Available online at www.sciencedirect.com
SCIENCE DIRECT®
Tetrahedron Letters 45 (2004) 6243–6247
Tetrahedron Letters

Efficient synthesis of extended guanine analogues designed for recognition of an A·T inverted base pair in triple helix based-strategy

Nathalie Van Craynest, Dominique Guianvarc'h,† Corinne Peyron and Rachid Benhida*

Laboratoire de Chimie Bioorganique UMR-CNRS 6001, Université de Nice-Sophia Antipolis, Parc Valrose, 06108 Nice Cedex 2, France

Received 15 June 2004; accepted 22 June 2004

Abstract—We report herein an efficient synthesis of new nucleosides **N1**, **N2**, and **N3** as extended guanine analogues, derived from aminobenzimidazole and thymine or 5-substituted uracil. These nucleosides were devised for the recognition of an A·T inverted base pair by three hydrogen bonds, in triple helix-based technology.

It is currently accepted that pyrimidine oligonucleotides can bind in parallel orientation to a double stranded DNA (ds-DNA) to form stable local triple-helical structures. This specific recognition process occurs in the major groove of ds-DNA by formation of T·A×T and C·G×C$^+$ triplets (i.e., pyrimidine motif) through specific Hoogsteen hydrogen bonds (Fig. 1).[1] Accordingly, triplex forming oligonucleotides (TFOs) have attracted a considerable interest because of their potential use in genes recognition-based technologies.[2] Unfortunately, the application of TFO approach in therapeutic and biotechnology has a major intrinsic limitation since triplex formation is restricted to oligopyrimidine·oligopurine tracts. Thus, a single purine·pyrimidine (A·T or G·C) interruption in the oligopyrimidine·oligopurine ds-DNA is sufficient to induce high triplex destabilization.[3]

Many efforts have been undertaken to extend the recognition pattern of TFOs to polypurine sequences, which are interrupted by a pyrimidine.[4] However, the most reported works in this area were addressed for the recognition of G·C rather than A·T. Indeed, the recognition of A·T is highly hampered by the bulky methyl group of thymine due to steric hindrance and hence, only a few systems were devised for this purpose.

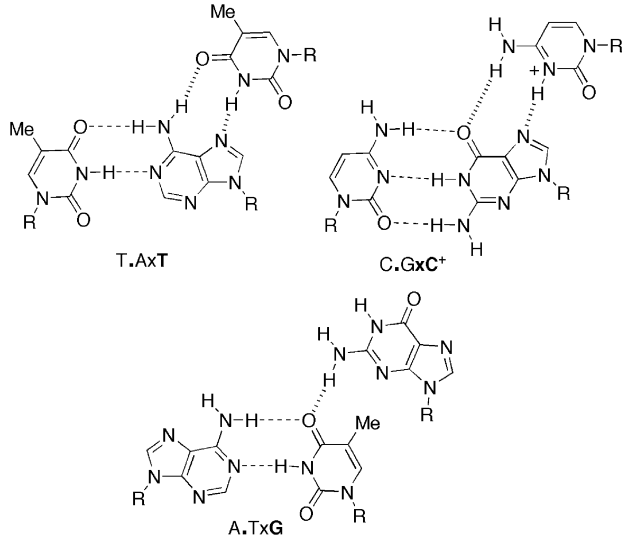


Figure 1. Hydrogen bonding within T·A×T, C·G×C$^+$ and A·T×G triplets.

We have previously reported on the recognition properties of several non-natural *C*-nucleosides, designed for

Keywords: Extended nucleosides; Triple helix; Molecular recognition.
* Corresponding author. Tel.: +33-(0)4-9207-6153; fax: +33-(0)4-9207-6151; e-mail: benhida@unice.fr
† Present address: Laboratoire de Chimie Organique Biologique, Université Paris VI, UMR CNRS 7613, 4 Place Jussieu, 75252 Paris cedex 05, France.

