# Hydrophilic modifications in peptide nucleic acid — Synthesis and properties of PNA possessing 5-hydroxymethyluracil and 5-hydroxymethylcytosine[1]

**Robert H.E. Hudson, Yuhong Liu, and Filip Wojciechowski**

**Abstract:** We have investigated the chemistry for the incorporation of C5-hydroxymethyl-uracil and -cytosine in peptide nucleic acid (PNA) and the subsequent effect of this modification on PNA hybridization behavior. Largely based on literature precedent, we prepared a peptide nucleic acid monomer, possessing 5-hydroxymethyuracil, which was compatible with Fmoc-based oligopeptide synthesis. An improved, large-scale synthesis of 5-hydroxymethylcytosine was developed, as a starting point for the synthesis of a monomer containing this nucleobase. In each case, the hydroxyl group was blocked as a *t*-butyldiphenylsilyl ether, and the exocyclic amino group of cytosine was additionally blocked with the benzoyl-group. The modified monomers were incorporated into isolated positions in the oligomer sequence using standard protocols. The modified oligomers showed that the 5-hydroxymethyl group is compatible with triplex and duplex formation.

*Key words:* peptide nucleic acid, hydroxymethyluracil, hydroxymethylcytosine, modified nucleobase, hybridization.

**Résumé :** On a étudié la chimie de l'incorporation d'un C5-hydroxyméthyl-uracile et d'une -cytosine dans un acide nucléique peptidique (ANP) ainsi que l'effet de cette modification sur le comportement d'hybridisation de l'ANP. En se basant principalement sur les précédents trouvés dans la littérature, on a préparé un acide nucléique peptidique monomère possédant un groupe 5-hydroxyméthyluracile qui était compatible avec une synthèse d'oligopeptide basée sur l'utilisation d'un groupe Fmoc. On a développé une synthèse à grande échelle améliorée de la 5-hydroxyméthylcytosine qui a été utilisée comme point de départ pour la synthèse d'un monomère contenant cette nucléobase. Dans chaque cas, le groupe hydroxyle a été bloqué par un éther *t*-butyldiphénylsilyle et on a de plus bloqué le groupe amino exocyclique de la cytosine à l'aide d'un groupe benzoyle. On a incorporé les monomères modifiés dans des positions isolées de la séquence de l'oligomère en faisant appel aux protocoles habituels. Les oligomères modifiés ont permis de montrer que le groupe 5-hydroxyméthyle est compatible avec la formation d'un triplex et d'un duplex.

*Mots-clés :* acide nucléique peptidique, hydroxyméthyluracile, hydroxyméthylcytosine, nucléobase modifiée, hybridisation.

[Traduit par la Rédaction]

## Introduction

Ready access to synthetic oligonucleotides has enabled much of the tremendous progress witnessed in molecular biology, forensics, diagnostics, and related areas over the past couple of decades. Some applications of oligonucleotides, especially in hostile environments containing DNA or RNA degrading enzymes, require structural modifications to resist degradation. Out of endeavors to make robust oligonucleotide analogs, peptide nucleic acid (PNA) was born (1). PNA is an oligonucleotide mimic based on a polyamide repeat unit. It exhibits very tight and specific binding to complementary nucleic acids and continues to be of interest for both in vitro diagnostic applications and in vivo therapeutic use (2).

Although the original design of PNA "works" quite well (see Fig. 1), possessing excellent hybridization affinity and selectivity, often-cited limitations are its poor water solubility, poor ability to remain in solution and a tendency to be adsorbed onto surfaces, sequence-dependent aggregation, and a general difficulty in penetrating cellular membranes. These biophysical properties are related in various degrees to the non-ionic structure of PNA. Judicious chemical modifications can improve these biophysical properties while maintaining the highly desirable hybridization properties. For example, the aqueous solubility of PNA is improved by the attachment of a cationic amino acid, such as lysine, to one or both ends of the oligomer (3). Alternatively, introduction of charged (guanidinium) groups along the backbone structure improves the aqueous solubility and also facilitates cellular-membrane penetration (4). Another strategy to improve the solubility of PNA is to introduce hydrophilic nucleobase replacements (5).

The modular nature of PNA is particularly amenable to structural variation, and this has led to the synthesis of a wide variety of modified PNAs. We have chosen to investi-

Received 22 December 2006. Accepted 5 February 2007. Published on the NRC Research Press Web site at canjchem.nrc.ca on 5 May 2007.

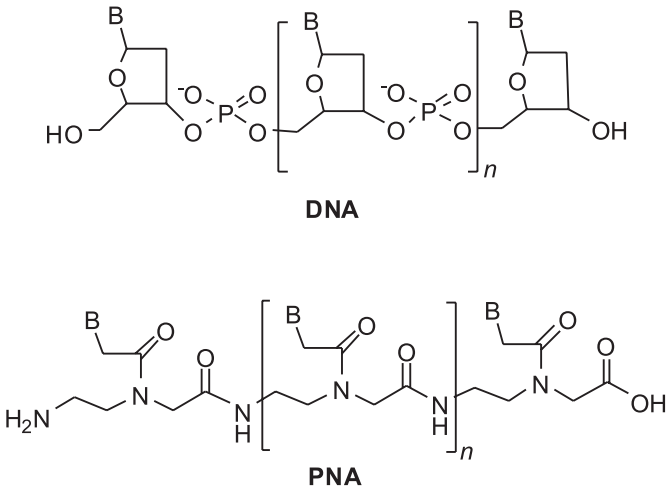
**R.H.E. Hudson,[2] Y. Liu, and F. Wojciechowski.** Department of Chemistry, The University of Western Ontario, London, ON N6A 5B7, Canada.

[1]This article is part of a Special Issue dedicated to Professor Kelvin K. Ogilvie.

[2]Corresponding author (e-mail: rhhudson@uwo.ca).

 doi:10.1139/V07-030 

**Fig. 1.** Structures of DNA and PNA.




gate nucleobase modifications as a way to introduce additional functionality, which may be distributed throughout the PNA oligomer without changing the *N*-2-aminoethylglycine backbone structure. Our approach, as reported herein, was to make a conservative structural modification that added hydrophilic character, which was synthetically accessible and would be sterically tolerated in a duplex or triplex structure. The 5-hydroxymethyl modification of pyrimidines (Fig. 2) fulfilled these requirements.

The 5-hydroxymethyl modification constitutes a relatively minor structural change to the nucleobases, representing the replacement of a hydrogen on 5-methyluracil (thymine) or 5-methylcytosine by a hydroxyl. The resulting 5-substituent is placed in the major groove and is expected to be sterically well tolerated (6). Other advantages are that the chemistry is accessible and starts with inexpensive reagents, and substitution does not produce a new stereocenter and thereby avoids problems of enantiomeric purity in the product.

## Experimental

### General

All reagents and solvents were obtained from common commercial sources and used without further purification, unless indicated. Anhydrous solvents were prepared by passing the reagent-grade solvent through columns of activated alumina (Innovative Technologies). Chromatography was carried out using silica gel 60 (particle size 0.063–0.200 mm; Merck). TLC was carried out on plastic sheets, silica gel 60 $F_{254}$, layer thickness of 0.2 mm (Merck). NMR spectra were recorded on Mercury 400 (400.1 MHz for $^1$H; 100.6 MHz for $^{13}$C) and Inova 400 spectrometers. Chemical shift values, reported in ppm, were referenced by the spectrometer to the residual proton in the deuterated solvents used to record the spectra. Resonances, attributed to conformers due to restricted rotation around amide bonds (i.e., rotamers), are reported as major (ma) and minor (mi) in the $^1$H and $^{13}$C NMR spectra. Assignments are shown for selected resonances that have been unambiguously identified by the use of 2D NMR correlation experiments. Multiplicity of signals are indicated where appropriate (s: single, d: doublet, t: triplet, q: quartet, m: mulitplet, br: broad, v: very). Melting points were determined on a Gallenkamp melting-point apparatus and are reported uncorrected.

### 5-Hydroxymethyluracil (1)

The synthesis of 5-hmU followed our previously reported method (7). Specifically, triethylamine (42 mL, 300 mmol) was added to a suspension of uracil (22.4 g, 200 mmol) and paraformaldehyde (18.0 g, 600 mmol) in water (600 mL). The mixture turned into a clear solution after heating to 60 °C. The solution was stirred at 60 °C overnight after which time the water was removed under vacuum until the volume reached 60 mL. To this residue, 60 mL of ethanol (95%) was added, and the resultant solution was left to stand at room temperature for 30 min. Once a precipitate formed, the mixture was placed in the refrigerator at 4 °C for several hours, a white solid was collected by filtration, and was washed with cold ethanol. The filtrate was cooled to –20 °C, and a second crop of **1** was collected. Further crops were yielded by reducing the volume of the filtrate by one half and diluting with an equal volume of 95% ethanol, followed by cooling to –20 °C. White solid; 26.1 g (92%); mp (from ethanol) 360 °C dec (lit. 300 °C dec.) (13). TLC ($CHCl_3$–$CH_3OH$, 4:1): $R_f$ 0.16. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 11.04 (s, 1H, NH), 10.69 (d, $J$ = 5.20 Hz, 1H, NH), 7.22 (d, $J$ = 5.20 Hz, 1H, H6), 4.85 (t, $J$ = 5.60 Hz, 1H, OH), 4.09 (d, $J$ = 4.80 Hz, 2H, –$CH_2$–). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 163.6, 151.2, 138.0, 112.6, 55.8. HR-MS (EI) calcd. for $C_5H_6O_3N_2$: 142.0378; found: 142.0376.

