Pyranosyl-RNA Supramolecules Containing Non-Hydrogen Bonding Base-Pairs

Christian Hamon,[a] Tilmann Brandstetter[1], Norbert Windhab[a*]

[a] Department for Bio-Organic Systems, Aventis Research & Technologies, D-65926 Frankfurt am Main, Germany
Fax +49(69)305 81535; E-mail: windhab@aventis.com

Received 9 February 1999

Abstract: Synthesis and properties of a new pyranosyl-RNA nucleoside using tryptamine as nucleo-base is reported. Incorporation of this unit into oligomers using standard phosphoamidite chemistry yielded self-complementary and non-selfcomplementary oligonucleotide pairs. Thermal melting experiments of these examples showed the sequence dependent stabilising characteristics of the incorporated base in the symmetric pairing constitution with a standard T_m near that of a similar A-T-pair as well as pairing selectivity with respect to non-symmetric pairing tolerating thymine but destabilizing if confronted to an adenine as complementary base in the antiparallel strand.

Key words: *N*-glycosides, synthetic nucleotides, supramolecular chemistry, base pairing, nanotechnology

With the invention of pyranosyl-RNA Eschenmoser et al. presented 1993 a new pairing system, a synthetic class of molecules designed for the antithetic approach to the question why furanoses and not pyranoses (in the nucleic acids' backbone)[2a] using synthetic molecules as hypothetic challenges to our understanding of basic principles. p-RNA surprisingly well matched the prerequisites for a genotypic function and has even quantifiable advantages in pairing selectivity and stability with respect to its isomer natural RNA and to DNA. The ongoing work on the phenotypic properties with respect to conformation and reactivity makes p-RNA to a distinguished representative of constitutionally prebiotically possible systems, studied in his laboratories.[2b,c]

p-RNA strands consisting of nucleosides of the same chirality pair exclusively in the Watson-Crick-modus in antiparallel strand orientation. No triplex formation has been reported and the idealized diamantoide backbone conformation (Figure 1) illustrates the pairing orthogonality of the planar base-pair repetition throughout a p-RNA duplex to natural systems, the latter leading to repetitive angular increments as twist and pitch of the helix. The exclusivity and orthogonality of its pairing properties make p-RNA to a model for pairing principles and to a synthetically manageable supramolecular entity being both a theoretical and a technical model system in the field of Life Science and Nanotechnology.[3a,b,c]

To increase the repertoire for controlled supramolecular synthesis for different technical applications and to fit changing conditions, the supramolecular chemist seeks for modular subunits in much higher variance than the natural gene encoding nucleosides do provide for self assembling systems.

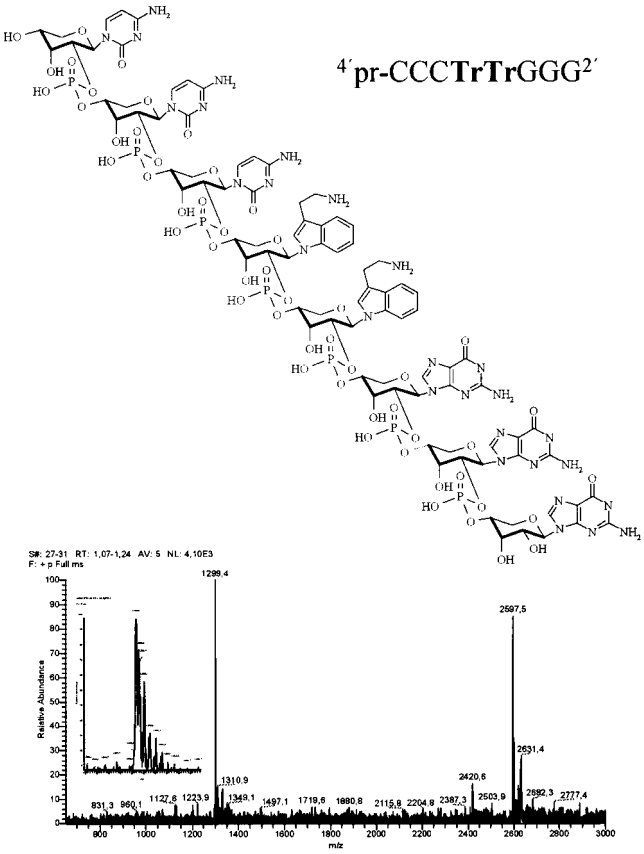


Figure 1 ESI-mass-spectrum and constitution of a p-RNA oligonucleotide containing the unnatural base **Tr**.