0040-4039/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tetlet.2004.06.095

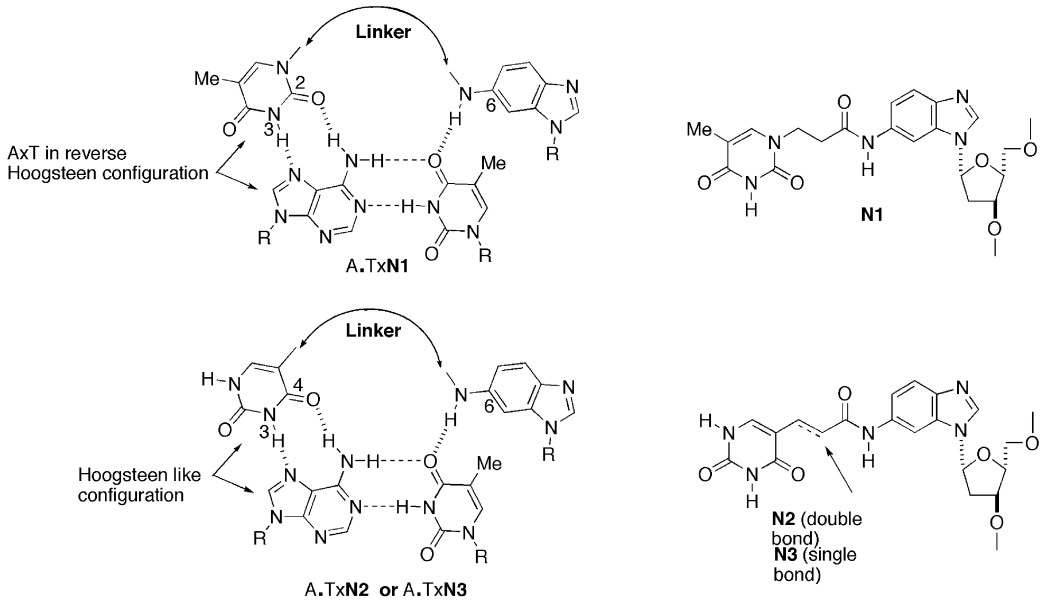
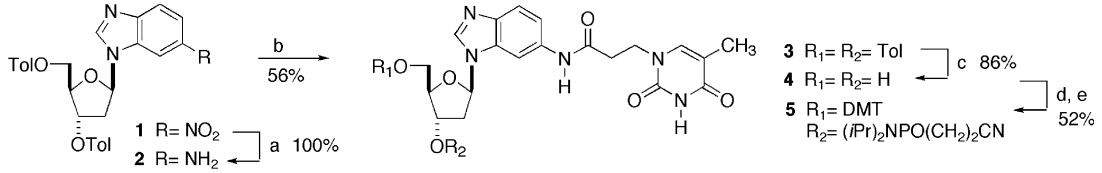


Figure 2. Proposed recognition model within A·T×**N1** and A·T×**N2**(**N3**) triplets.

the A·T inverted base pair recognition on the basis of molecular modeling studies, when incorporated into pyrimidine-motif TFOs.[5] Among these compounds, we found that an analogue containing two unfused aromatic rings (3-aminophenyl-thiazole) linked to a 2′-deoxyribose unit by an acetamide motif (named **S**) had interesting properties. Indeed, it was able to recognize an A·T inversion in ds-DNA with a high affinity, very close to those of canonical triplets (T·A×T and C·G×C^+), and a moderate but significant selectivity for A·T rather than G·C inversion.[5a,c] In continuation of our studies directed at overcoming the sequence limitations in triple helix strategy,[5] we investigated a novel class of modified nucleobases designed for specific recognition of A·T inversion. This study is based on the fact that the guanine is known to recognize a T interruption in oligopurine strand through formation of one hydrogen bond between the *2-amino* group of guanine and the *oxygen O(4)* of thymine (Fig. 1).[6] Moreover, the non-canonical A·T×G triplet was found to be the most stable triplet combination with natural nucleobases. However, this stabilizing effect by guanine is insufficient because it involves only one hydrogen bond. From this fact, a promising way to increase triplex stability at the A·T inverted site should be the use of an extended guanine analogue that can simultaneously and cooperatively bind to all free Hoogsteen sites of both thymine and adenine Watson–Crick base pair with increase of bases stacking.

