 of adenine is shown to highlight the spatial proximity to the ethynyl substituent.

RESULTS AND DISCUSSION

Figure 2 shows the structure of the *C*-nucleoside W and its base pair with a deoxyadenosine residue in a DNA duplex. The phosphoramidite building block for the incorporation of W in DNA strands was prepared in 15 synthetic steps. Among the challenges in synthesizing the *C*-nucleoside were the reactivity of the ethynylpyridone toward phosphitylating reagents, necessitating a protecting group compatible with oligonucleotide synthesis, the introduction and deprotection of the ethynyl substituent, and the coupling reaction introducing the aglycone to the deoxyribose. Exploratory work confirmed that free pyridone moieties react readily with standard phosphitylating reagents, and that the benzyl protecting group employed in the synthesis of E (31) is quite difficult to remove without side reactions at the alkyne or anomerization. Further, the trimethylsilylacetylene commonly used to introduce ethynyl substituents has a boiling point of 52°C, complicating Sonogashira reactions at elevated temperatures, and producing nucleosides prone to lose the silyl group prematurely in later steps, so that reactive intermediates are formed (33). Finally, Heck reactions under established reaction conditions were low yielding, all



Scheme 1. Synthesis of *C*-nucleoside **10** and its phosphoramidite building block **13**. (a) Piv-Cl, NEt_3, CH_3CN; (b) H_2, Pd/C, MeOH, 2.5 bar; (**c**) NBS, CH_3CN, 0°C, 74% over three steps; (d) $BF_3{\cdot}OEt_2$, *t*$BuNO_2$, THF, 0°C, 94%; (e) KI, CH_3CN, 76%; (f) **7**, $Pd(OAc)_2$, $P(PhF_5)_3$, Ag_2CO_3, CH_3CN; (g) 3 HF·NEt_3, THF; (h) $NaBH(OAc)_3$, CH_3CN/AcOH, 0°C, 68% over 3 steps; (i) tri*iso*propylsilylacetylene, $[Pd(PPh_3)_2Cl_2]$, CuI, NEt_3, DMF, 80°C, 90%; (j) DMT-Cl, pyridine, DMAP, 77%; (k) $(iPr_2N)_2P(OC_2H_4CN)$, DIPAT, CH_3CN, 88%, DIPAT = di*iso*propylammonium tetrazolide, DMT = 4,4′-dimethoxytrityl, NBS = *N*-bromosuccinimide, Piv = pivaloyl, TBAF = tetrabutylammoniumfluoride, TBDMS = *tert*butyldimethylsilyl, TIPS = tri*iso*propylsilyl.

but precluding the preparation of sufficient quantities of the nucleoside building block for routine DNA synthesis.

Scheme 1 shows the successful synthesis of phosphoramidite **13**. It starts with methylnitropyridone **1** (34), which is commercially available for less than $ 10/g. This compound was selectively *O*-acylated to **2**, using pivaloyl chloride in acetonitrile. Subsequent reduction with hydrogen on Pd/C gave aniline **3**, which was brominated with NBS in the cold to give **4** in 74% yield over three steps. Diazotization of the brominated aniline produced **5**, which readily precipitated from the reaction mixture under optimized reaction conditions. When **5** was used directly in the Heck reaction with glycal **7**, the low yields known from similar routes were obtained (31). Even after careful optimization, the yield of **10** never exceeded 13% over three steps. An important step was therefore to switch to iodide **6**, accessible from **5** in 76% yield, and to use the recently reported coupling conditions of Hocek *et al.* (35), which led to high-yielding reactions. Glycal **7** was synthesized in four steps, with an overall yield of 54%, following the route of Minuth and Richert (31). At the base concentration chosen, partial loss of the silyl group of enol ether **8** is common during Heck reactions, producing variable amounts of ketone **9**, which is readily separated from **8** by column chromatography. Remaining silyl enol ether **8** was converted to **9**, and the two fractions of **9** were pooled. The ketone was then diastereoselectively reduced with $NaBH(OAc)_3$ in the cold, furnishing **10** in 68% overall yield from **6** on a gram scale. The benzyl derivative of **10** crystallized readily, and an X-ray crystal structure confirmed its stereochemical configuration (see Figure S45, Supplementary Data). For the subsequent Sonogashira reaction to **11**, we used the less volatile and more stable tri*iso*propylsilylacetylene to complete the elaboration of the carbon framework of the *C*-nucleoside. Tritylation gave **12**, and was followed by phosphitylation, which then led to **13** in 12% yield over 15 steps. Batches of up to 700 mg of the phosphoramidite were readily prepared.

