Pergamon

Tetrahedron Letters 39 (1998) 547–550

TETRAHEDRON LETTERS

Synthesis of 4-guanidinopyrimidine nucleosides for triple helix-mediated guanine and cytosine recognition.

Svetlana O. Doronina and Jean-Paul Behr*.

Laboratoire de Chimie génétique associé CNRS/Université Louis Pasteur,
Faculté de Pharmacie de Strasbourg, BP 24, F-67401 Illkirch

Received 8 September 1997; accepted 12 November 1997

Abstract: Recognition of DNA sequences by triple helix-forming oligonucleotides is restricted to homopurine sequences. Based on geometrical considerations, a single nucleoside, 4-guanidinocytidine **1**, was designed to recognize both guanosine (as the natural 1β anomer) and cytidine (as the 1α anomer) in a new triple helix context where the third strand is centered in the major groove of the target duplex. Several methods of synthesis of this nucleoside are described.

Synthetic oligonucleotide-directed triple helix formation represents a promising way to interfere with gene expression and has been an area of intense investigation since it was first demonstrated in 1987[1]. Much effort has been devoted to overcoming the major limitation of this approach, i.e. the requirement of a long homopurine tract on the target DNA[2]. Triplex formation indeed involves bidentate Hoogsteen hydrogen bond formation between the nucleic bases of the third-strand oligonucleotide and purines in the major groove of the target DNA duplex. Pyrimidines only present a single hydrogen bonding site which makes their recognition tricky.
In our approach the difficulty of pyrimidine recognition is circumvented by binding the opposite purine of the Watson-Crick base pair in a non-natural triple helix context where the ribose-phosphate chain of the oligonucleotide is *centered* in the major groove of DNA (Fig. 1). In this particular geometry the targeting residue is shifted away relative to the known Py*Pu-Py and Pu*Pu-Py triple helix families to become equidistant from the strands of the target duplex. As a consequence, molecular recognition of a random sequence may be achieved with only two targeting residues able to 'read' guanine and adenine on either strand through the α/β anomerism[2b]. This approach leads to a perfect triplex isomorphism, yet at the expense of optimal base stacking at each purine/pyrimidine junction.

4-Guanidinopyrimidine nucleosides (**1**) match the extended third-strand-ribose-to-target-base distance and have the potential H-bond network of a planar guanine-binding residue (Fig. 1). Here we report several methods to synthesize **1**.