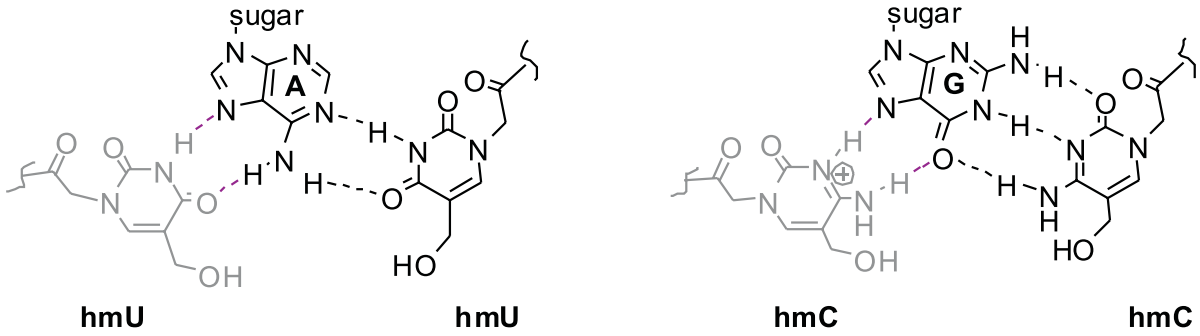
### Ethyl(5-hydroxymethyluracil-1-yl)acetate (2)

To a suspension of 5-hydroxymethyluracil (22.4 g, 158 mmol) and $K_2CO_3$ (21.8 g, 158 mmol) in DMF (475 mL) was added ethyl bromoacetate (98%, 28.3 g, 166 mmol). The reaction mixture was stirred overnight after which time the $K_2CO_3$ was removed by filtration, and the DMF was removed under vacuum, thus yielding a white solid. The solid was washed with $H_2O$ (180 mL) by stirring the slurry for approximately 2 h. The first crop of **2** (21.4 g) was obtained by filtration and drying. Further crops were recovered from the filtrate by reducing its volume. White solid; 82% (29.5 g); mp 178–180 °C. TLC ($CHCl_3$ – $CH_3OH$, 4:1): $R_f$ 0.46. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 11.42 (s, NH), 7.53 (s, 1H, H6), 5.07 (t, $J$ = 5.60 Hz, 1H, OH), 4.54 (s, 2H, –$CH_2$–), 4.13 (m, 4H, overlapping –$OCH_2$–), 1.21 (t, $J$ = 7.20 Hz, 3H, –$CH_3$). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 168.1, 163.0, 150.7, 141.7, 113.4, 61.2, 55.9, 48.8, 14.2. HR-MS (EI) calcd. for $C_9H_{12}O_5N_2$: 228.0746; found: 228.0748.

### Ethyl[5-(*t*-butyldiphenylsilanyloxymethyl)uracil-1-yl]acetate (3)

To a solution of ethyl(5-hydroxymethyluracil-1-yl)acetate (5.01 g, 22 mmol) and imidazole (3.29 g, 48 mmol) in DMF (110 mL) was added *tert*-butyldiphenylchlorosilane (TBDPS-Cl, 98%, 6.64 g, 23.7 mmol) with stirring. The solution was stirred overnight after which time the reaction was stopped by adding 5% $NaHCO_3$ in $H_2O$ (11 mL). The solvents were removed under vacuum to give a white solid.

**Fig. 2.** (*a*) Expected base-pairing of adenine (A, DNA strand) with 5-hydroxymethyluracil-containing PNA in duplex formation (black) and triplex formation (black and grey). (*b*) Expected base-pairing between guanine (G, DNA strand) with 5-hydroxymethylcytosine-containing PNA in duplex formation (black) and triplex formation (black and grey).




This material was dissolved in dichloromethane (250 mL), washed with $H_2O$ (100 mL × 3), and the organic phase was separated and dried over sodium sulfate. The crude product was obtained after removing the dicholormethane by vacuum and was subsequently purified by trituration with hexanes to remove the silanol byproduct (TBDPS-OH; $R_f$ 0.96; ethyl acetate–DCM, 1:2). The yield was 99% (10.2 g); mp 168–170 °C. TLC (ethyl acetate – DCM, 1:2): $R_f$ 0.71. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 11.52 (s, NH), 7.65–7.67 (m, 4H, Ar–H), 7.62 (s, 1H, C6–H), 7.43–7.48 (m, 6H, Ar–H), 4.59 (s, 2H, N–$CH_2$–), 4.36 (s, 2H, O–$CH_2$Ar), 4.16 (q, $J$ = 7.20 Hz, 2H, O–$CH_2$–), 1.21 (t, $J$ = 7.20 Hz, 3H, C–$CH_3$), 1.00 (s, 9H, SiC–$CH_3$). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 168.3 (C7), 163.0 (C2), 150.8 (C1), 142.6 (C4), 135.2 (C13), 132.7 (C12), 130.0 (C15), 128.0 (C14), 112.0 (C3), 61.3 (C8) 58.9 (C5), 48.9 (C6), 26.7 (C11), 18.9 (C10), 14.1 (C9). HR-MS (EI) calcd. for $C_{25}H_{30}O_5N_2Si$: $[M + H]^+$, 467.2002; found: 467.1994.

### [5-(*t*-butyldiphenylsilanyloxymethyl)uracil-1-yl]acetic acid (4)

To a solution of ethyl[5-(*t*-butyldiphenylsilanyloxymethyl)uracil-1-yl]acetate (10.6 g, 22.7 mmol) in THF (204 mL) and $H_2O$ (22 mL) at 0 °C was added 2.5 mol/L NaOH (47 mL, 6 equiv.). The mixture was stirred at 0 °C for 20 min, then 1 mol/L HCl (170 mL) was added to stop the reaction. The aqueous phase was extracted with ethyl acetate four times (400 mL × 1, 100 mL × 3). The ethyl acetate phase was washed with $H_2O$ (100 mL × 2) and then dried over sodium sulfate. The solvent was removed under vacuum to yield the title compound: white solid; 8.2 g, 96%; mp 169–170 °C. TLC (ethyl acetate – DCM, 1:2): $R_f$ 0.71. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 13.14 (br s, 1H, COOH), 11.47 (s, 1H, NH), 7.65–7.67 (m, 4H, Ar), 7.62 (s, 1H, C6–H), 7.41–7.50 (m, 6H, Ar), 4.49 (s, 2H, N–$CH_2$–), 4.35 (s, 2H, O–$CH_2$–), 1.00 (s, 9H, –$CH_3$). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 169.8, 163.0, 150.9, 142.8, 135.2, 132.8, 130.0, 128.0, 111.8, 58.9, 48.8, 26.7, 18.9. MS (ESI) calcd. for $[C_{23}H_{26}O_5N_2Si + Na]^+$: 461.15; found: 461.1.

### Methyl(*N*-(2-((9H-fluoren-9-yl)methoxy)carbonyl)aminoethylglycinate hydrochloride (5)

To a suspension of methyl 2-aminoethylglycinate dihydrochloride (6.0 g, 20.8 mmol) and Fmoc-*N*-hydroxysuccinimide (10.8 g, 1.1 equiv.) in dioxane (100 mL) was added aq. $NaHCO_3$ (3 equiv., 7.4 g/120 mL) with rapid stirring for 50 min. The reaction mixture was diluted with 400 mL $H_2O$, which produced an instant white precipitate that redissolved on the addition of DCM (350 mL). This aqueous phase was extracted with DCM (2 × 300 mL), and then the DCM phase was dried over sodium sulfate, acidified with $Et_2O$ – HCl (5 mol/L), and cooled in the freezer overnight. The desired product was collected by filtration: white solid; 7.0 g, 61%; mp 110–113 °C. Further product can be obtained by reducing the volume of the filtrate. TLC ($CH_3COOEt$ – $CH_3OH$, 95:5; visualized with UV and 0.3% ninhydrin (purple)): $R_f$ 0.24. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 9.44 (br s, 2H, $H_2N^+$), 7.90 (d, $J$ = 7.20 Hz, 2H, Ar), 7.70 (d, $J$ = 7.20 Hz, 2H, Ar), 7.58 (t, $J$ = 5.60 Hz, CONH), 7.42 (t, $J$ = 7.20 Hz, 2H, Ar), 7.33 (t, $J$ = 7.60 Hz, 2H, Ar), 4.33 (d, $J$ = 6.80 Hz, 2H, –$CH_2$–O), 4.23 (t, $J$ = 6.80 Hz, 1H, $Ar_2$–CH–), 4.00 (s, 2H, N–$CH_2$–), 3.74 (s, 3H, O–$CH_3$), 3.33 (overlapping with water, –$CH_2$–$CH_2$–), 3.02 (t, $J$ = 6.00 Hz, 1.8H, ma, –$CH_2$–$CH_2$–), 2.90 (mi, 0.2H). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 167.1, 156.3, 143.9, 140.8, 127.7, 127.2, 125.3, 120.2, 65.7, 52.7, 46.7, 46.6, 46.5, 36.6. MS (EI) calcd. for $[C_{20}H_{22}O_4N_2HCl – HCl]^+$: 354.16; found: 354.1.