Access to **13** allowed the exploration of the pairing properties of the ethynylmethylpyridone in different duplexes. We chose the sequences shown in Figure 3. They included selfcomplementary hexamer **14**, and non-selfcomplementary sequences **15**–**19** with one, two, four, or six W residues. Further, dodecamer **19** with a single T-to-W replacement was prepared to compare the UV-melting point of its duplex with those of other *C*-nucleosides incorporated in the same sequence (29–31).

Figure 4 shows the HPLC chromatogram of crude **19**, prepared by automated DNA synthesis on controlled-pore glass, followed by two-step deprotection, consisting of treatment with aqueous ammonia and then TBAF in THF, i.e. conditions similar to those used in RNA syntheses (36). After HPLC purification, oligonucleotide **19** was obtained in 33% yield. While heavily modified sequences often give low yields in oligonucleotide synthesis, even for undecamer **18** with its six neighboring W residues, the product peak was the most intense peak in the chromatogram of the crude product and the pure product was obtained in 2% overall yield, confirming that **13** is well-behaved in automated DNA synthesis.

With the W-containing strands in hand, we proceeded to studying duplex stabilities by UV-melting analysis. The melting point (T_m) of the duplex of dodecamer **19** with its target strand 3′-GAAAAGAAAGAA-5′ (**20**) was 4.4°C higher than that of the duplex featuring thymine at the position of W (Table 1, Figure 5A). The melting point of the duplex with the W:A base pair was also 1.2°C higher than that of the duplex with E at the same position (31) (Table S1, Supplementary Data). When a C:G base pair was placed at this position, the duplex had almost the identical thermal

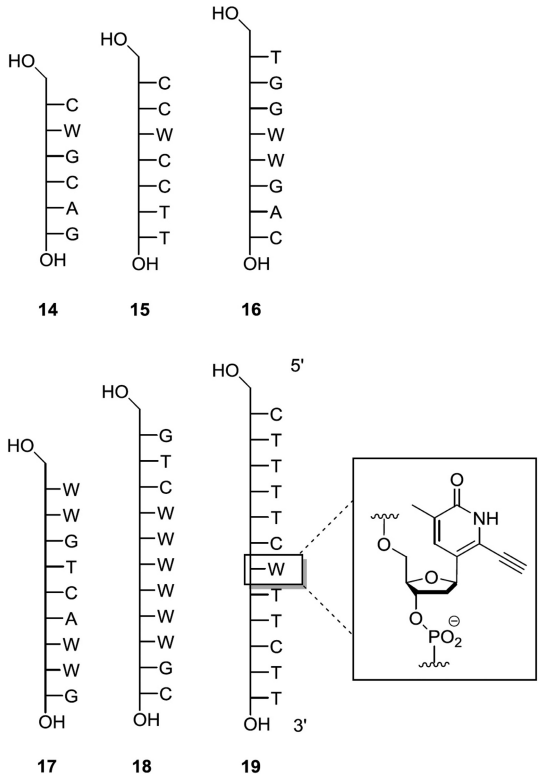


Figure 3. Oligonucleotides synthesized using phosphoramidite **13**.