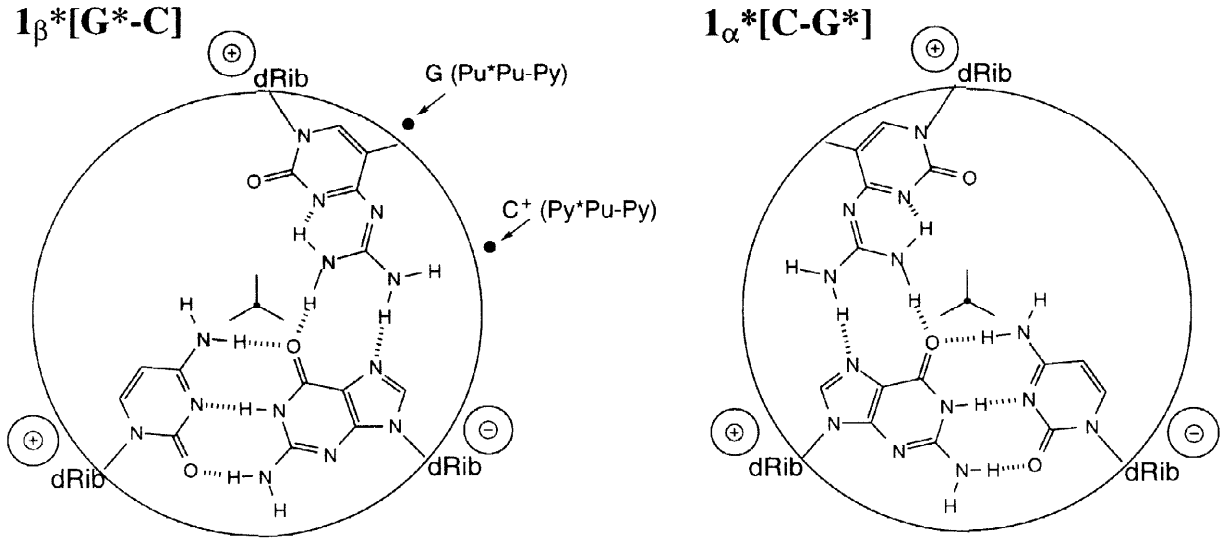


Fig. 1 Guanidinopyrimidine **1** from an oligonucleotide centered in the major groove would be able to bind guanine on either strand through the α/β anomerism. (•) is the position of the (C1')dRib within the known triple helix families. Hoogsteen hydrogen-bonded heterocycles are labeled with an asterisk (*).