Byproducts identified as dibenzofulvene: $R_f$ 0.96 (ethyl acetate – methanol, 95:5), and 2-oxo-3-*N*-(Fmoc)piperazine: $R_f$ 0.42 ($CHCl_3$ – $CH_3OH$ – $CH_3COOH$, 90:8:2). Mp 172–174 °C.

### Methyl(2-(((9H-fluoren-9-yl)methoxy)carbonyl)aminoethyl)glycinate (6)

The hydrochloride salt of the title compound (3.0 g, 7.8 mmol) was suspended in dichloromethane (350 mL) and shaken against a saturated aq. $NaHCO_3$ solution (150 mL × 3) to liberate the free base. The phases were separated, and the combined aqueous phases were back-extracted with dichloromethane (100 mL × 3). The organic phase was dried over sodium sulfate, and the solvent was removed under vacuum to yield the free base as a white wax-like product: 2.70 g (99%); mp 76–78 °C. TLC (ethyl acetate – methanol, 95:5; visualized by UV and 0.3% ninhydrin (purple)): $R_f$ 0.30. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 7.76 (d, $J$ = 7.60 Hz, 2H, Ar), 7.61 (d, $J$ = 7.60 Hz, 2H, Ar), 7.40 (t, $J$ = 7.60 Hz, 2H, Ar), 7.32 (t, $J$ = 7.20 Hz, 2H, Ar), 5.36 (br, ma, CONH), 5.12 (br, mi, CONH), 4.40 (d, $J$ = 7.20 Hz, 2H, –$CH_2$–O), 4.22 (t, $J$ = 6.80 Hz, 1H, $Ar_2$–CH–), 3.73 (s, 3H,

O–$CH_3$), 3.42 (s, 2H, N–$CH_2$–C(O)), 3.27 (q, $J$ = 5.60 Hz, ma, CONH–$CH_2$–), 3.17 (mi, CONH–$CH_2$–), 2.75 (t, $J$ = 5.60 Hz, ma, –$CH_2$–N), 2.62 (mi, –$CH_2$–N). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 173.1, 156.7, 144.2, 141.5, 127.8, 127.2, 125.2, 120.1, 40.8, 66.7, 52.0, 50.4, 48.8, 47.4. HR-MS (EI) calcd. for $C_{20}H_{22}O_4N_2$: 354.1580; found: 354.1581.

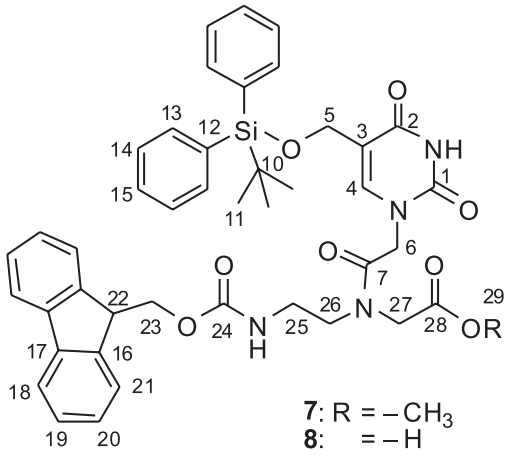
**Fmoc hmU-OMe (7)**

To a suspension of methyl(2-(((9H-fluoren-9-yl)methoxy)carbonyl)aminoethyl)glycinate (**6**) (1.61 g, 4.5 mmol) and the nucleobase acetic acid derivative **4** (3.18 g, 7.3 mmol) in DMF (45 mL) was added 1-ethyl-3-(3′-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 1.39 g, 7.3 mmol) in two portions within 0.5 h. The mixture was stirred for 16 h at room temperature, and then the reaction mixture was diluted with dichloromethane (1000 mL). This solution was washed with 0.05 mol/L HCl (200 mL × 4), saturated $NaHCO_3$ (200 mL × 3), and $H_2O$ (450 mL × 7), respectively. The organic layer was separated and dried over $Na_2SO_4$, and the solvent was removed under vacuum at room temperature. The crude product (white foam) was purified by gradient column chromatography (DCM, ethyl acetate, ethyl acetate – methanol (95:5), respectively) to yield the title compound as a white foam (Fig. 3): 3.10 g, 88% yield. TLC (ethyl acetate – methanol, 95:5; visualized by UV and 0.3% ninhydrin (purple)): $R_f$ 0.63. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 7.19–7.77 (m, overapping 19H, Ar), 6.06 (br t, ma, backbone NH) and 5.45 (br t, mi, backbone NH), 4.51–4.54 (br s, overlap H5 and its rotamer), 4.51–4.54 (br s, ma, H6) and 4.37–4.44 (overlap, mi, H6), 4.37–4.44 (overlap, H23 and its rotamer), 4.17–4.20 (m, H22), 4.17–4.20 (mi, H27) and 4.07 (s, ma, H27),[3] 3.74 (s, ma, H29) and 3.78 (s, mi, H29), 3.57 (br s, ma, 1.6 H, H26) and 2.88 (br s, mi, 0.4 H, H26), 3.45 (br m, ma, H25) and 3.38 (br m, mi, H25), 1.08 (s, ma, H11) and 1.09 (s, mi, H11). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 170.1 (C28), 167.3 (C7), 162.7 (C2), 157.0 (C24), 151.2 (C1), 144.0 (C16), 141.4 (C17), 141.2 (C4), 135.7 (C13), 133.0 (C12), 130.0 (C15), 128.0 (C14), 127.8 (C19), 127.2 (C20), 125.2 (C21), 120.1 (C18), 114.2 (C3), 66.9 (C23), 59.1 (C5), 53.0 (r), 52.6 (C29), 50.5 (r), 49.0 (C27), 48.9 (C26), 48.6 (r), 48.5 (C6), 47.3 (C22), 39.2 (r), 39.6 (C25), 27.1 (C11), 19.4 (C10). HR-MS (ESI-TOF) calcd. for $[C_{43}H_{46}O_8N_4Si + Na]^+$: 797.2983; found: 797.3022.

**Fmoc-hmU-OH (8)**

The ester precursor to the PNA monomer acid (2.93 g, 3.78 mmol) was dissolved in THF (32 mL), and then $H_2O$ (21 mL) was added to produce a turbid mixture. The mixture was cooled to 0 °C, and 2.5 mol/L NaOH (12 mL, 8 equiv.) was added. The mixture was stirred for 5 min at 0 °C during which time the mixture cleared to produce a homogeneous solution. The reaction was stopped by addition to a separatory funnel containing a mixture of ethyl acetate (400 mL) and 0.3 mol/L HCl (200 mL). The acidic aqueous phase was extracted by ethyl acetate (150 mL × 3). The ethyl acetate washes were combined (~850 mL), washed with $H_2O$ (100 mL × 3), and dried over $Na_2SO_4$. The ethyl acetate was removed under vacuum at room temperature to yield an off-white (yellowish) foam (Fig. 3). The product was purified by column chromatography using a gradient of solvents (DCM; DCM–methanol (90:10), (85:15), and (78:22); respectively) to give the title compound as an off-white solid: 2.60 g, 90%; mp 216–21s7 °C. TLC ($CHCl_3$ – $CH_3OH$ – $CH_3COOH$, 90:8:2; visualized with UV and 0.3% ninhydrin (grey)): $R_f$ 0.08. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 11.39 (br s, uracil NH) and 11.36 (br s, uracil NH), 7.29–7.88 (overlapping signals, 18H Ar, backbone NH), 4.75 (s, mi, 0.2H, H6) and 4.56 (s, ma, 1.8H, H6), 4.23–4.35 (m, overlap, 2H, H5) and (r, H23), 4.17–4.20 (m, overlap, H23, H22), 3.95 (s, 2H, H27),[3] 3.44 (br m, mi, H26) and 3.38 (br m, ma, H26), 3.36 (br m, mi, H25) and 3.16 (br m, ma, H25), 0.98 (s, 9H, H11). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 171.5 (C28), 166.9 (C7), 163.0 (C2), 156.2 (C24), 150.9 (C1), 143.9 (C16), 143.3 (C4), 140.7 (C17), 135.2 (C13), 132.8 (C12), 129.9 (C15), 128.0 (C14), 127.6 (C19), 127.1 (C20), 125.3 (C21), 120.1 (C18), 111.4 (C3), 65.6 (C23), 58.9 (C5), 51.4 (r), 48.5 (C27), 48.1 (C6), 47.6 (r), 46.9 (C26), 46.7 (C22), 38.7 (r), 38.0 (C25), 26.7 (C11), 18.8 (C10). HR-MS (ESI-TOF) calcd. for $[C_{42}H_{44}O_8N_4Si + Na]^+$: 783.28261; found: 783.2841.

**Fig. 3.** Structure of the 5-hydroxymethyluracil (hmU) PNA monomer ester ($R$ = –$CH_3$) and acid ($R$ = –H), defining the arbitrary numbering system used in NMR signal assignments.