Co-operative stability afforded a good stacking residue, our nanotechnological approach requested a versatile linker-functionalisation to link the pairing strand with any synthetic or biological molecule or solid support at any position in its sequence. Another aspect was to incorporate a fluorescent probe without additional labelling.

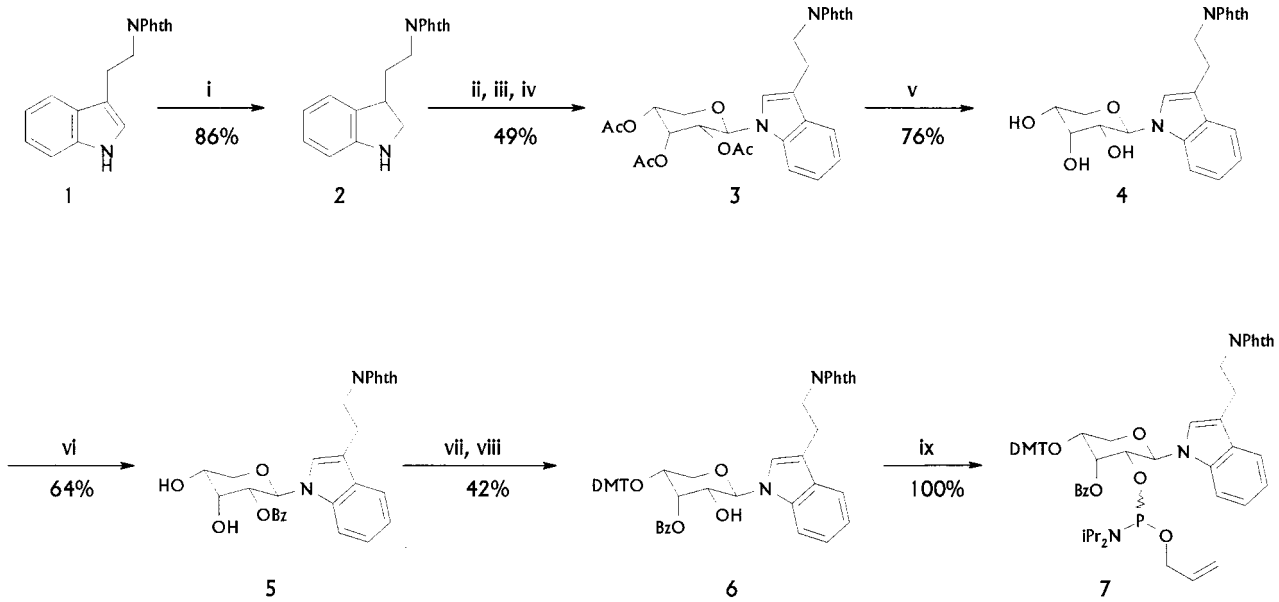
Excellent planar stacking properties can be expected from cristallographic data reported for tryptamine with different biogenic heterocycles such as adenine-9-acetic acid,[4a] and 7-methylguanosine.[4b]

Successful incorporation of different isosteric compounds have been reported for natural DNA and nonselective pairing for the case of pairing with natural bases.[5] Two major differences to the natural system must be considered: p-RNA shows a strong inter-strand stacking which

is 2'→ 4' directional and a lateral broadening of the duplex with respect to pairing residues is theoretically possible which would unwind any helical system. This is why we neglected the isosteric considerations and used tryptamine even as pseudo-symmetric base pair, pairing with itself.

The synthesis of the supramolecules starts with the synthesis of the amidite-monomer for synthesizer chemistry (Scheme 1). The phthalimide protecting group was chosen in order to get a fully deprotected oligo after nucleophilic cleavage from CPG-support. Direct catalysed *N*-glycosidation proved to be very unsatisfactory.