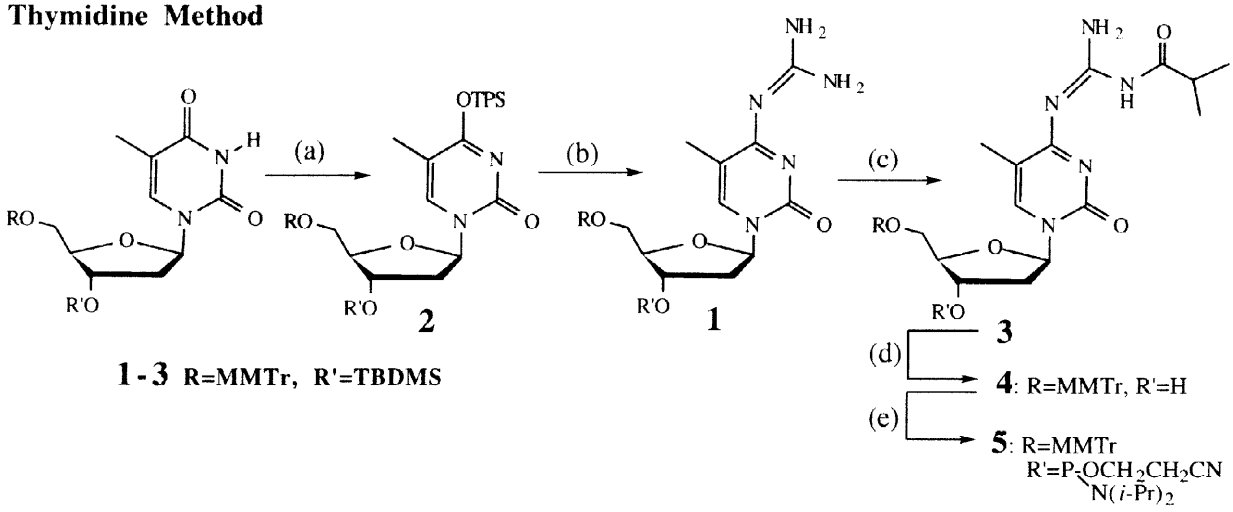
* e-mail: behr@pharma.u-strasbg.fr

0040-4039/98/$19.00
PII: S0040-4039(97)10627-X

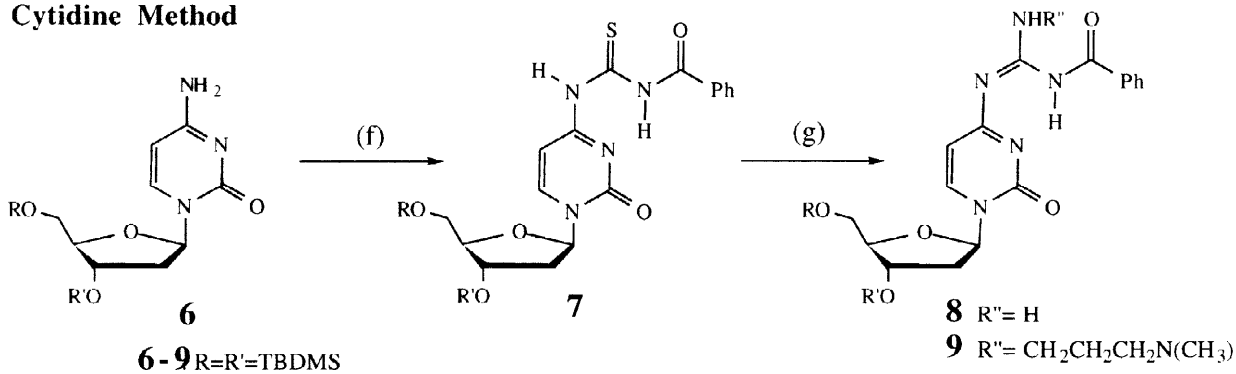
Straightforward mild methods of guanidinopyrimidine **1** synthesis *via* reaction of cytidine hydrochloride with hydrogen cyanamide (a general route to aromatic guanidines[3]), with *N,N'*-di(Boc)thiourea/Hg(II)[4] or with 1*H*-pyrazole-1-carboxamidines[5] were unsuccessful. Nucleophilic substitution of pyrimidines[6] (the Thymidine method) and guanidine synthesis *via* reaction of activated thioureas with amines in the presence of a carbodiimide[7] (the Cytidine method) gave **1**, yet with moderate yields.

In the Thymidine method, the 4-*O*-triisopropylphenylsulfonyl (OTPS) thymidine derivative **2** was guanylated directly in appropriate conditions. This type of reaction was described for various N-, O-, and C-nucleophiles[8]. The strongly basic nature of guanidine and low solubility of guanidinium salts in most solvents led us to choose *t*-BuOK/*t*-BuOH as homogeneous non-nucleophilic base. When **2** was treated with a large excess (10 eq.) of guanidinium chloride in *t*-BuOK/*t*-BuOH at room temperature for 4 hours, TLC revealed a fluorescent spot corresponding to the highly conjugated guanidino derivative **1**[9]. After workup, the reaction product was recovered by preparative silica gel chromatography in 18-40% yield. The main by-product, thymidine, resulted from preferential reaction of **2** with unavoidable trace amounts of water rather than with the weak nucleophile guanidine. Some base-induced desilylation also occured. The 5'-*O*-MMTr-3'-*O*-silyl derivative of **2** appeared to be more resistant to strong base, as compared to 5', 3'-*O*-di-TBDMS **2**. The Thymidine method was used to synthesize also the α-anomer of nucleoside **1** starting from commercial α-thymidine[10].

Thymidine Method



Cytidine Method



Scheme. (a). TPS-Cl(1.5eq.)/NaH(5eq.)/THF/12h/25°C/95%; (b). Guanidine·HCl(10eq.)/*t*-BuOK(10eq.)/*t*-BuOH/4h/25°C/18-40%; (c). (*i*-PrCO)$_2$O(1.5eq.)/Et$_3$N(5eq.)/4-DMAP/CH$_2$Cl$_2$/1h/25°C/85%; (d). Bu$_4$NF(10eq.)/THF/2h/25°C/90%; (e). ((*i*-Pr)$_2$N)$_2$POCH$_2$CH$_2$CN(2eq.)/Tetrazole(1eq.)/CH$_2$Cl$_2$/3h/25°C/75-80%; (f). Bzl-NCS(1.3eq.)/Et$_3$N/CH$_2$Cl$_2$/18h/25°C/15%; (g). NH$_2$R"(5-10 eq.)/EDC(1-2 eq.)/DMF/3-18h/25°C/60-70%.