We report herein the synthesis of novel extended nucleosides **N1**, **N2**, and **N3** designed to contain the complementary donor–acceptor–donor (**dad**) binding sites for the formation of base triplets as shown in Figure 2. In these nucleobases, the thymine was employed to achieve adenine recognition in Hoogsteen (**N2** and **N3**) or reverse Hoogsteen (**N1**) like configuration, and the 6-aminobenzimidazole was chosen as a guanine analogue for *NH(6)* benzimidazole-*O(4)* thymine interaction (Fig. 2). **N2** and **N3** only differ by the nature of their linker (between aminobenzimidazole and thymine), which is conformationally restricted in **N2** and flexible in **N3** (more degrees of freedom).

First, the common starting material 2′-deoxynucleoside **1** was prepared by simple glycosylation of 5(6)-nitrobenzimidazole and flash chromatography separation of the obtained regioisomers.[7] Catalytic hydrogenation of **1** provided quantitatively the amine derivative **2**, which was then coupled with 1-thyminyl-propionic acid[8] using Mukaiyama's reagent (2-chloro-1-methyl-pyridinium iodide)[9] to give the extended nucleoside **3** in 56% non-optimized yield (Scheme 1). The use of standard HOBT/EDC or pentafluorophenol activating agents



Scheme 1. Reagents and conditions: (a) H_2–Pd/C, THF; (b) 1-thyminyl-propionic acid (1.2 equiv), Mukaiyama's reagent (2 equiv), Et_3N (2 equiv), CH_2Cl_2, 50 °C; (c) MeONa (3 equiv), MeOH; (d) DMTCl (3 equiv), Et_3N (3 equiv), DMF; (e) $(CN(CH_2)_2O)(iPr_2N)PCl$ (2 equiv), Hunig's base (2 equiv).

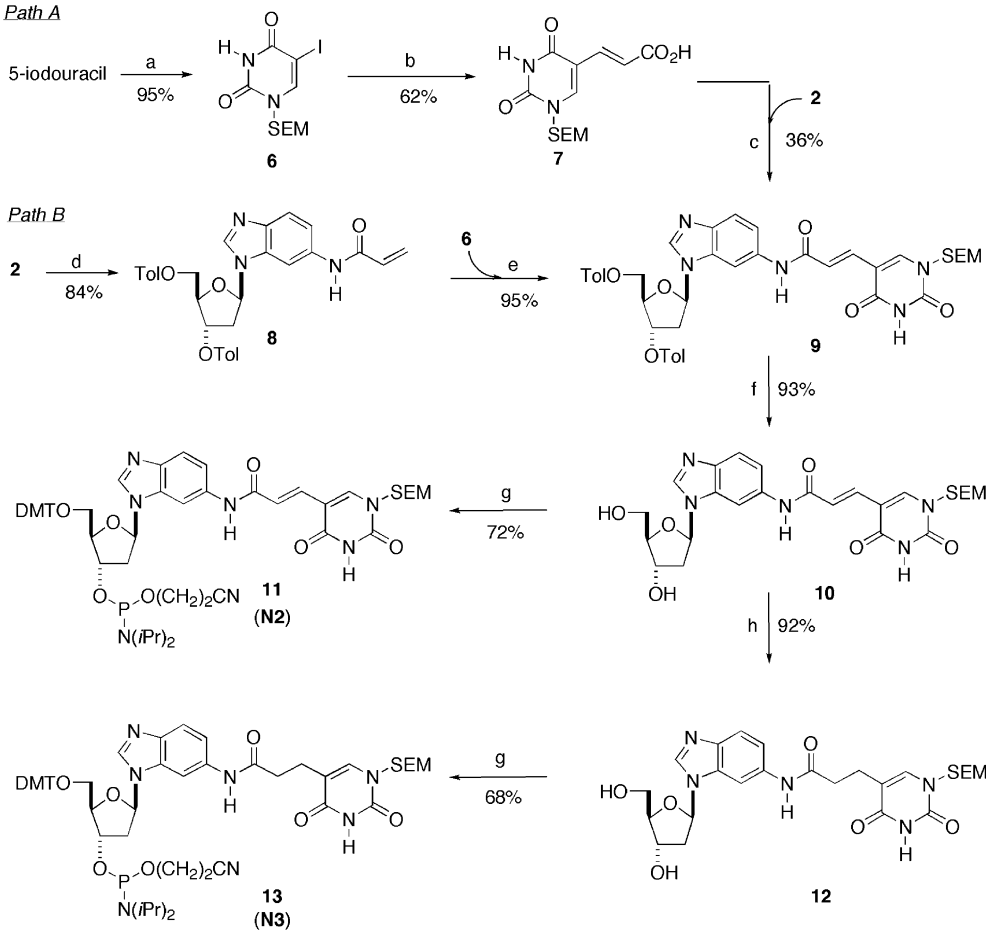
was found to be ineffective in such a reaction. The toluoyl protecting groups were then cleaved (MeONa/MeOH) leading to the free nucleoside **4** in 86% yield. Successive 5′-dimethoxytritylation and 3′-phosphitylation afforded the phosphoramidite derivative **5** in 52% overall yield (Scheme 1).[10]