N-Phthaloyl tryptamine **1** was synthesized from phthalic anhydride and tryptamine[7]. Reduction with borane-THF[6] gave the indoline derivative **2**. The 3-substituted indoline **2** was reacted with ribose in ethanol to yield the nucleoside triol. The hydroxyl groups were protected with acetic anhydride, and the indoline was oxidized with DDQ to give nucleoside **3**. Treatment with sodium methoxide removed the acetates. The 2'-position of **4** was benzoylated selectively at –78 °C. Reaction with dimethoxytrityl chloride took place predominantly at position 4' of the hydroxyl group of **5**. After rearrangement of the benzoyl group to position 3' yielding compound **6**,



i: 1M-borane-THF, CF_3CO_2H, 0°C, 30 min; ii: D-ribose, dry ethanol, reflux, 4h; iii: Ac_2O, py., rt, 18h; iv: DDQ, CH_2Cl_2, rt, 1.5h; v: MeONa, dry methanol, rt, 18h; vi: C_6H_5COCl, CH_2Cl_2, DMAP, py., -78°C, 15 min; vii: DMTCl, CH_2Cl_2, DMAP, py. DIPEA, molecular sieves, rt, 4.25h; viii: DMAP, py., DIPEA, n-propanol, p-nitrophenol, 75-80°C, 96h; ix: ClP(Oallyl)(iPr_2N), DIPEA, CH_2Cl_2, rt, 2h.

Scheme 1[(13-18)]

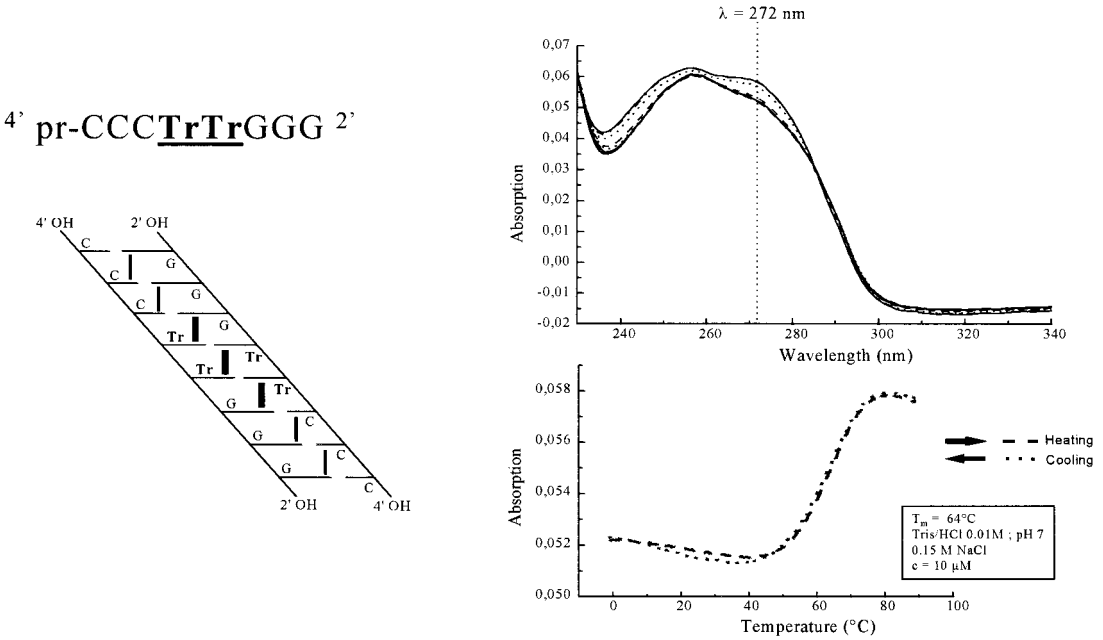


Figure 2. Example of the reversible hypochromicity profile at 10 mM Tris/ HCl (pH 7.0) and 150 mM NaCl, recorded at 272 nm. The pairing pattern could show three strong interstrand stacking interactions with the tryptamine base.

the phosphoramidite **7** was obtained under standard conditions and could be used for automatic oligonucleotide synthesis without changing the protocol of p-RNA synthesis.[2a, 13-18]