The Cytidine method is based on a recently described guanidine synthesis *via* reaction of acyl thioureas with amines in the presence of a carbodiimide[11]. 3',5'-*O*-Protected cytidine **6** was converted to the thiourea derivative **7**[12] by reaction with benzoylisothiocyanate. Benzoylisothiocyanate was chosen as a reagent because the benzoyl group could act as a protecting group in the subsequent oligomerization procedure. Although reaction of cytosine with isocyanates has been used for the preparation of various 4-ureidopyrimidines with moderate to good yields[13], the reaction with acylisothiocyanates led to partial decomposition with loss of

thiocyanic acid (HSCN)[14]. The yield of **7** did not exceed 15%, whereas the decomposition product, benzoylcytidine, was obtained with 60% yield. Changing the solvent polarity or the temperature did not improve the yield. The cytidine method was nevertheless attractive since the low-yield step was that converting readily available cytidine into an already base-protected nucleoside. Reaction of **7** with ammonium hydrogencarbonate in the presence of 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC) proceeded smoothly[15] at room temperature to give guanidino derivative **8**. Reaction with another amine (3-(*N*-dimethyl)-1,3-diaminopropane to yield **9**) confirmed this approach to be of general use for the synthesis of substituted guanidinopyrimidines. Debenzoylation of **8** was either performed in K_2CO_3/aqu. MeOH (1 h) or in standard oligonucleotide deprotection conditions (32% ammonia) to give **10**[16].

Prior to testing the triple helix formation according to Fig. 1, the protonation state of the guanidine had to be assessed and the nonnatural nucleoside had still to be converted into an oligonucleotide synthon.
In contrast to third-strand cytidine binding to guanine in a Py*Pu-Py triple helix, pyrimidinoguanidine **1** does not require protonation to establish two hydrogen bonds with guanine (Fig. 1). Moreover, adjacent cationic charges on an oligonucleotide are detrimental to triplex stability. The pK of **1** was derived from the pH dependence of the UV spectrum (**1**: ε(303 nm) = 25.0 mL/cm·μmol ; $\mathbf{1H^+}$: ε(300 nm) = 12.2 mL/cm·μmol) and found to be equal to 7.5 ± 0.5 (for comparison, benzoylguanidine[20] has pK = 7.0). This value would be shifted to the basic side for an isolated site in an oligonucleotide and probably lead to an interesting case of self-regulation of the protonation in a guanidinopyrimidine tract under physiological pH conditions.
The guanidino group of nucleoside **1** was protected ($(i\text{-PrCO})_2O/Et_3N/CH_2Cl_2$)[17], the silyl group of **3** was removed with Bu_4NF/THF to provide intermediate **4** (yield 90%)[18]. Phosphitylation with ($(i\text{-Pr})_2N)_2POCH_2CH_2CN$/tetrazole following a known procedure afforded synthon **5**[19] in 75-80% yield.
The automated synthesis of oligonucleotides made of fully modified pyrimidine nucleosides is now being optimized.

Acknowledgements. We warmly thank Mrs Elisabeth Krempp for her enthusiasm and help with the NMR techniques. S.O.D. was supported by a Bourse Doctorale du Gouvernement Français.

References and Notes

1. Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645-650; Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.; Thuong, N. T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *19*, 7749-7760.

2. a) Thuong, N. T.; Hélène, C. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 666-690; b) Doronina, S. O.; Behr, J.-P. *Chem. Soc. Reviews* **1997**, *26*, 63-71.

3. Hughes, J. L.; Liu, R. C.; Enkoji, T. *J. Med. Chem.* **1975**, *18*, 1077-1088.

4. Kim, K. S.; Qian, L. *Tetrahedron Lett.* **1993**, *34*, 7677-7680.

5. Bernatowicz, M. S.; Wu, Y.; Matsueda, G. R. *J. Org. Chem.* **1992**, *57*, 2497-2502; Dutta, S. P.; Chheda, G. B. in *Nucleic Acid Chemistry*, part 4; Townsend, L. B.; Tipson, R. S.; John Wiley and Sons, Inc.: New York, 1991; pp. 245-247.

6. Bischofberger, N. *Tetrahedron Lett.* **1987**, *28*, 2821-2824.

7. Atwal, K. S.; Ahmed, S. Z.; O'Reilly, B. *Tetrahedron Lett.* **1989**, *30*, 7313-7316.

8. For N-nucleophiles see Perbost, M.; Sanghvi, Y. S. *J. Chem. Soc., Perkin Trans. 1* **1994**, 2051-2052. For O-nucleophiles, Xu, Y.-Z.; Swann, P. F. *Nucleic Acids Res.* **1990**, *18*, 4061-4065. For C-nucleophiles, ref. 6.