**5-Hydroxymethylcytosine (9)**

Triethylamine (16.5 mL, 118 mmol) was added to a suspension of cytosine (5.27 g, 47.4 mmol) and paraformaldehyde (3.28 g, 109 mmol) in water (140 mL). The mixture was stirred and heated under reflux for 72 h. After 24 h and 32 h of reaction, a second and third portion of paraformaldehyde (1.14 g, 38.0 mmol) and (0.57 g, 19.0 mmol) was added, respectively. After completion of the reaction, the volatile components (water and triethylamine) were removed under vacuum to give a dark-coloured residue (12.8 g), which represented greater than 100% theoretical yield. This residue was dissolved to produce a clear solution by heating with 35 mL of 95% ethanol and 18–24 mL $H_2O$, depending on solubility. The solution was left at room tem-

[3] Variable position (3.85 – 4.20 ppm), depending on the protonation state of the carboxylic acid.

perature for a short time during which a precipitate formed. Further precipitation was induced by cooling to 4 °C. The first crop of 5-hydroxymethylcytosine, 4.2 g (5-hmC, yellow solid, purity ≈ 90% (wt%), 7%–8% (wt%) cytosine, as determined by $^1$H NMR spectroscopy) was collected by filtration and washed three times with 20 mL of 95% ethanol. This first crop represents 57% chemical yield of **9** from cytosine. Reduction in the volume of the filtrate gave subsequent crops of 5-hmC. The later crops contained proportionally more cytosine and other unidentified compounds. Nonetheless, these impure fractions of 5-hmC were suitable for further use. The total yield of 5-hmC calculated from all recovered fractions, based on determining the purity by $^1$H NMR spectroscopy, was 85%. This experiment was repeated many times and proved to be reproducible. Mp 300 °C dec. TLC (*t*-butyl alcohol – methyl ethyl ketone – $H_2O$ – $NH_4OH$, 4:3:2:1): $R_f$ 0.32, compared with $R_f$ 0.46 for cytosine. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 10.60 (v br, ring NH), 7.28 (s, 1H, C6–H), 7.10 (br, $NH_2$), 6.54 (br, $NH_2$), 4.99 (v br, OH), 4.16 (s, 2H, O–$CH_2$–Ar). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 165.6, 157.0, 140.8, 104.9, 57.1. HR-MS (EI) calcd. for $[C_5H_7O_2N_3 + H]^+$: 142.0616; found; 142.0620.

**Ethyl(5-hydroxymethylcytosin-1-yl)acetate (10)**

To a suspension of crude 5-hydroxymethylcytosine (**9**) (8.97 g, 68% (wt%), 43 mmol) and $K_2CO_3$ (5.97 g, 43.2 mmol) in dry DMF (140 mL), was added ethyl bromoacetate (5.6 mL, 98% pure, 47.5 mmol). The reaction mixture was stirred overnight, and a dark-coloured oil was obtained after removal of the $K_2CO_3$ by filtration through a pad of Celite and evaporation of DMF under vacuum. Ethyl acetate (120 mL) was added to triturate the residue, which was sonicated (often 1–2 h, sometimes up to 5 h), and then a brown solid was isolated by filtration. The solid crude product was washed with a small amount of ethyl acetate (50 mL × 3). All of the mass was accounted for in the crude material, which was found to contain the desired product (**10**, 60% (wt%)) along with the dialkylated byproduct (19% (wt%)), some unreacted starting material **9** (~20% (wt%); $R_f$ 0.06; DCM – $CH_3OH$, 4:1), and small amounts of unidentified compounds, as estimated by examination of $^1$H NMR spectra. The crude material was suitable for the next reaction. The title compound **10** could also be purified by solvent gradient silica-gel column chromatography (DCM; DCM – $CH_3OH$ (90:10), (85:15), and (80:20); respectively). Mp 175–177 °C, as isolated from chromatography. TLC (DCM – $CH_3OH$, 4:1): $R_f$ 0.25. $^1$H NMR (399.8 MHz, DMSO-$d_6$, ppm) δ: 7.54 (s, 1H, C6–H), 7.33 (br s, $NH_2$), 6.62 (br s, $NH_2$), 5.11 (t, $J$ = 5.20 Hz, OH), 4.45 (s, 2H, –$CH_2$–N), 4.17 (d, $J$ = 5.20 Hz, 2H, O–$CH_2$–Ar), 4.11 (q, $J$ = 7.20 Hz, 2H, O–$CH_2$–), 1.19 (t, $J$ = 7.20 Hz, 3H, –$CH_3$). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 168.8, 165.0, 155.7, 144.2, 105.7, 60.9, 57.1, 50.0, 14.1. HR-MS (EI) calcd. for $[C_9H_{13}O_4N_3]^+$: 227.0906; found: 227.0905.

Identified byproduct: *N*1,*N*4-bisalkylated nucleobase: ethyl(*N*4-(2-ethoxy-2-oxoethyl)-5-(hydroxymethylcytosin-1yl)acetate. TLC (DCM – $CH_3OH$, 4:1): $R_f$ 0.52. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 7.88 (s, 1H, C6–H), 5.44 (t, $J$ = 5.2, –OH), 4.72 (s, 2H, N1–$CH_2$–), 4.32 (d, $J$ = 5.3, 2H, O–$CH_2$–Ar), 4.29 (s, 2H, N4–$CH_2$–), 4.16 (q, $J$ = 7.1, 4H, 2 × –$CH_2$–), 1.22 (t, $J$ = 7.2, 6H, 2 × –$CH_3$). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 185.2, 170.9, 167.9, 145.9, 144.6, 108.8, 61.6, 61.5, 56.1, 49.7, 49.0, 14.1.

**Ethyl[5-(*t*-butyldiphenylsilanyloxymethyl)cytosin-1-yl]acetate (11)**

Crude ethyl(5-hydroxymethylcytosin-1-yl)acetate (**10**), which contained 5-hmC, was used as the starting material for this transformation. As a byproduct, 5-(*t*-butyldiphenylsilanyloxymethyl)cytosine was obtained, which could be alkylated by ethyl bromoacetate to give the title compound, as a supplementary route to the following one.

As an example, a particularly impure sample was utilized. To a solution of crude ethyl(5-hydroxymethylcytosin-1-yl)acetate (**10**) (5.08 g, 48% (wt%), ~11 mmol) and imidazole (3.66 g, 54 mmol) in DMF (115 mL) was added TBDPS-Cl (8 mL, 27.0 mmol) with stirring. The solution was stirred overnight, and then a saturated solution of $NaHCO_3$ (10 mL) was added to stop the reaction. The DMF and $H_2O$ were removed under vacuum to yield a dark-brown oil, which was dissolved in ethyl acetate (300 mL) and washed with $H_2O$ (150 mL × 4). Unreacted starting materials (**9** and **10**), as well as dialkylated hmC, were removed by the aqueous washes. Compound **9** was recovered by back-extraction with ethyl acetate, the efficacy of which was monitored by TLC. The ethyl acetate phase was dried over $Na_2SO_4$, and the solvent was removed under vacuum to yield compound **9** along with silylated 5-hmC. The title compound **11** was isolated as a slightly orange solid, 40% yield, by flash column chromatography, using a gradient of solvents (hexane, ethyl acetate, and ethyl acetate – methanol (96:4) for **9**; ethyl acetate – methanol (75:25) for 5-(*t*-butyldiphenylsilanyloxymethyl)cytosine). Mp 183–185 °C, as isolated from chromatography. TLC (ethyl acetate – methanol, 95:5) $R_f$ 0.18. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 7.64 (d, $J$ = 8.00 Hz, 4H, Ar), 7.38–7.48 (m, 6H, Ar), 6.48 (s, 1H, C6–H), 4.42 (s, 2H, SiO–$CH_2$–Ar), 4.31 (s, 2H, –$CH_2$–N), 4.19 (q, $J$ = 6.40 Hz, 2H, –O–$CH_2$–), 1.26 (t, $J$ = 6.40 Hz, 3H, –$CH_3$), 1.04 (s, 9H, SiC–$CH_3$). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 143.6, 168.2, 166.0, 156.2, 135.8, 132.7, 130.3, 128.1, 105.0, 62.0, 61.4, 50.2, 27.0, 19.3, 14.3. HR-MS (EI) calcd. for $[C_{25}H_{31}O_4N_3Si + H]^+$: 466.2162; found: 466.2155.

Identified byproduct: 5-(*t*-butyldiphenylsilanyloxymethyl)cytosine; white solid. TLC (ethyl acetate – methanol, 95:5): $R_f$ 0. $^1$H NMR (400.1 MHz, DMSO-$d_6$, ppm) δ: 7.63 (d, $J$ = 6.00 Hz, 4H, Ar), 7.42–7.50 (m, 6H, Ar), 7.10 (s, 1H, C6–H), 4.46 (s, 2H, SiO–$CH_2$–Ar), 1.00 (s, 9H, SiC–$CH_3$). MS (TOF-MS, ES$^+$) calcd. for $[C_{21}H_{25}O_2N_3Si + H]^+$: 380.1794; found: 380.1794.