Melting studies were carried out at 272 nm on a computer-interfaced Perkin-Elmer Lamba 2 UV-visible spectrophotometer in stoppered 1 cm path length quartz cells for the 1, 3 and 5 µM concentration of oligomer and with 0,1 cm path length for the 10, 25, 50, 100 µM concentration. Samples of 1 - 100 µM concentration of oligomer were prepared in a buffer solution composed of 10 mM Tris/HCl (pH 7.0) and 150 mM NaCl. Degasing and heating up for one time above 80 °C yielded fully reversible hypochromicity profiles. The melting curves (Figure 2) were measured using a temperature gradient from 5 °C to 90 °C for heating and cooling curves at a ramp rate of 1 °C/40s with a tailor-made micro-thermoelement in the sample cell read by a programmed Keithley Instruments DAS-801-AT-Bus measure card (Quick-Basic 4.0 driver). Uncertainty in the T_m data is estimated at ± 0.3 °C based on the repetitions of experiments. Thermodynamic parameters were estimated by the determination of T_m fitting on the hypochromicity profile to a two state transition model as described[8,10] and a van't Hoff interpretation of the concentration-dependent equilibria involving only duplex formation and transitions that proceed in a two-state manner (Figure 3, Table 1).

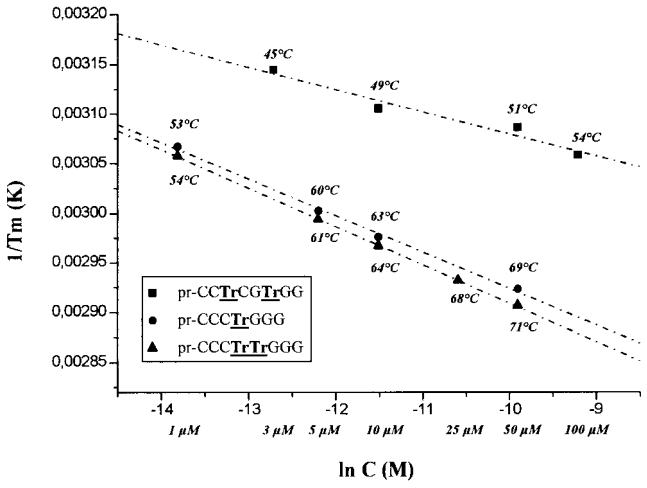


Figure 3 Concentration dependence of T_m for three self-complementary dimerisations.

Table 1 Comparison of the thermodynamic melting data. pr-CCCGGG* data reported.[9]

Sequence	Melting Temperature	Thermodynamic data (kcal/Mol)			
	T_m (°C, 10µM)	ΔH_{VH}	$T\Delta S_{VH}^{298\,K}$	$\Delta G_{VH}^{298\,K}$	K $(M)^{298\,K}$
pr - CCCGGG *	68	-48,5	-35,5	-13,0	-
pr - CCC**Tr**GGG	63	-54,1	-41,2	-12,9	$2{,}8 \cdot 10^9$
pr - CCC**TrTr**GGG	64	-51,4	-38,6	-12,8	$2{,}3 \cdot 10^9$
pr - CC**Tr**CG**Tr**GG	49	-88,5	-75,3	-13,2	$4{,}7 \cdot 10^9$
pr - C**Tr**C**Tr**G**Tr**G	no pairing detected				

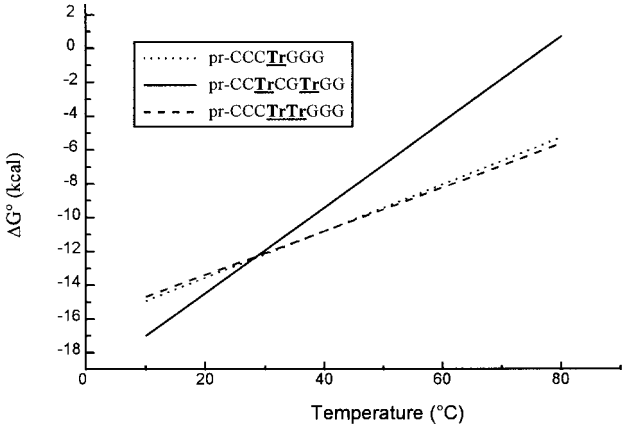


Figure 4 Calculated coalescence of the three different equilibria around 25 °C.