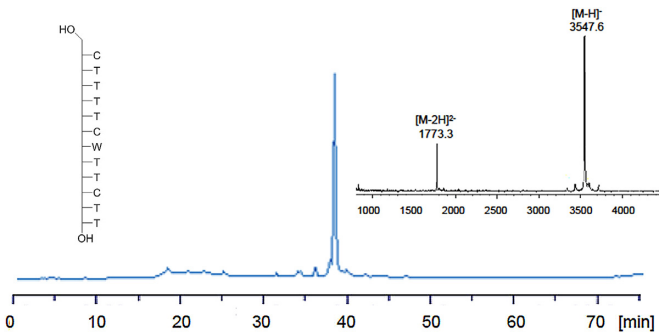


Figure 4. HPLC chromatogram of crude **19** (C18 column, TEAA buffer/CH_3CN, 55°C), and MALDI-TOF mass spectrum of the fraction containing the oligonucleotide (inset).

stability as that with the W:A pair (44.5 versus 44.6°C, Table 1). When RNA target strand *r*(3′-GAAAAGAAAGAA-5′) (**24**) was used as complementary strand, the W for T replacement still gave a ΔT_m of +2.9°C. Finally, the stabilizing effect of W residues in a short duplex was measured, which was found to be a ΔT_m of +11.5°C (+5.7°C per nucleotide) for self-complementary hexamer **14**, compared to control duplex formed by **25** (Table 1). Exchanging the W:A base pair with C:G base pairs in the self-complementary hexamer **25** to give **14** again confirmed the similarity in base pairing strength of the W:A and the C:G base pairs, with duplex melting points of 42.7 versus 41.0°C (Table 1).

Next, we determined the base pairing selectivity of W in melting curves with dodecamer **19** and its unmodified ver-

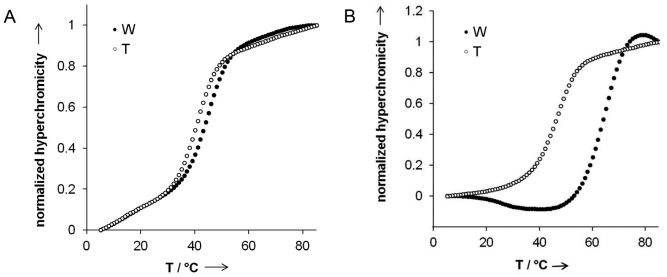


Figure 5. Representative UV-melting curves of W-containing duplexes and control duplexes of unmodified DNA, detected at 260 nm; (A) duplexes of **19** (filled circles) and **21** (open circles) with target strand **20**; (B) duplexes of **18** (filled circles) and **44** (open circles) with target strand **43**. The legend states which nucleobase is found in the probe strand of the respective duplex. Conditions: oligonucleotide concentration 1.6 μM (**19**/**20**/**21**) or 4.9 μM (**18**/**43**/**44**), 10 mM PIPES, pH 7.0, 100 mM NaCl, 10 mM $MgCl_2$. See Tables 1 and 3 for more details.

sion **21** (Table 2). A melting point decrease of 20.5°C was found for a single mismatched G opposite W (Table 2), indicating excellent discrimination against wobble pairing. The ΔTm is larger than that measured for E (19.8°C; 31) and much larger than that reported for 2-thiothymidine (12.0°C, 23). Further, not a single one of the different mismatches tested gave a ΔT_m of less than 14°C, confirming that W high fidelity in a broad range of contexts. In contrast, T-containing control strand **21** showed the expected low selectivity for A:T versus G:T pairs, with a melting point decrease of 9.5°C in the latter case. To the best of our knowledge W is the most selective replacement for T known to date.

In a last set of experiments, the ability of W to pair with adenine was studied in a more diverse set of sequence contexts. This included oligonucleotides **15**–**18** forming duplexes with **31**, **35**, **39** or **43** (Table 3). The UV-melting point increase, relative to the duplex with natural T:A pairs, was between +5.8°C per W residue for the shortest duplex (**15**:**31**) and +2.9°C per residue for the longest duplex, featuring six W residues, (**18**:**43**, overall increase of 17.5°C). Figure 5B shows that even for this six-fold modified duplex, a sharp sigmoidal transition is found. This confirms highly cooperative melting, even with multiple neighboring W, a trait that is typical for natural DNA and not often found for heavily modified strands. Taken together, the melting point data confirmed that W is strongly stabilizing and well behaved in DNA duplexes, even when multiple modified bases are placed in a sequence.