**Ethyl[4-benzoyl-5-(*t*-butyldiphenylsilanyloxymethyl)cytosin-1-yl]acetate (12)**

To a solution of ethyl[5-(*t*-butyldiphenylsilanyloxymethyl)cytosin-1-yl]acetate (3.42 g, 7.3 mmol) in anhyd. pyridine (35 mL) was added benzoyl chloride (1 mL, 8.0 mmol) with stirring. The solution was stirred overnight, ethyl acetate (300 mL) was added, and the solution was washed with 1 mol/L HCl (200 mL × 2) to remove most of

the pyridine.[4] The ethyl acetate phase was subsequently washed with 0.02 mol/L HCl (200 mL × 3) and then water (200 mL × 3) to remove traces pyridine. The acidic aqueous phase was back-extracted by ethyl acetate (200 mL). The ethyl acetate phases were combined and dried over $Na_2SO_4$. The crude product had the appearance of an orange oil after removal of the solvent. Column chromatography (hexane, hexane–EtOAc (95:5) and (92:8) for **12**; EtOAc – $CH_3OH$ (95:5) to recover **11**; $R_f$ 0.04 (EtOAc – DCM, 1:6) was used to purify the crude material. White solid; 92% (4.16 g, after column-chromatography purification); mp 130–132 °C. TLC (EtOAc – DCM, 1:6): $R_f$ 0.76. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 13.19 (s, 1H, NH), 8.03 (d, $J$ = 7.20 Hz, 2H, Ar), 7.63 (d, $J$ = 6.40 Hz, 4H, Ar), 7.28–7.40 (m, 10H, Ar and C6–H), 4.70 (s, 2H, O–$CH_2$–Ar), 4.43 (s, 2H, N–$CH_2$–), 4.20 (q, $J$ = 7.20 Hz, 2H, O–$CH_2$–), 1.24 (t, $J$ = 7.20 Hz, 3H, –$CH_3$), 1.04 (s, 9H, –Si–$CCH_3$). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 179.6, 167.2, 158.2, 148.5, 141.3, 136.8, 135.6, 133.0, 132.6, 130.2, 130.0, 128.2, 128.1, 115.2, 62.4, 59.4, 49.9, 27.0, 19.5, 14.3. HR-MS (EI) calcd. for $[C_{32}H_{35}O_5N_3Si]^+$: 569.2346; found: 569.2352.

**[4-*N*-Benzoyl-5-(*t*-butyldiphenylsilanyloxymethyl)cytosin-1-yl]acetic acid (13)**

To a solution of **12** (2.78 g, 4.88 mmol) in THF (55 mL) and $H_2O$ (14 mL) at 4 °C was added 2.5 mol/L NaOH (16 mL, 8 equiv.). The mixture was stirred at 0 °C for 15 min, and the reaction was stopped by decanting into a separation funnel containing ethyl acetate (280 mL) and 0.25 mol/L HCl (190 mL) in which the reaction mixture was acidified, and **13** was obtained. The ethyl acetate phase was washed with $H_2O$ (140 mL × 2). The aqueous phase was extracted with ethyl acetate (280 mL × 2). The ethyl acetate phases were combined and washed with $H_2O$ (400 mL), dried over $Na_2SO_4$, and the solvent was removed under vacuum to yield the title compound as a white foam: 2.74 g (99% yield). TLC (ethyl acetate – DCM, 1:6; same eluent used for **12**): $R_f$ 0. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 8.06 (d, $J$ = 7.20 Hz, 2H, Ar), 7.68 (m, 4H, Ar), 7.34–7.49 (m, 10H, Ar and C6–H), 4.73 (s, 2H, O–$CH_2$–Ar), 4.54 (s, 2H, –$CH_2$–N), 1.09 (s, 9H, SiC–$CH_3$). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 177.7, 170.8, 158.7, 149.6, 142.8, 136.2, 135.7, 132.9, 132.8, 130.2, 129.8, 128.4, 128.1, 114.5, 59.6, 50.5, 27.1, 19.5. HR-MS (ESI-TOF, positive ion mode) calcd. for $[C_{30}H_{31}O_5N_3Si + H]^+$: 542.2111; found: 542.2091.

**Fmoc hmC-OMe (14)**

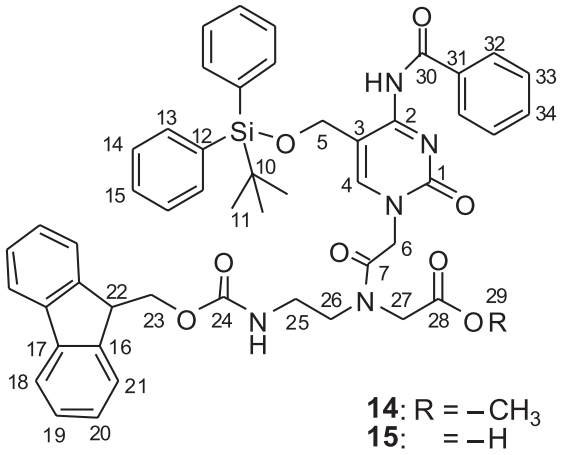
Compound **13** (3.83 g, 7.1 mmol) and hydroxybenzotriazole (HOBt, 1.03 g, 7.6 mmol) were dissolved in anhyd. DMF (39 mL). The solution was cooled to 0 °C. After adding *N*,*N*′-dicyclohexylcarbodiiimide (DCC, 1.58 g, 7.6 mmol) in one portion, the mixture was taken off the ice and stirred at room temperature for 1 h during which time a thick white precipitate of dicyclohexylurea (DCU) was formed. Methyl(2-(((9H-fluoren-9-yl)methoxy)carbonyl)aminoethyl)glycinate (2.52 g, 7.1 mmol) was added, and the mixture was stirred overnight. The precipitated DCU was filtered out by passing the solution through a pad of Celite. The filtrate was diluted with DCM (450 mL) and washed with acidic water twice ($H_2O$, 250 mL; 1 mol/L HCl, 6 mL) to remove the unreacted backbone submonomer, then washed with $NaHCO_3$ (satd. 250 mL × 2) and $H_2O$ (300 mL × 3), respectively. All aqueous phases were back-extracted with DCM. The combined organic phases were dried over $Na_2SO_4$, and the solvent was removed under vacuum (RT). The crude product (yellow oil) was purified by column chromatography (hexane, hexane – ethyl acetate (80:20), (60:40), and (10:90) for the title compound, whereas ethyl acetate – methanol (98:2) and (95:5) was used to recover the backbone). White solid; 3.61 g, (58%); mp 125–126.5 °C (softens at 92 °C). TLC (ethyl acetate – methanol, 95:5; visualized with UV and 0.3% ninhydrin (green)): $R_f$ 0.62. $^1$H NMR (400.1 MHz, $CDCl_3$, ppm) δ: 13.26 (v br, NH), 8.13 (d, $J$ = 7.20 Hz, 2H, H32), 7.29–7.74 (m, 22H, aromatic protons), 6.06 (br t, ma, backbone NH) and 5.48 (br t, $J$ = 4.80 Hz, mi, backbone NH), 4.77 (s, 2H, H5), 4.64 (s, ma, H6) and 4.52 (s, mi, H6), 4.47 (d, $J$ = 6.80 Hz, ma, H23) and 4.40 (d, $J$ = 7.20 Hz, mi, H23), 4.22–4.25 (m, 1H, H22), 4.09 (s, ma, H27) and 4.22–4.25 (m, mi, H27), 3.73 (s, ma, H29) and 3.79 (mi H29), 3.73 (br, ma, H26) and 2.95 (br, mi H26), 3.46 (br, ma, H25) and 3.41 (br, mi H25), 1.15 (s, mi, H11) and 1.13 (s, ma, H11). $^{13}$C NMR (100.6 MHz, $CDCl_3$, ppm) δ: 179.3 (C30), 169.9 (C28), 166.9 (C7), 158.5 (C2), 156.9 (C24), 148.5 (C1), 143.8 (C16), 142.8 (C4), 141.2 (C17), 136.7 (C31), 135.5 (C13), 133.0 (C12), 132.4 (C34), 130.0 (C15), 129.8 (C32), 128.1 (C33), 127.9 (C14), 127.7 (C19), 127.1 (C20), 125.1 (C21), 119.9 (C18), 114.2 (C3), 66.7 (C23), 59.3 (C5), 52.4 (ma, C29) and 52.9 (mi, C29), 48.9 (ma, C6) and 49.1 (mi, C6), 48.8 (ma, C27) and 50.3 (mi, C27), 48.7 (C26), 47.2 (C22), 39.0 (mi, C25) and 39.4 (ma, C25), 26.9 (C11), 19.3 (C10). HR-MS (ESI-TOF) calcd. for $[C_{50}H_{51}O_8N_5Si + Na]^+$: 900.3405; found: 900.3430.