In the case of nucleosides **N2** and **N3**, their incorporation into oligonucleotides requires the protection of the reactive *N-1* position of pyrimidine base. Hence, we have selected for our synthesis the trimethylsilylethoxymethyl (SEM) as a protecting group. This group seems to be a good candidate because it could be easily introduced and cleaved from oligonucleotides. The synthesis of the targeted phosphoramidites **11** and **13**, relative to **N2** and **N3**, respectively, is outlined in Scheme 2.

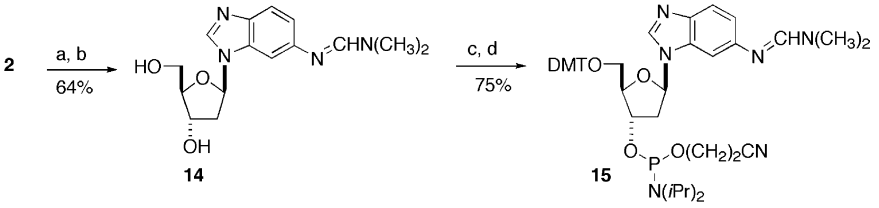
We first tried to obtain **7** following Heck coupling of **6**[11] and methylacrylate then saponification (aqNaOH, dioxane) but, surprisingly, we observed the removal of SEM protecting group during NaOH treatment.[12] To avoid the saponification step, transformation of **6** to **7** was achieved in 62% yield by using acrylic acid in one operation. However, when **7** was coupled with amine **2**, the desired pyrimidine ended nucleoside **9** was obtained in 36% yield (path A). Alternatively **9** was best obtained from **2** by simply inverting the reaction sequences (path B). Thus, treatment of **2** with acryloyl chloride in THF gave **8** which, after Pd coupling with **6** afforded **9** in 80% overall yield. Compound **9** was then quantitatively converted to its free analogue **10** by using NH_3/EtOH solution instead of MeONa/MeOH, which induced, as expected, the cleavage of SEM protecting group. The 5′-dimethoxytritylation and 3′-phosphitylation steps on compound **10** were achieved in 72% overall yield, providing the fully protected phosphoramidite **11**.