In order to better explain the high cooperativity in duplex formation, we then studied the steric and electrostatic profile of W, as compared to T, its natural counterpart. For this, the structures of the free bases were calculated *ab initio* on the B3LYP/def2-SVP level in TURBOLOME, version 7.2 (see: http://www.turbomole.com). Figure 6A shows electrostatic maps of the two bases. The settings for the graphical representation are similar to those used for nucleobases in the past (37). It can be discerned that the electron density at the hydrogen bonding sites and the methyl group are near-identical for T and W. Only at the 1-, 2- and 6-positions

Table 1. UV-melting points of DNA:DNA or DNA:RNA duplexes

Sequences[a]	T_m[b] [°C]	ΔT_m[c] [°C]
3′-GAAAAG**A**AAGAA-5′ (**20**)		
5′-CTTTTC**T**TTCTT-3′ (**21**)	40.2	
5′-CTTTTC**W**TTCTT-3′ (**19**)	44.6	4.4
3′-GAAAAG**G**AAGAA-5′ (**22**)		
5′-CTTTTC**C**TTCTT-3′ (**23**)	44.5	4.3
r(3′-GAAAAG**A**AAGAA-5′) (**24**)		
5′-CTTTTC**T**TTCTT-3′ (**21**)	41.4	
5′-CTTTTC**W**TTCTT-3′ (**19**)	44.5	2.9
(5′-C**T**GC**A**G-3′)$_2$ (**25**)	31.3	
(5′-C**W**GC**A**G-3′)$_2$ (**14**)	42.7	11.5
(5′-C**C**GC**G**G-3′)$_2$ (**26**)	41.0	9.8

[a]Target strand listed first, probe strands shown with indentation; self-complementary sequences listed with a double indentation. Bases at the position varied are in boldface.
[b]Average of four curves, detected at 260 nm; conditions: 5.6 ± 0.7 μM strands, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl and 10 mM $MgCl_2$.
[c]To control duplex with T:A base pair.

(pyrimidine numbering) is the *C*-nucleoside less polar than T. The lipophilicity at positions 1 and 6 should be of little consequence for base pairing, as these positions are directed toward the deoxyribose ring, not any pairing partner. The ethynyl group at position 2 is well suited to interact with the lower Watson–Crick face of adenine, with a likely van der Waals contact to the CH fragment at position 2 of the purine, as shown in Figure 6B.

The shape of the ethynyl group of W also explains the exquisite selectivity for A versus G. While the alkynyl substituent provides shape complementarity to A, it induces a steric conflict with G, when a similar hydrogen bonding geometry is attempted as in a T:G or U:G wobble base pair (Figure 6C). Apparently, W combines a size and hydrogen bonding capability similar to that of T, so that it can be integrated seamlessly into DNA duplexes, with improved shape complementarity and stacking capabilities, making it well suited for oligonucleotides designed to bind A-containing target strands.

Table 2. UV-melting points of DNA duplexes with a matched or mismatched nucleobase opposite a T or a W residue

Sequences[a]	Pairing[b]	T_m[c] [°C]	ΔT_m[d] [°C]
5′-CTTTTC**T**TTCTT-3′ (**21**)			
3′-GAAAAG**A**AAGAA-5′ (**27**)	T:A	40.2	-
3′-GAAAAG**T**AAGAA-5′ (**28**)	T:T	25.5	−14.7
3′-GAAAAG**C**AAGAA-5′ (**29**)	T:C	26.1	−14.1
3′-GAAAAG**G**AAGAA-5′ (**30**)	T:G	30.7	−9.5
5′-CTTTTC**W**TTCTT-3′ (**19**)			
3′-GAAAAG**A**AAGAA-5′ (**27**)	W:A	44.6	-
3′-GAAAAG**T**AAGAA-5′ (**28**)	W:T	30.4	−14.2
3′-GAAAAG**C**AAGAA-5′ (**29**)	W:C	26.7	−17.9
3′-GAAAAG**G**AAGAA-5′ (**30**)	W:G	24.1	−20.5