**Fmoc-hmC-OH (15)**

Compound **14** (3.40 g, 3.9 mmol) (Fig. 4) was dissolved in THF (42 mL) and $H_2O$ (11 mL). The solution was cooled to 4 °C in an ice-bath, and then 2.5 mol/L NaOH (12.1 mL, 8 equiv.) was added. The solution was stirred for 5 min at 4 °C, and then the reaction was stopped by decanting it into a separation funnel containing ethyl acetate (500 mL) and acidified water (36 mL of 1 mol/L HCl and 200 mL of $H_2O$). The ethyl acetate phase was washed with $H_2O$ (200 mL × 2). The aqueous phase was extracted with ethyl acetate (150 mL × 2). The organic phase was separated and dried over $Na_2SO_4$. The solvent was removed under vacuum (RT). The crude product (off-white solid, 3.7 g) was purified by flash column chromatography (hexane, hexane – ethyl acetate (50:50), ethyl acetate, and ethyl acetate – methanol (95:5), (80:20), and (75:25)). White solid; 2.95 g (90%); mp 186–187 °C, as isolated from chromatography (softens at 130 °C). TLC (EtOAc – $CH_3OH$, 95:5; visualized with UV and 0.3% ninhydrin (green)): $R_f$ 0.04. $^1$H NMR (400.1 MHz,

[4] Caution: care must be taken in the amount of HCl used, which should be calculated based on complete neutralization of pyridine; excess acid removed the *N*4-Bz group.

**Fig. 4.** Structure of the 5-hydroxymethylcytosine (hmC) PNA monomer ester ($R$ = –$CH_3$) and acid ($R$ = –H), defining the arbitrary numbering system used in NMR signal assignments.




DMSO-$d_6$, ppm) δ: 7.96 (d, $J$ = 7.60 Hz, 2H, H32), 7.31–7.89 (22H, aromatic protons), 7.28–7.55 (overlap, 1H, backbone NH), 4.92 (s, mi, H6) and 4.72 (s, ma, H6), 4.60 (s, mi, H5) and 4.58 (s, ma, H5), 4.34 (d, $J$ = 6.80 Hz, mi, H23) and 4.24 (d, $J$ = 6.40 Hz, ma, H23), 4.19–4.20 (m, 1H, H22), 3.99–4.01 (br, depending on the state of protonation, 2H, H27), 3.47 (b, ma, H26) and 3.30 (br m, mi, H26), 3.39 (br m, ma, H25) and 3.17 (br m, mi, H25), 0.97 (s, 9H, H11). $^{13}$C NMR (100.6 MHz, DMSO-$d_6$, ppm) δ: 172.1 (C30), 171.5 (C28), 167.3 (C7), 158.7 (C2), 156.6 (C24), 148.7 (C1), 146.3 (C4), 144.0 (C16), 140.8 (C17), 136.3 (C31), 135.3 (C13), 132.8 (C12), 132.5 (C34), 130.0 (C32), 129.3 (C15), 128.2 (C33), 128.0 (C14), 127.7 (C19), 127.2 (C20), 125.3 (C21), 120.1 (C18), 111.9 (C3), 65.7(C23), 59.2(C5), 51.1 (mi, C27) and 49.4 (ma, C27), 48.6 (C6), 47.7 (mi, C26) and 47.1 (ma, C26), 46.9 (C22), 38.2 (C25), 26.7 (C11), 18.9 (C10). HR-MS (ESI-TOF) calcd. for $[C_{49}H_{49}O_8N_5Si - H]^+$: 862.32722; found: 862.3262**.**

### Synthesis of PNA oligomers

PNA oligomers were synthesized on a 5 μmol scale using the standard FastMoc module on an Applied Biosystems 433A synthesizer employing a NovaSyn TGR (Rink amide) resin. The hexameric PNA (**P1**) was synthesized using the Rink amide resin with a loading of 0.23 mmol/g. Oligomers (**P2**–**P5**) were prepared on a Rink amide resin preloaded with lysine (approximate loading: 0.065 mmol/g).

### Resin cleavage and deprotection

To the resin bound oligomer (**P1**–**P5**) was added an ice-cold solution of TFA (950 μL) and triethylsilane (25 μL). The resin was agitated for 2 h. After filtration, TFA was removed with a stream of nitrogen, and the oligomer was precipitated, as a white solid, with cold diethyl ether (15 mL). Under these conditions, the TBDPS protecting group is also removed.

Oligomer **P4**, which contains an hmC residue, was additionally treated with concentrated aq. ammonia (1 mL, 55 °C for 12 h). The solution was dried under reduced pressure to give an off-white solid.

### Oligomer purification

The crude, solid oligomers were dissolved in aq. TFA (0.1%), filtered, and purified by RP-HPLC, using a 250 mm × 4.6 mm, C18-bonded phase, 300 Å pore, and 5 μm particle-size column.

### Characterization of PNAs

Mass spectral analysis of oligomers was performed at the UWO Biological Mass Spectrometry Laboratory. Electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) was performed with mass calibration provided by myoglobin ($M^+$ = 16951.49 Da, average mass) or histatin 5 ($M^+$ = 3036.33 Da, monoisotopic mass) (see Table 1).

### Binding analysis

Poly(riboadenylic acid) (poly(rA)) was purchased from Sigma (Milwaukee, WI). Mixed sequence oligonucleotides were synthesized and purified by standard methods or purchased from the University Core DNA Services (Calgary, AB). Thermal denaturation of PNA:NA complexes (~2 μmol/L in nucleobases) was measured at 260 nm with a temperature ramp of 0.5 °C/min under ionic conditions (150 mmol/L NaCl, 10 mmol/L $Na_2HPO_4$, and 0.1 mmol/L EDTA at pH 7.0). A minimum of two independent samples were run many times, and the $T_m$ was estimated by determining the maximum in the first derivative of the denaturation curve. The error associated with the $T_m$ values is ±0.5 °C.

## Results and discussion

The goal of this work was to achieve the synthesis of 5-hydroxymethyl-derivatized uracil and cytosine PNA monomers, incorporate these modified bases into PNA, and study their impact on hybridization. As a starting point, we were interested in a scalable and straightforward approach to the synthesis of the monomers. Key considerations were on the preparation of the 5-hydromethyluracil and -cytosine and subsequently the use of an appropriate protecting-group strategy for the hydroxyl functional group.

### 5-Hydroxymethyluracil-containing PNA

The synthesis of 5-hydroxymethyluracil monomer (hmU), suitable for Fmoc-based peptide chemistry, is outlined in Scheme 1. We took the approach of preparing the appropriate (pyrimidin-1-yl)acetic acid ester derivative, which would be condensed with the methyl[*N*-2-(Fmoc-amino)ethyl]glycinate "backbone" submonomer in a convergent fashion, which we have previously reported (8). The synthesis starts with the inexpensive uracil and a one-step hydroxymethylation. Despite the potential biological importance of 5-hydroxymethyluracil and nucleoside derivatives thereof (9–12),[5]

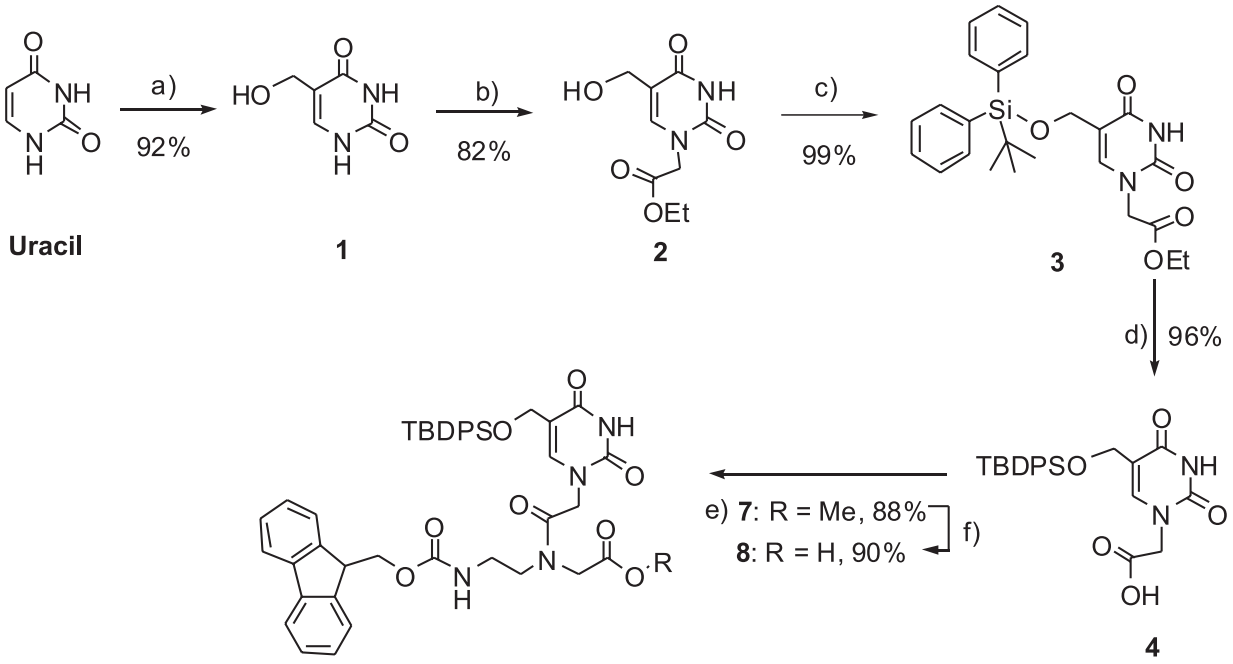
[5] 5-hmU is a component of bacteriophage DNA, as well, there exists hmU specific binding proteins. The DNA of the dinoflagellate *Gyrodinium cohnii* contains 5-hmU. Glycosylated hmU is found in *Typanosoma Brucei*, the causative organism of African sleeping sickness. 5hmU is caused by oxidative damage to DNA