9. **1**: R_f=0.3 ($MeOH/CH_2Cl_2$, 10/90). MS (FAB+), m/z: 670 (8 %, $[M+H]^+$). ^1H NMR ($CDCl_3$), δ (ppm): -0.06 and 0.00 (2s, 6H, 3'-Si(CH_3)$_2$), 0.81 (s, 9H, 3'-Si-C(CH_3)$_3$), 1.60 (s, 3H, CH_3), 2.10-2.28 (m, 1H, H2'), 2.32-2.54 (m, 1H, H2"), 3.19 -3.55 (m, 2H, H5', H5"), 3.79 (s, 3H, OCH_3), 3.94-4.03 (m, 1H, H4'), 4.46 (m, 1H, H3'), 6.29 (t, 1H, H1'), 6.84 (d, 2H, *J* = 8.8 Hz, *o*-CH_3O-phenyl), 7.18-7.44 (m, 12H, phenyl), 7.84 (s, 1H, H6).

10. α **1**: from conversion of α-5'-*O*-MMT-3'-*O*-silyl **2** which should be used immediately after preparation. Yield 23% as a faint yellow oil. R_f=0.25 ($MeOH/CH_2Cl_2$, 10/90). ^1H NMR ($CDCl_3$), δ (ppm): -0.05 and 0.04 (2s, 6H, 3'-Si(CH_3)$_2$), 0.78 (s, 9H, 3'-Si-C(CH_3)$_3$), 1.98 (s, 3H, CH_3), 2.03-2.11 and 2.53-2.64 (2m, 2H, H2', H2"), 3.81 (s, 3H, OCH_3), 3.15 (m, 2H, H5', H5"), 4.26 (m, 1H, H4'), 4.41 (m, 1H, H3'), 6.75 (dd, 1H, *J* = 5.4 Hz, H1'), 6.85 (d, 2H, *J* = 8.8 Hz, *o*-CH_3O-phenyl), 7.15-7.46 (m, 12H, phenyl), 7.64 (s, 1H, H6).

11. Poss, M. A.; Iwanowicz, E.; Reid, J. A.; Lin, J.; Gu, Z. *Tetrahedron Lett.* **1992**, *33*, 5933-5936.

12. **7**: R_f=0.7 ($MeOH/CH_2Cl_2$, 5/95). MS (FAB+), m/z: 619 (10 %, $[M+H]^+$). RT ^1H NMR of **7** showed no resonance for H5 whereas H6 still gave a doublet (*J* = 7.5 Hz). The H5/H6 proton correlation peak was also absent from

the COSY spectrum. Variable temperature spectra showed evidence for hindered rotation about N4-C(S) bond. Two isomers in an equimolar ratio were observed in the 230°K spectrum.