[a]Target strand listed first, probe strands below, with indentation; bases of interest in boldface.
[b]Base combination at the position of interest.
[c]Average of four curves; conditions: 4.0 μM strands, 10 mM PIPES, pH 7.0, 100 mM NaCl and 10 mM $MgCl_2$.
[d]To *Tm* of fully matched duplex.

Table 3. UV-melting points of DNA:DNA duplexes with one or several W residues together with data from control duplexes

Sequences[a]	T_m[b] [°C]	ΔT_m[c] [°C]
3′-GGAGGAA-5′ (**31**)		
5′-CC**T**CCTT-3′ (**32**)	26.5	
5′-CC**W**CCTT-3′ (**15**)	32.3	5.8
3′-GG**G**GGAA-5′ (**33**)		
5′-CC**C**CCTT-3′ (**34**)	33.6	7.1
3′-ACC**A**ACTG-5′ (**35**)		
5′-TGG**T**TGAC-3′ (**36**)	36.3	
5′-TGG**W**WGAC-3′ (**16**)	45.3	9.0
3′-ACC**G**GCTG-5′ (**37**)		
5′-TGG**C**CGAC-3′ (**38**)	48.2	11.9
3′-**A**ACAGT**A**AC-5′ (**39**)		
5′-**T**TGTCA**T**TG-3′ (**40**)	32.3	
5′-**W**WGTCA**W**WG-3′ (**17**)	45.5	13.2
3′-CAG**A**AAAAAACG-5′ (**43**)		
5′-GTC**T**TTTTTGC-3′ (**44**)	46.5	
5′-GTC**W**WWWWWGC-3′ (**18**)	64.0	17.5

[a]Target strand listed first, probe strands with an indentation. The bases of interest in boldface.
[b]Average of four curves, detected at 260 nm. Conditions: strand concentration 5.6 ± 0.7 μM, 10 mM PIPES buffer, pH 7.0, 100 mM NaCl and 10 mM $MgCl_2$.
[c]To control duplex with T:A base pair(s).