**Table 1.** Mass spectral characterization of PNA oligomers.

| Oligomer | Sequence | Found | Calculated |
|---|---|---|---|
| **P1** | Ac-(hmU)T(hmU)T(hmU)T-Lys-$NH_2$ | 1832.50 | 1832.76 |
| **P2** | H-Gly-GTA GAT CCC T-Lys-$NH_2$ | 2887.86 | 2887.80 |
| **P3** | H-Gly-GTA GAT CTC T-Lys-$NH_2$ | 2902.78 | 2902.82 |
| **P4** | H-Gly-GTA GAT C(hmC)C T-Lys- $NH_2$ | 2917.68 | 2917.83 |
| **P5** | H-Gly-GTA GAT C(hmU)C T-Lys- $NH_2$ | 2918.93 | 2918.82 |

**Note:** Abbreviations in use: 5-hydroxymethyluracil (hmU), 5-hydroxymethylcytosine (hmC).

**Scheme 1.** Reagents and conditions: (*a*) $(HCHO)_n$, 0.5 mol/L $Et_3N$, 60 °C, 16 h; (*b*) $BrCH_2CO_2Et$, $K_2CO_3$, DMF, RT, 16 h; (*c*) TBDPS-Cl, imidazole, DMF, RT, 16 h; (*d*) 0.43 mol/L NaOH, 0 °C, 20 min; (*e*) methyl[*N*-2-(Fmoc-amino)ethyl]glycinate (**6**), EDC, DMF, RT, 16 h; (*f*) (*i*) 0.46 mol/L NaOH, 0 °C, 5 min; (*ii*) $H^+$.



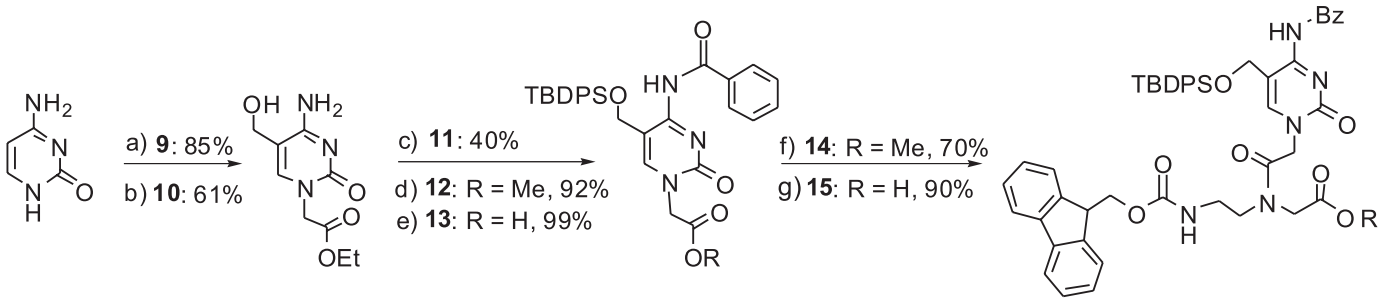

there exist but a few synthetic routes to the modified nucleobase. Hydroxymethyluracil has been prepared by the reaction of uracil with paraformaldehyde in water, using an inorganic base (13, 14). However, we wished to avoid the potential problem of separating the nucleobase from inorganic salts and were inspired by the report of triethylamine-catalyzed hydroxymethylation of the deoxynucleoside (15). The reaction of paraformaldehyde with uracil in aq. $NEt_3$ (0.5 mol/L) proceeds smoothly overnight, and the product conveniently crystallizes from the reaction solvent.

As with uracil, thymine, and related derivatives, we next proceeded to the alkylation of the nucleobase with ethyl bromoacetate in anhyd. DMF with $K_2CO_3$ (16), suspecting that we would achieve the desired chemoselectivity, based on the pKa differences between the 5-hydroxyl group and the lactam/imide groups of the heterocycle. We observed complete selectivity for N-alkylation and recovered the desired N1-derivative in good yield. The only byproduct that we were able to identify from this reaction was the *N*1,*N*3-bisalkylated nucleobase.

Next, we chose a silyl ether as the masking group for the 5-hydroxy functionality for ease of installation and removal as well as orthogonality to the Fmoc- group and the methyl ester of the backbone. Initially, we prepared the *t*-butyldimethylsilyl ether, but this proved not robust enough in model studies; thus, we turned to the *t*-butyldiphenylsilyl ether. This group was stable to alkyl ester hydrolysis (0.43 mol/L NaOH, THF – $H_2O$, 0 °C, >40 min), Fmoc-removal (30% piperidine in DMF, 40 min), and normal workup conditions (pH 1–12), yet was completely and rapidly removed by treatment with TFA. After settling on a hydroxyl-protecting group, standard transformations yielded the hmU monomer **8** with some noteworthy points. The condensation of the backbone submonomer **6** with the modified nucleobases **4** was mediated by the use of the carbodiimide reagent EDC without the use of additional base or additives. As well, the best yield and least elimination of the Fmoc-group were obtained when the acid component was used in excess (1.6 equiv.) relative to the backbone submonomer free base. In this instance, the isolated yields of **7** were reliably high (88%) with a minimal loss of the base-labile Fmoc-group (<4%).

The penultimate step, the saponification of the monomer ester (**7**), was also carefully done to avoid loss of the Fmoc-

**Scheme 2.** Reagents and conditions: (*a*) $(HCHO)_n$, 0.75 mol/L $Et_3N$, $H_2O$, 3 d, 95 °C; (*b*) $K_2CO_3$, DMF, $BrCH_2CO_2Et$, 24 h; (*c*) TBDPS-Cl, imidazole, DMF, 16 h; (*d*) BzCl, pyridine; (*e*) 0.46 mol/L NaOH, 15 min, 4 °C; (*f*) **6**, DCC, HOBT, DMF; (*g*) (*i*) 0.46 mol/L NaOH, 0 °C, 5 min; (*ii*) $H^+$.




group. The reaction conditions (8 equiv. of NaOH, 0.46 mol/L, 5 min) were optimized to give complete hydrolysis, while minimizing concomitant removal of the Fmoc-group. Nonetheless, some Fmoc- groups were lost (4%–8%), yet the final monomer was recovered in 90% yield from column chromatography. The optimized conditions for the condensation and hydrolysis reactions were then used for the synthesis of the hmC monomer **15**.

## 5-Hydroxymethylcytosine-containing PNA

Although 5-hydroxymethylcytosine has been known for some time in the context of a modified base found in the DNA of the *Escherichia coli* T-even bacteriophages (17) and in the nucleoside antibiotic mildiomycin (18), there has been no report of its synthesis from cytosine, via hydroxymethylation, until quite recently. Aside from isolation from these natural sources, early syntheses are based on multistep de novo heterocycle formation (19).

Our initial attempts to produce hmC under similar conditions of reaction to those used for the synthesis of hmU were unproductive. During this period of time, Abdel-Rahman and El Ashry reported a small-scale (1 mmol) hydroxymethylation of cytosine with paraformaldehyde in 0.5 N KOH solution under microwave irradiation (20). Our attempts to repeat this synthesis failed. However the report of the reaction of 2-deoxycytosine monophosphate with paraformaldehyde in the presence of 5 mol/L KOH (75 °C, 4 d) (12% yield) (21) motivated our further efforts on the direct hydroxymethylation of cytosine. Although the reaction of cytosine in aq. KOH with paraformaldehyde failed miserably, hmC was formed in good chemical yield (~85%) when triethylamine was employed along with higher reaction temperature and longer reaction time. We found that a molar excess of paraformaldehyde was necessary for the reaction to proceed in good chemical yields, but the addition must be made portion-wise during the course of the reaction. For instance, addition of greater than 3 equiv. of paraformaldehyde at the beginning of the reaction causes the formation of undesired and unidentified byproducts and made the recovery of hmC by recrystallization unfeasible. The reaction completely fails to give the desired product at 10 equiv. of paraformaldehyde. To overcome these problems, the amount of paraformaldehyde was carefully controlled and added in several portions during the reaction ($(HCHO)_n$ – cytosine (mol) 3.5:1), Scheme 2. The hmC is recovered from the reaction solution by crystallization. Although later crops are significantly contaminated with unreacted cytosine (up to approximately 20% (wt%)), they are still suitable for use in the next step.

The next necessary transformations are installation of nucleobase-protecting groups and alkylation. For cytosine, either of two routes may be used: N4-protection followed by N1-alkylation in the presence of either $K_2CO_3$ or NaH (22, 23); or the reverse, N1-alkylation in the presence of NaH and then protection of the exocyclic amino group (24). Both routes were attempted and gave identical yields; thus, the alkylation was simply performed in DMF with $K_2CO_3$, as was done with 5-hydroxymethyluracil.

Following alkylation, the nuclebase was protected, first by silylation of the hydroxyl group and then benzoylation of the exocyclic amino group. The silylation step gave a poor yield, likely by interference by the adjacent amino group, which could not be improved even by use of an excess TBDPS-Cl (2 equiv.). The remaining transformation to yield the hmC PNA monomer **15** proceeded without incident.