Under the assumption that enthalpy and entropy are not temperature dependent in the experimental range, the slope of the free enthalpy as shown in Figure 4 illustrates the difference of the pairing behaviour although the free enthalpy at 25 °C is the same in the range of experimental error. The strong entropic disadvantage of two isolated "non-polar"-tryptamine groups in a single strand (pr-CC**Tr**CG**Tr**GG) can be explained due to the surface shown to the rest of the molecule. The sequential repetition (pr-CCC**TrTr**GGG) will have less "unnatural" base interaction along the co-operative unit and in fact shows a temperature dependence of the free enthalpy as in the case of a single **Tr**-residue (pr-CCC**Tr**GGG) incorporated. Repetition of **Tr** seems to overcompensate (Figure 4). However this difference could occur through a different mechanism: the mode of transition. The data presented here are derived using equilibria without taking any population of intermediate states into account. This could be the case for a strong enthalpic stabilising entity which decreases the length of the co-operative unit.[10] This non-co-operative quality, as well as the strong unfavorable entropic effect, could be an explanation for the absence of any detectable pairing for the sequence with three **Tr**-residues (pr-C**Tr**C**Tr**G**Tr**G) up to a concentration of 500 µM (Table 1). The sequence with the same number of G-C-pairs was not investigated as the appropriate p-RNA sequences with a number of bases superior to 8 (such as pr-CC**Tr**C**Tr**G**Tr**GG) would be likely to form a hairpin structure.[2c]

A second group of non self-complementary oligomers reveals first insights into the pairing selectivity with respect to mismatch discrimination (Table 3). The clear unfavourable pairing to an adenine complement with full toleration of a thymine and a tryptamine complement can be explained by the water cloud that the adenine presents into the pairing pocket - one explanation of a ground state selectivity of natural pairing, with hydrogen bonds that are destabilising if not saturated (by the Watson-Crick-complement) but not leading to a significant (> 1 kcal/mol) stabilisation if water is present *per se*, as the interstrand stack of adenine is still possible.[11]

This interstrand stacking behaviour is in contrast to another unnatural non-hydrogen base designed for interstrand stack in DNA.[12] Kool's pyrene nucleoside destabilises the DNA duplex for any canonical pairing partner. Interestingly two pyrenes paired against one another show nearly the same stability as a natural A-T-pair. This very same comparison was obtained in our first mismatch study (Table 3).

Table 2. Mismatch examples synthesized.[a)]

Sequence	[MH$^+$] found	[MH$^+$] calc.
pr - NH_2---TAGGCA**A**T	2728,0	2728,5
pr - NH_2---A**T**TGCCTA	2678,6	2678,4
pr - Cy3---A**Tr**TGCCTA	3053,0	3053.7
pr - Biotine---TAGGCA**Tr**T	2991,1	2991,6

a) For technical reasons different 4'-labels are used. Reference experiments show minimal impact on pairing thermodynamics.

Table 3. Non self-complementary pairs and their T_m at 5+5 µM strand concentration. **X** corresponds to either a dye-label (Cy-3) or a biotine or an amino-linker. Our previous studies have shown that only minor differences for T_m (< ± 1 °C) are obtained when the 4' end of the pRNA is occupied with either the dye (Cy-3), biotine, aminolinker or simply a free 4'-OH group. Strand orientation: →.

Duplex	T_m (°C)
→ 4' pr- X T A G G C A **A** T 2' 2' pr- A T C C G T **T** A X 4' ←	51
→ 4' pr- X T A G G C A **Tr** T 2' 2' pr- A T C C G T **Tr** A X 4' ←	52
→ 4' pr- X T A G G C A **Tr** T 2' 2' pr- A T C C G T **T** A X 4' ←	52
→ 4' pr- X T A G G C A **A** T 2' 2' pr- A T C C G T **Tr** A X 4' ←	46

Tryptamine may not be a potentially prebiotic substance, but it is a bio- molecule. Incorporation of this new p-RNA nucleoside with no donor- or acceptor-sites for the exclusive Watson-Crick pairing constitution shows that different rules control this sequentially selective model pairing system. The naive question about a natural use of similar compounds as pseudo bases or about any possible genotypic function of non-hydrogen-bonding nucleobases would lead to the basic problem of a quantifiable reach of the incremental optimisation in evolution with respect to a genetic code.