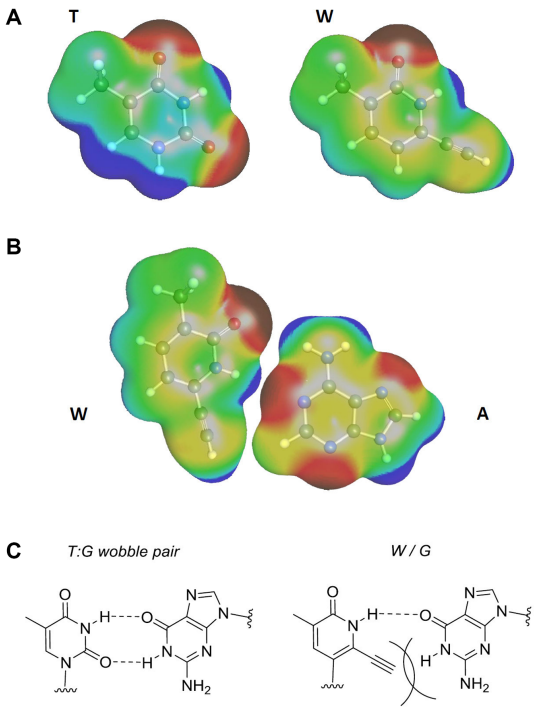


Figure 6. Gaining insights into the structural basis for the pairing and stacking properties of W. (A) Calculated structures of ethynylmethylpyridine W and thymine (T), optimized at the B3LYP/def2-SVP-level with the molecular electrostatic potential (MEP) mapped onto the 0.001 electron density isosurface; (B) Shape complementarity between W and A, as visualized by graphically placing the two bases into close proximity. (C) Structural model explaining the suppression of wobble base pairing: While T and G can forma a wobble pair without steric conflicts, the ethynyl group of W produces a steric clash with G, and the O4 atom is not well positioned to form a hydrogen bond. The electron density and the MEP shown in (A) and (B) were computed with TURBOMOLE and visualized using TmoleX Version 4.3.1 (44).The color code ranges from –0.05 au (red) to +0.05 au (blue).

CONCLUSIONS

In conclusion, we describe a new *C*-nucleoside analog of thymidine that pairs about as strongly with adenine as cytosine pairs with guanine. This analog, abbreviated W, can be readily incorporated in oligonucleotides using phosphoramidite **13** and automated chain assembly. With W replacing T in their sequence, oligodeoxynucleotides form duplexes with excellent base pairing selectivity. Wobble base pairing with G is suppressed, with UV-melting point depression of more than 20°C for a single W:G mismatch in the sequence context studied here. The consistently high affinity for adenine in complementary strands and the predictable pairing properties of W make it promising for applications ranging from high fidelity microarrays (14) to functional DNA nanostructures with improved nucleation during folding (38–40), capture agents for mRNAs, or high-affinity primers. Studies aimed at using W for reading out nucleic acid sequences via enzyme-free copying (41–43) are also under way. Independent of such practical considerations, it will be interesting to ask whether 'W is better than T' in a biological context, and if so, what the consequences were, if nature had access to improved binders for adenine.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank Dr B. Miehlich for *ab initio* calculations, D. Göhringer for expert technical assistance, Dr M. Minuth for discussions, and Nicholas Birchall and Helmut Griesser for a review of the manuscript.

FUNDING

DFG [RI 1063/15-1]; University of Stuttgart. Funding for open access charge: University of Stuttgart.
Conflict of interest statement. None declared.