## Synthesis and hybridization properties of PNA containing C5-hydroxymethylpyrimidines

The hmU and hmC PNA monomers were used for both automated and manual PNA-oligomer synthesis. Initially, standard conditions were employed (resin loading 0.23 mmol/g), and this was suitable for the incorporation of hmU in isolated positions (**P1**) by automated methods. However, upon attempts to prepare contiguously modified oligomers by manual methods, the synthesis failed after 5 or 6 residues. Although these oligomers were identified by mass spectrometry, they were inseparable, and further sequence elongation was not possible. Therefore, we turned to lower-loading resin (0.06 mmol/g), which did not resolve this problem. However, the lower-loading resin was subsequently used to prepare singly modified oligomers **P2**–**P4**, which allowed us to evaluate the effect of the hmU or hmC substitution for T or C, respectively.

The oligomers were examined for their ability to bind complementary nucleic acids (Table 2). The poly(hmU/T) hexamer was used for triplex formation with poly(rA), wherein two strands of PNA bind to the RNA strand, thus giving each complex a total of six modifications. Compared with the unmodified oligomer, there was a slight drop in $T_m$ (–1.5 °C) per insert. Interestingly, the same modest $T_m$ drop

**Table 2.** Summary of $T_m$ values for complexes formed between PNA and complementary DNA or RNA.

| Oligomer | Sequence | Complement | $T_m$ (°C) | $\Delta T_m$/mod[d] (°C) |
|---|---|---|---|---|
| **P0** | Ac-TTT TTT-K-$NH_2$[a] | poly(rA) | 62.0 | — |
| **P1** | Ac- (hmU)T(hmU)T(hmU)T-Lys-$NH_2$ | poly(rA) | 53.0 | –1.5 |
| **P2** | H-Gly-GTA GAT CCC T-Lys-$NH_2$ | DNA[b] | 50.7 | — |
| **P3** | H-Gly-GTA GAT CTC T-Lys-NH | DNA[c] | 47.3 | — |
| **P4** | H-Gly-GTA GAT C(hmC)C T-Lys- $NH_2$ | DNA[b] | 49.3 | –1.4 |
| **P5** | H-Gly-GTA GAT C(hmU)C T-Lys- $NH_2$ | DNA[c] | 46.0 | –1.3 |

[a]From ref. 7.
[b]Complementary deoxyoligonucleotide sequence: 5′-AGA GAT CTA C-3′.
[c]Complementary deoxyoligonucleotide sequence: 5′-AGG GAT CTA C-3′.
[d]$\Delta T_m$/mod: the difference in the melting temperature of modified oligomer compared with the unmodified oligomer on a per-substitution basis.

was observed in the singly modified duplexes (**P4** and **P5**), regardless of the identity of the nucleobase. These investigations represent the first study of the effect of 5-hydroxymethylpyrimidines on complex stability in either DNA or analogs. Previous studies have focused on the glycosylated derivatives (25), wherein only the diaminoglycoside conjugate was stabilizing towards duplex formation (26, 27). Relevant studies to 5-hydroxymethylcytosine have been focused on the interaction between oligonucleotides bearing this modification and enzymes and have not included $T_m$ measurements (28, 29).

In summary, we have described a convenient, large-scale synthesis of 5-hydroxymethylcytosine and 5-hydroxymethyluracil and the transformation of these modified nucleobases in monomers compatible with PNA synthesis. The modifications were placed in isolated positions in sequences capable of either triplex formation or duplex formation. Dense functionalization of PNA oligomers was not achieved, likely because of the hydrophobic/steric nature of the protected hmU and hmC monomers. A single modification had a small negative impact on the stability of the complex, independent of the particular nucleobase, and multiple substitutions had an additive effect. Overall, the substitution of a hmU or hmC for T or C, respectively, is tolerated in both the context of a duplex or a triplex, within the sequences examined. Further investigation on the optimization of the protecting-group strategy and oligomerization chemistry is needed to fully evaluate this modification and its impact on the biophysical properties of PNA.

## Acknowledgements

The authors thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Academic Development Fund at The University of Western Ontario for support of this work. Prof. G. Lajoie and Ms. Paula Pittock of the Biological Mass Spectrometry Laboratories (UWO) are thanked for their expert assistance with mass spectral characterization of the polyamides. Mr. Doug Hairsine, Department of Chemistry, is thanked for performing HR-MS of small molecules. Mr. David Dodd is acknowledged for performing polyacrylamide gel electrophoresis of PNA (data not shown).

## References

1. P.E. Nielsen, M. Egholm, R.H. Berg, and O. Buchardt. Science (Washington, D.C.), **254**, 1497 (1991).
2. P.E. Nielsen, U. Koppelhus, and F. Beck. *In* Pseudo-peptides in drug discovery. *Edited by* P.E. Nielsen. Wiley-VCH, Weinheim, Germany. 2004. Chap. 4.
3. F. Beck. *In* Methods in molecular biology: Peptide nucleic acids, methods and protocols. Vol. 208. *Edited by* P.E. Nielsen. Humana Press Inc., Totowa, NJ, USA. 2002. Chap. 2.
4. P. Zhou, A. Dragulescu-Andrasi, B. Bhattacharya, H. O'Keefe, P. Vatta, J.J. Hyldig-Nielsen, and D.H. Ly. Bioorg. Med. Chem. Lett. **16**, 4931 (2006).
5. B.D. Gildea, S. Casey, J. MacNeil, H. Perry-O'Keefe, D. Sørensen, and J.M Coull. Tetrahedron Lett. **39**, 7255 (1998).
6. G.R. Revandkar and T.S. Rao. *In* Comprehensive natural products: DNA and aspects of molecular biology. Vol. 7. *Edited by* E.T. Kool. Elsevier Sciences Ltd., Oxford, UK. 2002. Chap. 2.
7. R.H.E. Hudson and F. Wojciechowski. Can. J. Chem. **83**, 1731 (2005).
8. R.H.E. Hudson, M. Goncharenko, A. Wallman, and F. Wojciechowski. Syn. Lett. **9**, 1442, (2005).
9. (*a*) G.G. Johnson and E.P. Geiduschek. Biochemistry, **16**, 1473 (1977); (*b*) J.R. Greene and E.P. Geiduschek. EMBO J. **4**, 1345 (1985).
10. P.M.M. Rae. Proc. Natl. Acad. Sci. U.S.A. **70**, 1141 (1973).
11. F. VanLeeuwen, M.C. Taylor, A. Mondragon, H. Moreau, W. Gibson, R. Kieft, and P. Borst. Proc. Natl. Acad. Sci. U.S.A. **95**, 2366, (1998).
12. M. Berger, J. Cadet, and J. Ulrich. Can. J. Chem. **63**, 6, (1985).
13. R.E. Cline, R.M. Fink, and K. Fink. J. Am. Chem. Soc. **81**, 2521 (1959).
14. U. Chiacchio, A. Corsaro, J. Mates, P. Merino, A. Piperno, A. Rescifina, G. Romeo, R. Giovanni, and T.T. Roberto. Tetrahedron, **59**, 4733, (2003).
15. C.J. LaFrancois, J. Fujimoto, and L.C. Sowers. Chem. Res. Toxicol. **11**, 75, (1998).
16. R.H.E. Hudson and J.M. Moszynski. Synlett, **18**, 2997, (2006).
17. G.R. Wyatt and S.S. Cohen. Nature (London), **170**, 1072, (1952).
18. S. Harada and T. Kishi. J. Antibiot. (Tokyo), **31**, 519, (1978).
19. T.L.V Ulbricht and C.C. Price. J. Am. Chem. Soc. **21**, 567, (1956), and refs. cited therein.
20. A.A.-H. Abdel-Rahman and E.-S.H. El Ashry. Synlett, **12**, 2043, (2002).
21. A.H. Alegria. Biochim. Biophys. Acta, **149**, 317, (1967).

22. K.L. Dueholm, M. Egholm, C. Behrens, L. Christensen, H.F. Hansen, T. Vulpius, K.H. Petersen, R.H. Berg, P.E. Nielsen, and O. Buchardt. J. Org. Chem. **59**, 5767, (1994).
23. Z. Timar, L. Kovacs, G. Kovacs, and Z. Schmel. J. Chem. Soc., Perkin Trans. 1, 19, (2000).
24. L. Christensen, H.F. Hansen, T. Koch, and P.E. Nielsen. Nucleic Acids Res. **26**, 2735, (1998).
25. No $T_m$ data reported: J.J. Turner, N.J. Meeuwenoord, A. Rood, P. Borst, G.A. van der Marcel, and J.H. van Boom. Eur. J. Org. Chem. 3832, (2003).
26. J. Hunziker. Bioorg. Med. Chem. Lett. **9**, 201, (1999).
27. R. Tona, R. Bertolini, and J. Hunziker. Org. Lett. **2**, 1693, (2000).
28. S. Tardy-Planechaud, J. Fujimoto, S.S. Lin, and L.C. Sowers. Nucleic Acids Res. **25**, 553, (1997).
29. M. de Kort, P.C. de Visser, J. Kurzeck, N.J. Meeuwenoord, G.A. van der Marcel, W. Rüger, and J.H. van Boom. Eur. J. Org. Chem. 2075, (2001).