Acknowledgement

The authors are grateful for the expert technical assistance of Petra Ickstadt for the monomer synthesis, Daniela Kammann, Joachim Schmidt and Manfred Wildt for the synthesizer synthesis of the oligos.

References and Notes

(1) New address: GPC-AG, Fraunhoferstr. 20, D-82151 Martinsried / München, Germany. E-mail: tilmann.brandstetter@gpc-ag.com

(2) (a) Pitsch, S.; Wendeborn, S.; Jaun, B.; Eschenmoser A. *Helv. Chim. Acta* **1993**, *76*, 2161. (b) Schlönvogt, I.; Pitsch, S.; Lesueur, C.; Eschenmoser, A.; Jaun, B.; Wolf, R.M. *Helv. Chim. Acta* **1996**, *79*, 2316. (c) Micura, R.; Bolli, M.; Windhab, N.; Eschenmoser, A. *Angew. Chem.* **1997**, *109*, 899; *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 870.

(3) (a) Mucic, C. R.; Storhoff, J.; Mirkin, C. A.; Letsinger, R. L. *J. Am. Chem. Soc.*, **1998**, *120*, 12674. (b) Seeman, N. C. *Annu. Rev. Biomol. Struct.* **1998**, *27*, 225. (c) Holmlin, R. E.; Dandliker, P. J.; Barton, J. K. *Angew. Chem.* **1997**, *109*, 2831; *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2714.

(4) (a) Ishida, T.; Tomita, K.-I.; Inoue, M. *Arch. Biochem. Biophys*. **1980**, *200*, 492. (b) Ishida, T.; Doi, M.; Ueda, H.; Inoue, M.; Scheldrick, G.M. *J. Am. Chem. Soc.* **1988**, *110*, 2286.

(5) Schweitzer, B.; Kool E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863.

(6) Giannis, A.; Sandhoff, K. *Angew. Chem.* **1989**, *101*, 220; *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 218.

(7) Kuehne, M. E.; Huebner, J. A.; Matsko, T. H. *J. Org. Chem.* **1979**, *44*, 2477.

(8) Marky, L.A.; Breslauer, K.J. *Biopolymers*, **1987,** *26*, 1601.

(9) Krishnamurty, R.; Pitsch, S.; Minton, M.; Miculka, C.; Windhab, N.; Eschenmoser, A. *Angew. Chem.* **1996**, *108*, 1619; *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1537.

(10) Windhab, N.; Ohms, J.; Ackermann, Th. *Biophys. Chem.* **1993**, *47*, 225.

(11) Kool, E. *Chem. Rev.* **1997**, *97*, 1473.

(12) Matray, T.; Kool, E. *J. Am. Chem. Soc.* **1998**, *120*, 6191.

(13) **2**: ^{1}H NMR ($CDCl_3$, 300 MHz): 1.85-2.00, 2.14-2.28 (2 m, 2 x 1 H, C**H**$_2$CH$_2$NPhth), 2.70 (bs, 1 H, NH), 3.24-3.38, 3.66-3.86 (2 m, 5 H, CH$_2$C**H**$_2$NPhth, H-2a, H-2b, H-3), 6.62 (d, J = 8.0 Hz, 1 H, H-7), 6.66-6.72 (m, 1 H, H-5), 6.99 (app t, J = 7.5 Hz, 1 H, H-6), 7.14 (d, J = 8.0 Hz, 1 H, H-4), 7.64-7.74, 7.78-7.86 (2 m, 2 x 2 H, Phth). ^{13}C NMR($CDCl_3$, 75 MHz): 32.70, 36.10 (2 t, C-2, **C**H$_2$CH$_2$NPhth), 39.62 (d, C-3), 53.04 (t, **C**H$_2$NPhth), 109.65 (d, C-7), 118.74 (d, C-5), 123.25 (d, Phth), 123.92, 127.72 (2 d, C-4, C-6), 131.81 (s, C-3a), 132.14 (s, Phth), 133.99 (d, Phth), 151.26 (s, C-7a), 168.38 (s, C = O). Cal: C: 73.96, H: 5.52, N: 9.58; found: C: 73.89, H: 5.57, N: 9.55. MS (ES$^+$): 293 (MH$^+$, 100%)