13. Dutta, S. P.; Chheda, G. B. *J. Carbohydrates Nucleosides Nucleotides* **1980**, *7*, 217-240.

14. Capuano, L.; Schrepfer H. J. *Chem. Ber.* **1971**, *104*, 3039-3047.

15. Typical procedure: 1.2-2 eq. of EDC was added to a mixture of 1 eq. of thiourea derivative **7** and 5-10-fold excess of amine in 4 ml of dry DMF. The reaction mixture was stirred overnight. After workup the compound was purified by preparative chromatography (0-20 % MeOH in CH_2Cl_2).
Compound 8: yield 72%. R_f=0.45 (MeOH/CH_2Cl_2, 5/95). MS (FAB+), m/z: 602 (8 %, $[M+H]^+$). ^1H NMR (methanol-d_4), δ (ppm): 0.08 and 0.14 (2s, 12H, 3'- and 5'-Si$(CH_3)_2$), 0.88 and 0.96 (2s, 18H, 3'- and 5'-Si-C$(CH_3)_3$), 2.12-2.23 (m, 1H, H2'), 2.36-2.54 (m, 1H, H2"), 3.84-3.97 (m, 3H, H5', H5", H4'), 4.90 (m, 1H, H3'), 6.18 (d, 1H, J = 7.4 Hz, H5), 6.20 (t, 1H, J = 6.0 Hz, H1'), 7.41-7.61 (m, 3H, *m*-H and *p*-H of benzoyl), 8.10-8.15 (m, 2H, *o*-H of benzoyl), 8.24 (d, 1H, J = 7.4 Hz, H6).
Compound 9: yield 65%. R_f=0.1 (MeOH/CH_2Cl_2, 5/95), 0.5 (MeOH/CH_2Cl_2, 10/90). MS (FAB+), m/z: 687 (89 %, $[M+H]^+$). In the ^1H NMR spectra of **9** H5, H6, N-H and H1' signals splitting was observed, probably due to isomerism relative to the exocyclic double bond.^1H NMR ($CDCl_3$), δ (ppm): 0.06 and 0.12 (2s, 12H, 3'- and 5'-Si$(CH_3)_2$), 0.86 and 0.94 (2s, 18H, 3'- and 5'-Si-C$(CH_3)_3$), 1.86-2.02 (m, 2H, C-CH_2-C), 2.09-2.23 (m, 1H, H2'), 2.32 and 2.33 (2s, 6H, N$(CH_3)_2$), 2.45-2.58 (m, 3H, H2" and C-CH_2-N$(CH_3)_2$), 3.62 -3.85 (m, 3H, H5', H5", H4'), 3.92-4.03 (m, 2H, NH-CH_2-C), 4.41 (m, 1H, H3'), 5.92 (B, J = 7.4 Hz) and 6.09 (A, J = 7.2 Hz) (d, 1H, H5), 5.96 (B, J_1 = 6.4 Hz, J_2 = 4.5 Hz) and 6.34 (A, J_1 = 6.4 Hz, J_2 = 5.2 Hz) (dd, 1H, H1'), 7.41-7.54 (m, 1H, *p*-H of benzoyl), 7.57-7.65 (m, 2H, *m*-H of benzoyl), 8.21-8.32 (m, 2H, *o*-H of benzoyl), 8.11 (A, J = 7.2 Hz) and 8.35 (B, J = 7.4 Hz) (d, 1H, H6), 9.42 (A) and 10.6 (B) (1H, bt, NH-CH_2-), 13.6 (B) and 15.9 (A) (bs, 1H, NH-CO).

16. **10**: MS (FAB+), m/z: 498 (15 %, $[M+H]^+$). ^1H NMR ($CDCl_3$), δ (ppm): 0.04 and 0.10 (2s, 12H, 3'- and 5'-Si$(CH_3)_2$), 0.82 and 0.94 (2s, 18H, 3'- and 5'-Si-C$(CH_3)_3$), 2.02-2.16 and 2.32-2.47 (2m, 2H, H2', H2"), 3.67-3.98 (m, 3H, H5', H5", H4'), 4.39 (m, 1H, H3'), 5.88 (d, J = 7.4 Hz, H5), 6.27 (dd, 1H, J = 5.8 Hz, H1'), 7.96 (d, 1H, J = 7.4 Hz, H6).