In the same way, phosphoramidite building block **13** (i.e., **N3**), a flexible analogue of **N2**, was obtained in good yield from intermediate **10** following (i) catalytic hydrogenation (Pd/C, THF/AcOH) and (ii) successive 5′- and 3′-functionalization (Scheme 2).[13]

In order to point out the implication of the pyrimidine ring of **N1** and **N2** in the proposed recognition process (Fig. 2), we have efficiently synthesized the phosphoramidite guanine analogue **15** containing only one hydrogen donor site (NH_2), as a control. The 6-amino group was



Scheme 2. Reagents and conditions: (a) *N*,*O*-bis(trimethylsilyl)acetamide (BSA) (2.2 equiv), SEMCl (1.2 equiv), $Bu_4N^+I^-$ (10% molar), CH_2Cl_2, 50 °C; (b) acrylic acid (2.5 equiv), $Pd(OAc)_2$ (0.1 equiv), PPh_3 (0.2 equiv), Et_3N (2 equiv), DMF, 100 °C; (c) Mukaiyama's reagent (2 equiv), Et_3N (2 equiv), CH_2Cl_2/DMF; (d) acryloyl chloride (1.5 equiv), Et_3N (2 equiv), CH_2Cl_2; (e) $Pd(OAc)_2$ (0.1 equiv), PPh_3 (0.2 equiv), Et_3N (2 equiv), dioxane, reflux; (f) EtOH, aq NH_3; (g) DMTCl (2 equiv), Et_3N (3 equiv), DMF then $(CN(CH_2)_2O)(iPr_2N)PCl$ (1.2 equiv), Hunig's base (2 equiv), CH_2Cl_2; (h) Pd/C (10% molar), AcOH/THF (1/1, v/v).



Scheme 3. Reagents and conditions: (a) DMA–DMF, 80 °C; (b) MeONa (2.2 equiv), MeOH; (c) DMTCl (2 equiv), Et_3N (3 equiv), DMF; (d) $(CN(CH_2)_2O)(iPr_2N)PCl$ (1.2 equiv), Hunig's base (2 equiv), CH_2Cl_2.

protected, for oligonucleotides synthesis, using dimethylformamide dimethylacetal (DMA) in DMF (Scheme 3).[14]

In summary, we have designed and synthesized new extended nucleosides holding three specific **dad** sites for triple helix mediated recognition of both thymine and adenine base pair in DNA duplexes having a polypurine strand interrupted by one or more thymidine. The incorporation of the efficiently protected phosphoramidites **5**, **11**, **13**, and **15** into TFO for triple helix hybridization studies is in progress.

Acknowledgements

We gratefully acknowledge the CNRS, Région PACA and Université de Nice-Sophia Antipolis for financial support.