(14) **3**: ^{1}H NMR ($CDCl_3$, 300 MHz): 1.64, 1.98, 2.19 (3 s, 3 x 3 H, Ac), 3.06 (t, J = 8.0 Hz, 2 H, C**H**$_2$CH$_2$NPhth), 3.81-4.00 (m, 4 H, H-5ax, H-5eq, C**H**$_2$NPhth), 5.13 (ddd, J = 2.5, 6.0, 10.5 Hz, 1 H, H-4), 5.36 (dd, J = 3.5, 9.5 Hz, 1 H, H-2), 5.71 (d, J = 9.5 Hz, 1 H, H-1), 5.74 (app t, J = 3.0 Hz, 1 H, H-3), 7.02 (s, 1 H, H-2), 7.04-7.10, 7.13-7.19 (2 m, 2 x 1 H, H-5, H-6), 7.33 (d, J = 8.0 Hz, 1 H, H-7), 7.58-7.66, 7.72-7.80 (2 m, 5 H, Phth, H-4). ^{13}C NMR($CDCl_3$, 75 MHz):20.23, 20.65, 20.87 (3 q, Ac), 24.41, 38.28 (2 t, CH_2CH_2), 63.53 (t, C-5), 66.24, 68.00, 68.64 (3 d, C-2, C-3, C-4), 80.33 (d, C-1), 109.79 (d, C-7), 113.95 (s, C-3), 119.33, 120.39, 122.04, 122.47 (4 d, C-4, C-5, C-6, C-7), 123.18 (d, Phth), 128.70, 132.17 (2 s, C-3a, Phth), 133.87 (d, Phth), 136.78 (s, C-7a), 168.24, 168.77, 169.44, 169.87 (4 s, C = O). Cal: C: 63.50, H: 5.15, N: .11; found.: C: 63.48, H: 5.16, N: 5.05. MS (ES$^+$): 566 (M+NH$_4^+$, 82%), 549 (MH$^+$, 74%), 114 (100%).

(15) **4**: ^{1}H NMR (MeOD, 300 MHz): 3.09 (app. t, J = 7.0 Hz, 2 H, C**H**$_2$CH$_2$NPhth), 3.64-3.98 (m, 5 H, H-4, H-5ax, H-5eq, C**H**$_2$NPhth), 4.05 (dd, J = 3.5, 9.5 Hz, 1 H, H-2), 4.22 (app t, J = 3.0 Hz, 1 H, H-3), 5.65 (d, J = 9.5 Hz, 1 H, H-1), 6.95-7.05, 7.09-7.16 (2 m, 2 x 1 H, H-5, H-6), 7.25 (s, 1 H, H-2), 7.44 (d, J = 8.0 Hz, 1 H, H-7), 7.60 (d, J = 8.0 Hz, 1 H, H-4), 7.74-7.84 (m, 4 H, Phth). ^{13}C NMR (d_6-DMSO, 75 MHz):

23.87, 37.79 (2 t, $\underline{C}H_2\underline{C}H_2$NPhth), 64.82 (t, C-5), 66.74 (d, C-4), 68.41 (d, C-2), 71.42 (d, C-3), 81.37 (d, C-1), 110.42 (d, C-7), 111.05 (s, C-3), 118.17, 119.21, 121.36, 122.92, 123.80 (5 d, C-2, C-4, C-5, C-6, NPhth), 127.86, 131.59 (2 s, C-3a, Phth), 134.27 (d, Phth), 136.62 (s, C-7a), 167.72 (s, C = O). MS (ES^-):457 ($M+OH^-+H_2O$, 49%), 439 ($M+OH^-$, 100%), 421 ($M-H^+$, 28%)