REFERENCES

1. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer Verlag, NY.
2. Beier,M., Reck,F., Wagner,T., Krishnamurthy,R. and Eschenmoser,A. (1999) Chemical etiology of nucleic acid structure: comparing pentopyranosyl (2′-4′) oligonucleotides with RNA. *Science*, **283**, 699–703.
3. Borer,P.N., Dengler,B., Tinoco,I. and Uhlenbeck,O.C. (1974) Stability of ribonucleic acid double-stranded helices. *J. Mol. Biol.*, **86**, 843–853.
4. Egli,M. (1997) In vitro selected receptors rationalized: the first 3D structures of RNA aptamer/substrate complexes. *Angew. Chem. Int. Ed. Engl.*, **36**, 480–482.
5. Richert,C., Roughton,A.L. and Benner,S.A. (1996) Non-ionic analogs of RNA with dimethylenesulfone bridges. *J. Am. Chem. Soc.*, **118**, 4518–4531.
6. Lindahl,T. (1994) Instability and decay of primary structure of DNA. *Nature*, **362**, 709–715
7. Tye,B.K., Nyman,P.O., Lehman,I.R, Hochhauser,S. and Weiss,B. (1977) Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proc. Natl. Acad. Sci. U.S.A*, **74**, 154–157.
8. Barnes,D.E., Lindahl,T. and Sedgwick,B. (1993) DNA repair. *Curr. Opin. Cell Biol.*, **5**, 424–433.
9. Pranata,J., Wierschke,S.G. and Jorgensen,W.L. (1991) OPLS potential functions for nucleotide bases - relative association constants of hydrogen-bonded base pairs in chloroform. *J. Am. Chem. Soc.*, **113**, 2810–2819.
10. SantaLucia,J and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
11. Kennedy,G.C., Hatsuzaki,H., Dong,S., Liu,W.M., Huang,J., Liu,G., Su,X., Cao,M., Chen,W., Zhang,J. *et al.* (2003) Large-scale genotyping of complex DNA. *Nat. Biotechnol.*, **21**, 1233–1237.
12. Matsuzaki,H., Loi,H., Dong,S., Tsai,Y., Fang,J., Law,J., Di,X., Liu,W., Yang,G., Liu,G. *et al.* (2004) Parallel genotyping of over 10,000 SNPs using a One-Primer assay on a High-Density oligonucleotide array. *Genome Res.*, **14**, 414–425.
13. Nguyen,H.-K., Fournier,O., Asseline,U., Dupret,D. and Thuong,N. T. (1999) Smoothing the thermal stability of DNA duplexes by using modified nucleosides and chaotropic agents. *Nucleic Acids Res.*, **27**, 1492–1498.
14. Ahlborn,C., Siegmund,K. and Richert,C. (2007) Isostable DNA. *J. Am. Chem. Soc.*, **129**, 15218–15232.
15. Lin,K.-Y. and Matteucci,M.D. (1998) A cytosine analogue capable of clamp-like binding to a guanine in helical nucleic acids. *J. Am. Chem. Soc.*, **120**, 8531–8532.
16. Okamoto,A., Tainaka,K. and Saito,I. (2003) Clear distinction of purine bases on the complementary strand by a fluorescence change of a novel fluorescent nucleoside. *J. Am. Chem. Soc.*, **125**, 4972–4973.
17. Wagner,R.W., Matteucci,M.D., Lewis,J.G., Gutierrez,A.J., Moulds,C. and Froehler,B.C (1993) Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science*, **260**, 1510–1513.
18. Chaudhuri,N.C. and Kool,E.T. (1995) Very high-affinity DNA recognition by bicyclic and cross-linked oligonucleotides. *J. Am. Chem. Soc.*, **117**, 10434–10442.
19. Wagner,R.W., Matteucci,M.D., Grant,D., Huang,T. and Froehler,B.C. (1996) Potent and selective inhibition of gene expression by an antisense heptanucleotide. *Nat. Biotechnol.*, **14**, 840–844.
20. Kottysch,T., Ahlborn,C., Brotzel,F. and Richert,C. (2004) Stabilizing or destabilizing oligodeoxynucleotide duplexes containing single 2′-deoxyuridine residues with 5-alkynyl substituents. *Chem. Eur. J.*, **10**, 4017–4028.
21. Crick,F.H.C. (1966) Codon—anticodon pairing: the wobble hypothesis. *J. Mol. Biol.*, **19**, 548–555.
22. Allawi,H.T. and SantaLucia J. Jr (1997) Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry*, **36**, 10581–10594.
23. Sintim,H.O. and Kool,E.T. (2006) Enhanced base pairing and replication efficiency of thiothymidines, expanded-size variants of thymidine. *J. Am. Chem. Soc.*, **128**, 396–397.
24. Hassan,A.E.A., Sheng,J., Zhang,W. and Huang,Z. (2010) High fidelity of base pairing by 2-selenothymidine in DNA. *J. Am. Chem. Soc.*, **132**, 2120–2212.
25. Wagner,R.W., Matteucci,M.D., Lewis,J.G., Gutierrez,A.J., Moulds,C. and Froehler,B. C. (1993) Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science*, **260**, 1510–1513.
26. Chaput,J.C., Sinha,S. and Switzer,C. (2002) 5-Propynyluracil.diaminopurine: an efficient base-pair for non-enzymatic transcription of DNA. *Chem. Comm.*, 1568–1569.
27. Gyi,J.I., Gao,D., Conn,G L., Trent,J.O., Brown,T. and Lane,A.N. (2003) Solution structure of a DNA-RNA duplex containing 5-propynyl U and C; comparison with 5-Me modifications. *Nucleic Acids Res.*, **31**, 2683–2693.
28. Barnes,T.W. and Turner,D.H. (2001) Long-range cooperativity in molecular recognition of RNA by oligodeoxynucleotides with multiple C5-(1-propynyl) pyrimidines. *J. Am. Chem. Soc.*, **123**, 4107–4118.
29. Moran,S., Ren,R.X.-F., Rumney,S. and Kool,E.T. (1997) Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.*, **119**, 2056–2057.
30. Griesang,N. and Richert,C. (2002) Oligonucleotides containing a nucleotide analog with an ethynylfluorobenzene as nucleobase surrogate. *Tetrahedron Lett.*, **43**, 8755–8758
31. Minuth,M. and Richert,C. (2013) A nucleobase analog that pairs strongly with adenine. *Angew. Chem. Int. Ed.*, **52**, 10874–10877.