References and notes

1. (a) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J. L.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749–7760; (b) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645–650; (c) Praseuth, D.; Guieysse, A. L.; Hélène, C. *Biochim. Biophys. Acta* **1999**, *1489*, 181–206.
2. (a) Casey, B. P.; Glazer, P. M. *Prog. Nucleic Acid Res. Mol. Biol.* **2001**, *67*, 163–192; (b) Faria, M.; Giovannangeli, C. *J. Gene Med.* **2001**, *3*, 299–310.
3. (a) Mergny, J. L.; Sun, J. S.; Rougée, M.; Garestier, T.; Barcelo, F.; Chomilier, J.; Hélène, C. *Biochemistry* **1991**, *30*, 9791–9798; (b) Greenberg, W. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 5016–5022; (c) Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 1187–1193.
4. Recent reviews: (a) Gowers, D. M.; Fox, K. R. *Nucleic Acids Res.* **1999**, *27*, 1569–1577; (b) Purwanto, M. G. M.; Weisz, K. *Curr. Org. Chem.* **2003**, *7*, 427–446; (c) Buchini, S.; Leumann, C. J. *Curr. Opin. Chem. Biol.* **2003**, *7*, 717–726, and references cited therein.
5. (a) Guianvarc'h, D.; Benhida, R.; Fourrey, J.-L.; Maurisse, R.; Sun, J. S. *Chem. Commun.* **2001**, 1814–1815; (b) Guianvarc'h, D.; Fourrey, J.-L.; Maurisse, R.; Sun, J. S.; Benhida, R. *Org. Lett.* **2003**, *4*, 4209–4212; (c) Guianvarc'h, D.; Fourrey, J.-L.; Maurisse, R.; Sun, J. S.; Benhida, R. *Bioorg. Med. Chem.* **2003**, *11*, 2751–2759.
6. (a) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967–971; (b) Kiessling, L. L.; Griffin, L. C.; Dervan, P. B. *Biochemistry* **1992**, *31*, 2829–2834; (c) Wang, G.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 2514–2523; (d) Chandller, S. P.; Fox, K. R. *FEBS Lett.* **1993**, *322*, 189–192; (e) Radhakrishnan, I.; Patel, D. J. *Structure* **1994**, *2*, 17–32; (f) Gowers, D. M.; Fox, K. R. *Nucleic Acids Res.* **1997**, *25*, 3787–3794.
7. In our case, the glycosylation of 5(6)-nitrobenzimidazole in the presence of α-2-deoxy-3,5-ditoluoylribosyl chloride using NaH in CH_3CN yielded the β-nucleoside **2** as a mixture of 5- and 6-regioisomers in nearly 1/1 ratio. The structure of both regioisomers was assigned by 2D COSY-NOESY ^1H NMR, in agreement with the previously reported data Seela, F.; Bourgeois, W. *Synthesis* **1989**, *12*, 912–918.
8. 1-Thyminyl-propionic acid was synthesized as previously reported Mokhir, A. A.; Richert, C. *Nucleic Acids Res.* **2000**, *28*, 4254–4265.