(16) **5**: ^{1}H NMR ($CDCl_3$, 300 MHz): 2.45, 2.70 (2 bs, 2 x 1 H, OH), 3.04 (t, J = 8.0 Hz, 2 H, $C\underline{H}_2CH_2$NPhth), 3.80-4.20 (m, 5 H, H-4, H-5ax, H-5eq, $C\underline{H}_2$NPhth), 4.63 (bs, 1 H, H-3), 5.46 (dd, J = 3.5, 9.5 Hz, 1 H, H-2), 6.03 (d, J = 9.5 Hz, 1 H, H-1), 7.08-7.31 (m, 5 H, H-2, H-5, H-6, Bz-m-H), 7.41-7.48 (m, 1 H, H-Bz-p-H), 7.50 (d, J = 8.0 Hz, 1 H, H-7), 7.64-7.79 (m, 7 H, Phth, H-4, Bz-o-H). ^{13}C NMR (d_6-DMSO, 75 MHz): 24.40, 38.22 (2 t, $\underline{C}H_2\underline{C}H_2$NPhth), 65.95 (t, C-5), 66.65 (d, C-4), 69.55 (d, C-3), 71.87 (d, C-2), 79.57 (d, C-1), 109.96 (d, C-7), 113.70 (s, C-3), 119.21, 120.21, 122.11, 122.41, 123.14 (5 d, C-2, C-4, C-5, C-6, NPhth), 128.28 (d, Bz), 128.58, 128.59 (2 s, C-3a, Bz), 129.62 (d, Phth), 132.05 (s, Phth), 133.81 (Bz), 136.97 (s, C-7a), 165.12, 168.29 (2 s, C = O). MS (ES^-): 525 ($M-H^+$, 12%), 421 ($M-PhCO^+$, 23%), 107 (100%).

(17) **6**: ^{1}H NMR ($CDCl_3$, 300 MHz): 2.64 (bs, 1 H, OH), 2.68 (dd, J = 5.0, 11.5 Hz, 1 H, H-5eq), 2.94 (dd, J = 7.5, 16.0 Hz, 1 H, $CH_2C\underline{H}_2$NPhth), 3.03 (dd, J = 8.0, 16.0 Hz, 1 H, CH_2CH_2NPhth), 3.67-3.74 (m, 1 H, H-5ax), 3.69, 3.70 (2 s, 2 x 3 H, OMe), 3.85 (t, J = 7.5 Hz, 2 H, $CH_2C\underline{H}_2$NPhth), 3.94 (ddd, J = 3.0, 5.0, 10.5 Hz, 1 H, H-4), 4.03 (dd, J = 3.5, 9.0 Hz, 1 H, H-2), 5.51 (d, J = 9.0 Hz, 1 H, H-1), 5.86 (bs, 1 H, H-3), 6.68-7.66 (m, 25 H), 8.19-8.30 (m, 2 H). ^{13}C NMR($CDCl_3$, 75 MHz): 24.16, 38.80 (2 t, $\underline{C}H_2\underline{C}H_2$NPhth), 55.25, 55.26 (2 q, Ome), 65.58 (t, C-5), 68.29, 69.19, 73,83 (3 d, C-2, C-3, C-4), 83.03 (d, C-1), 87.31 ($\underline{C}Ar_3$)110.03 (d, C-7), 113.37, 113.47 (2 d), 113.53 (s, C-3), 118.95, 120.20, 122.28, 122.31, 123.10, 127.07, 128.02, 128.08, 128.68 (9 d), 128.74 (s), 130.02, 130.19, 130.22 (3 d), 130.37, 131.95 (2 s), 133.40, 133.83 (2 d), 135.98, 136.14, 136.56, 145.12, 158.82, 166.76, 168.52 (7 s, C-7a, 2 $\underline{C}$OMe, 2 C = O).

(18) **7**: ^{1}H NMR ($CDCl_3$, 300 MHz): characteristic peaks:2.42, 2.53 (2 dd, J = 5.0, 11.0 Hz, 2 H, 2 H-5eq), 3.76, 3.77, 3.78, 3.79 (4 s, 4 x 3 H, OMe), 5.70, 5.73 (2 d, J = 9.0 Hz, 2 H, 2 H-1), 6.16, 6.29 (2 bs, 2 H, 2 H-3). ^{31}P NMR ($CDCl_3$): 150.6, 151.0

Article Identifier:
1437-2096,E;1999,0,S1,0940,0944,ftx,en;W07199ST.pdf