17. ß **3**: R_f=0.52 (MeOH/CH_2Cl_2, 5/95). MS (FAB+), m/z: 740 (6 %, $[M+H]^+$). Calc. C 66.58, H 7.17, N 9.47 %. $C_{41}H_{53}N_5Si0_6$. Found C 66.16, H 7.25, N 9.37 %. ^1H NMR ($CDCl_3$), δ (ppm): -0.60 and 0.06 (2s, 6H, 3' -Si$(CH_3)_2$), 0.79 (s, 9H, 3' -Si-C$(CH_3)_3$), 1.28 (d, 6H, J = 8.0 Hz, -CH$(CH_3)_2$), 1.62 (s, 3H, CH_3), 2.13-2.27 and 2.40-2.72 (2m, 3H, H2', H2", -CH$(CH_3)_2$), 3.24-3.54 (m, 2H, H5', H5"), 3.79 (s, 3H, OCH_3), 3.99 (m, 1H, H4'), 4.45 (m, 1H, H3'), 6.32 (t, 1H, J = 6.0 Hz, H1'), 6.84 (d, 2H, J = 8.9 Hz, *o*-CH_3O-phenyl), 7.21-7.41 (m, 12 H, phenyl), 7.92 (s, 1H, H6).
α **3**: R_f=0.50 (MeOH/CH_2Cl_2, 5/95). MS (FAB+), m/z: 740 (7 %, $[M+H]^+$). Calc. C 66.58, H 7.17, N 9.47 %. $C_{41}H_{53}N_5Si0_6$. Found C 66.57, H 7.29, N 9.29 %. ^1H NMR ($CDCl_3$), δ (ppm): -0.06 and 0.00 (2s, 6H, 3'-Si$(CH_3)_2$), 0.80 (s, 9H, 3'-Si-C$(CH_3)_3$), 1.28 (d, 6H, J = 8.0 Hz, -CH$(CH_3)_2$), 1.98 (s, 3H, CH_3), 2.03-2.11 and 2.53-2.64 (2m, 3H, H2', H2", -CH$(CH_3)_2$), 3.15 (m, 2H, H5', H5"), 3.81 (s, 3H, OCH_3), 4.28 (m, 1H, H4'), 4.41 (m, 1H, H3'), 6.28 (dd, 1H, J = 5.4 Hz, H1'), 6.85 (d, 2H, J = 8.8 Hz, *o*-CH_3O-phenyl), 7.15-7.46 (m, 12H, phenyl), 7.64 (s, 1H, H6).

18. ß **4**: R_f=0.35 (MeOH/CH_2Cl_2, 5/95). MS (FAB+), m/z: 626 (5 %, $[M+H]^+$). ^1H NMR ($CDCl_3$), δ (ppm): 1.26 (d, 6H, J = 8.0 Hz, -CH$(CH_3)_2$), 1.64 (s, 3H, CH_3), 2.15-2.30 and 2.55-2.69 (2m, 3H, H2', H2", -CH$(CH_3)_2$), 3.31-3.52 (m, 2H, H5', H5"), 3.79 (s, 3H, OCH_3), 4.15 (m, 1H, H4'), 4.52 (m, 1H, H3'), 6.46 (t, 1H, J = 6.0 Hz, H1'), 6.86 (d, 2H, J = 8.9 Hz, *o*-CH_3O-phenyl), 7.26-7.41 (m, 12 H, phenyl), 7.82 (s, 1H, H6).
α **4**: R_f=0.30 (MeOH/CH_2Cl_2, 5/95). MS (FAB+), m/z: 625 (18 %, $[M+H]^+$). ^1H NMR ($CDCl_3$), δ (ppm): 1.27 (d, 6H, J = 8.0 Hz, -CH$(CH_3)_2$), 1.98 (s, 3H, CH_3), 2.45-2.60 and 2.75-2.95 (2m, 2H, H2', H2"), 2.60-2.72 (m, 1H, -CH$(CH_3)_2$), 3.10-3.35 (m, 2H, H5', H5"), 3.82 (s, 3H, OCH_3), 4.47 (m, 2H, H3', H4'), 5.90 (dd, 1H, J = 5.4 Hz, H1'), 6.85 (d, 2H, J = 8.8 Hz, *o*-CH_3O-phenyl), 7.15-7.43 (m, 13H, H6, phenyl).

19. R_f=0.62, 0.65 (ß); 0.58, 0.60 (α) (MeOH/CH_2Cl_2, 10/90). MS (FAB+), m/z: 830 (4 %, $[M+H]^+$). ^{31}P NMR (acetonitrile-d_3), δ (ppm): 149.19, 149.39 (ß); 149.74, 150.05 (α).

20. Taylor, P. J.; Wait, A. R. *J. Chem. Soc., Perkin Trans. 2* **1986**, 1765-1770.