9. Mukaiyama, T.; Usui, M.; Shimada, E.; Saigo, E. *Chem. Lett.* **1975**, 1045.
10. Analytical and spectral data for selected products.
3: ^1H NMR ($CDCl_3$, 200 MHz) δ 1.59 (3H, br s, Me), 2.35 (3H, s, Me), 2.43 (3H, s, Me), 2.60–3.30 (4H, m, CH_2CO and 2′-H), 3.97 (2H, m, CH_2N), 4.50–4.80 (m, 3H, H-4′ and H-5′), 5.74 (m, 1H, H-3′), 6.42 (m, 1H, H-1′), 7.10–7.40 (6H, m, 4×H-Tol., H-5 and H-6-thymine), 7.58 (1H, d, J = 8.5 Hz, H-4), 7.86 (2H, d, J = 8.2 Hz, H-Tol.), 7.98 (2H, d, J = 8.2 Hz, H-Tol.), 8.13 (1H, s, H-7), 8.29 (1H, s, H-2), 9.62 (1H, br s, NH), 10.95 (1H, br s, NH); ^{13}C NMR ($CDCl_3$) δ 11.98, 21.60, 21.70, 35.72, 37.87, 45.66, 64.13, 74.84, 82.36, 85.23, 102.12, 110.02, 116.02, 120.22, 126.37, 126.64, 129.26, 129.58, 129.81, 132.90, 134.67, 140.45, 142.26, 144.09, 144.49, 151.64, 165.28, 165.97, 166.19, 168.95; MS (ES^+) *m/z* 688 (MNa^+).
4: ^1H NMR (CD_3OD, 200 MHz) δ 1.81 (3H, br s, Me), 2.37–2.53 (m, 1H, H-2′), 2.58–2.75 (1H, m, H-2′), 2.81 (2H, t, J = 6.4 Hz, CH_2CO), 3.67–2.87 (2H, m, H-5′), 3.97–2.15 (3H, m, H-4′ and CH_2N), 4.52 (1H, m, H-3′), 6.35 (1H, t, J = 6.7 Hz, H-1′), 7.20 (1H, dd, J = 8.7 and 1.9 Hz, H-5), 7.44 (1H, br s, H-6-thymine), 7.57 (1H, d, J = 8.7 Hz, H-4), 8.13 (1H, d, J = 1.9 Hz, H-7), 8.39 (1H, s, H-2′); ^{13}C NMR (CD_3OD) δ 12.20, 36.58, 41.17, 46.33, 63.20, 72.38, 86.49, 88.95, 104.14, 110.83, 117.44, 120.21, 134.09, 135.66, 141.06, 142.96, 143.73, 152.83, 166.92, 171.07; MS (ES^+) *m/z* 430 (MH^+), 314 (MH^+ sugar).
5: ^{31}P NMR ($CDCl_3$, 200 MHz) δ 148.88 and 148.96; MS (ES^+) *m/z* 932 (MH^+), 303 (DMT^+).
11. The SEM-protected 5-iodouracil **7** was obtained in quantitative yield from 5-iodouracil and used without purification. ^1H NMR ($CDCl_3$, 250 MHz) δ 0.02 (9H, s, $SiMe_3$), 0.90 (2H, t, J = 8.0 Hz, CH_2), 3.58 (2H, t, J = 8.0 Hz, CH_2), 5,12 (2H, s, CH_2), 7.77 (1H, s, H-6), 10.09 (1H, br s, NH). ^{13}C NMR ($CDCl_3$) δ −1.42, 17.83, 67.34, 76.18, 94.09, 147.82, 150.94, 160.71; MS (ES^+) *m/z* 802 (MNa^+), 780 (MH^+), 242 (MH^+ I).
12. Under these conditions only the deprotected 3-(5-thyminyl)-acrylic acid was obtained.
13. Analytical and spectral data for selected products. **8**: ^1H NMR ($CDCl_3$, 200 MHz) δ 2.37 (3H, s, Me), 2.44 (3H, s, Me), 2.82–2.20 (2H, m, H-2′), 4.64 (3H, m, H-4′ and H-5′), 5.70 (2H, m, H-3′ and H-acryl.), 6.37 (2H, m, H-acryl.), 6.54 (1H, t, J = 6.1 Hz, H-1′), 7.10–7.40 (2H, m, H-