32. Barone,A.D., Tang,J.-Y. and Caruthers,M.H. (1984) In situ activation of bis-dialkylaminophosphines – a new method for synthesizing deoxyoligonucleotides on polymer supports. *Nucleic Acids Res.*, **12**, 4051–4061.
33. Ingale,S.A., Mei,H., Leonard,P. and Seela,F. (2013) Ethynyl side chain hydration during synthesis and workup of "Clickable" oligonucleotides: bypassing acetyl group formation by triisopropylsilyl protection. *J. Org. Chem.*, **78**, 11271–11282.
34. Shvekhgeimer,M.-G.A. (1996) Synthesis of halogenopyridines (review). *Chem. Heterocyclic Comp.*, **32**, 987–1015.
35. Kubelka,T., Slavětínská,L., Eigner,V. and Hocek,M. (2013) Synthesis of 2,6-disubstituted pyridin-3-yl *C*-2′-deoxyribonucleosides through chemoselective transformations of bromo-chloropyridine *C*-nucleosides. *Org. Biomol. Chem.*, **11**, 4702–4718.
36. Usman,N., Ogilvie,K.K., Jiang,M.Y. and Cedergren,R.J. (1987) The automated chemical synthesis of long oligoribuncleotides using 2′-O-silylated ribonucleoside 3′-O-phosphoramidites on a controlled-pore glass support: synthesis of a 43-nucleotide sequence similar to the 3′-half molecule of an Escherichia coli formylmethionine tRNA. *J. Am. Chem. Soc.*, **109**, 7845–7854.
37. Mignon,P., Loverix,S., Steyaert,J. and Geerlings,P. (2005) Influence of the $\pi-\pi$ interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.*, **33**, 1779–1789.
38. Rothemund,P.W.K. (2006) Folding DNA to create nanoscale shapes and patterns. *Nature*, **440**, 297–302.
39. Park,S.H., Yin,P., Liu,Y., Reif,J.H., LaBean,T.H. and Yan,H. (2005) Programmable DNA self-assemblies for nanoscale organization of ligands and proteins. *Nano Lett.*, **5**, 729–733.
40. Dunn,K.E., Dannenberg,F., Ouldridge,T.E., Kwiatkowska,M., Turberfield,A.J. and Bath,J. (2015) Guiding the folding pathway of DNA origami. *Nature*, **525**, 82–86.
41. Griesang,N., Giessler,K., Lommel,T. and Richert,C. (2006) Four color, enzyme-free interrogation of DNA sequences with chemically activated, 3′-fluorophore-labeled nucleotides. *Angew. Chem. Int. Ed.*, **45**, 6144–6148.
42. Gießler,K., Griesser,H., Göhringer,D., Sabirov,T. and Richert,C. (2010) Synthesis of 3′-BODIPY-labeled active esters of nucleotides and a chemical primer extension assay on beads. *Eur. J. Org. Chem.*, **19**, 3611–3620.
43. Hänle,E and Richert,C. (2018) Enzyme-free replication with two or four bases. *Angew. Chem. Int. Ed.*, **57**, 8911–8915.
44. Steffen,C., Thomas,K., Huniar,U., Hellweg,A., Rubner,O. and Schroer,A. (2010) TmoleX—a graphical user interface for TURBOMOLE. *J. Comput. Chem.*, **31**, 2967–2970.