Ar), 7.21 (2H, d, $J=8.2$ Hz, 2×H-Tol.), 7.33 (2H, d, $J=8.2$ Hz, 2×H-Tol.), 7.61 (1H, d, $J=9.2$ Hz, H-Ar), 7.81 (2H, d, $J=8.2$ Hz, 2×H-Tol.), 7.84 (1H, m, H-Ar), 8.00 (2H, d, $J=8.2$ Hz, 2×H-Tol.); ^{13}C NMR ($CDCl_3$) δ 21.68, 38.29, 64.14, 74.84, 82, 85.30, 102.28, 116.09, 120.37, 126.31, 126.59, 127.48, 129.35, 129.59, 131.26, 133.08, 134.51, 140.36, 144.29, 144.58, 163.97, 166.07, 166.25; MS (ES^+) *m/z* 562 (MNa^+), 540 (MH^+). **9**: ^{1}H NMR ($CDCl_3$, 250 MHz) δ 0.02 (9H, s, $SiMe_3$), 0.96 (2H, t, $J=8.4$ Hz, CH_2Si), 2.36 (3H, s, Me), 2.46 (3H, s, Me), 2.70–2.10 (2H, m, H-2′), 3.66 (2H, t, $J=8.4$ Hz, OCH_2), 4.60–2.80 (3H, m, H-5′ and H-4′), 5.21 (2H, br s, NCH_2O), 5.81 (1H, m, H-3′), 6.45 (1H, t, $J=6.5$ Hz, H-1′), 7.18 (2H, d, $J=8.2$ Hz, 2×H-Tol.), 7.30 (2H, d, $J=8.2$ Hz, 2×H-Tol.), 7.20–7.50 (2H, m, 5-H and H-acryl.), 7.58 (1H, d, $J=13.1$ Hz, H-acryl.), 7.60 (1H, s, H-7), 7.70 (1H, d, $J=8.5$ Hz, H-4), 7.89 (2H, d, $J=8.0$ Hz, 2×H-Tol.), 8.00 (2H, d, $J=8.0$ Hz, 2×H-Tol.), 8.32 (1H, s, 6-H-thymine), 8.56 (1H, s, H-2), 9.52 (1H, br s, NH); ^{13}C NMR ($CDCl_3$) δ −1.39, 8.63, 17.96, 21.79, 38.06, 46.27, 64.21, 67.45, 74.87, 82.41, 85.09, 102.06, 110.73, 116.21, 120.87, 123.41, 126.42, 126.66, 129.32, 129.70, 129.93, 132.99, 133.27, 135.09, 140.22, 140.70, 144.15, 144.49, 145.51, 150.01, 162.33, 165.07, 166.14, 166.42; MS (ES^+) *m/z* 802 (MNa^+), 780 (MH^+). **10**: ^{1}H NMR (CD_3OD, 250 MHz) δ 0.02 (9H, s, $SiMe_3$), 0.91 (2H, t, $J=8.1$ Hz, CH_2Si), 2.45 (1H, m, H-2′), 2.64 (1H, m, H-2′), 3.65 (2H, t, $J=8.1$ Hz, OCH_2), 3.76 (2H, m, H-5′), 4.02 (1H, q, $J=3.8$ Hz, H-4′), 4.51 (1H, m, H-3′), 5.17 (2H, s, NCH_2O), 6.35 (1H, t, $J=6.7$ Hz, H-1′), 7.20–7.35 (3H, m, H-4, H-5 and H-acryl.), 7.57 (1H, d, $J=8.6$ Hz, H-acryl.), 7.94 (1H, s, H-7), 8.27 (1H, br s, H-6 thymine), 8.34 (1H, s, H-2);^{13}C NMR (CD_3OD) δ −3.16, 16.83, 39.39, 61.22, 66.16, 70.35, 75.96, 83.60, 84.54, 86.88, 101.92, 109.50, 115.50, 118.32, 121.36, 132.82, 134.24, 139.20, 140.85, 143.15, 145.89; MS (ES^+) *m/z* 566 (MNa^+), 544 (MH^+). **12**: ^{1}H NMR (CD_3OD/$CDCl_3$, 250 MHz) δ 0.02 (9H, s, $SiMe_3$), 0.88 (2H, t, $J=7.9$ Hz, CH_2Si), 2.51–2.96 (6H, m, 2×CH_2 and H-2′), 3.66 (2H, t, $J=7.9$ Hz, OCH_2), 3.90 (2H, m, H-5′), 4.16 (1H, m, H-4′), 4.65 (1H, m, H-3′), 5.21 (2H, s, NCH_2O), 6.51 (1H, t, $J=6.0$ Hz, H-1′), 7.03 (2H, m, H-4 and H-5), 7.63 (1H, s, H-7), 7.84 (1H, br s, H-6-thymine), 8.36 (1H, s, H-2), 8.81 (1H, br s, NH); ^{13}C NMR (CD_3OD/$CDCl_3$) δ −3.27, 16.60, 19.45, 22.98, 33.32, 35.95, 39.67, 46.00, 60.88, 65.59, 70.07, 75.34, 85.03, 87.10, 102.02, 112.22, 116.16, 117.64, 127.49, 134.60, 136.92, 140.95, 150.81, 163.99, 171.19; MS (ES^+) *m/z* 568 (MNa^+), 546 (MH^+).

14. **14**: two imine-isomers: ^{1}H NMR ($CDCl_3$, 250 MHz) δ 2.30–2.80 (2H, m, H-2′), 2.92 (6H, s, 2×Me), 3.62 (2H, m, H-5′), 4.01 (m, 1H, H-4′), 4.51 (1H, m, H-3′), 6.19 (1H, t, $J=6.4$ Hz, H-1′), 6.85 (1H, d, $J=8.5$ Hz, H-5), 7.22 (1H, br s, H-7), 7.31 (1H, m, H-4), 7.51 (1H, s, H-2), 8.13 and 8.31 (1H, s, NCH); ^{13}C NMR ($CDCl_3$) δ 40.07, 40.30, 61.70, 62.09, 70.67, 71.03, 84.73, 85.10, 87.33, 87.91, 109.91, 110.41, 118.86, 128.99, 129.24, 140.96, 142.00, 144.14, 147.79, 153.85, 155.59, 162.41; MS (ES^+) *m/z* 305 (MH^+). **15**: imine and phosphoramidite isomers: ^{31}P NMR ($CDCl_3$, 200 MHz) δ 147.27, 147.50, 147.74, and 147.82.