Tetrahedron Letters,Vol.25,No.45,pp 5197-5200,1984 0040-4039/84 $3.00 + .00
Printed in Great Britain

SYNTHESIS OF A NATURALLY OCCURRING NUCLEOPEPTIDE FRAGMENT VIA A PHOSPHOTRIESTER APPROACH

Cecile Schattenkerk, C.T.J. Wreesmann, Marja J. de Graaf, G.A. van der Marel and J.H. van Boom*

Gorlaeus Laboratory, P.O. Box 9502, 2300 RA Leiden, The Netherlands

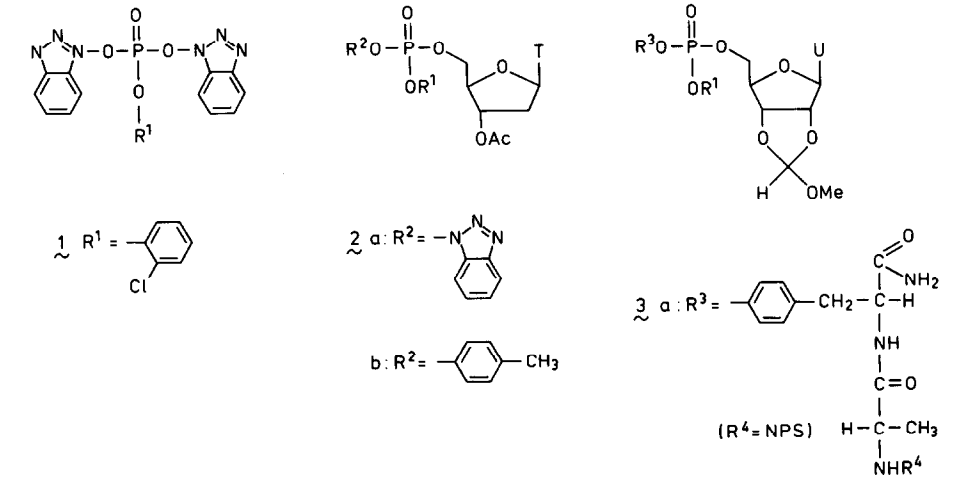
Abstract. The phosphorylating agent 2-chlorophenyl-O,O-bis[1-benzotriazolyl]phosphate has been used for the introduction of a phosphodiester bond between the hydroxyl group of the tyrosine containing dipeptide (H-Ala-Tyr-NH_2) and the 5'-OH of the RNA dimer UpU.

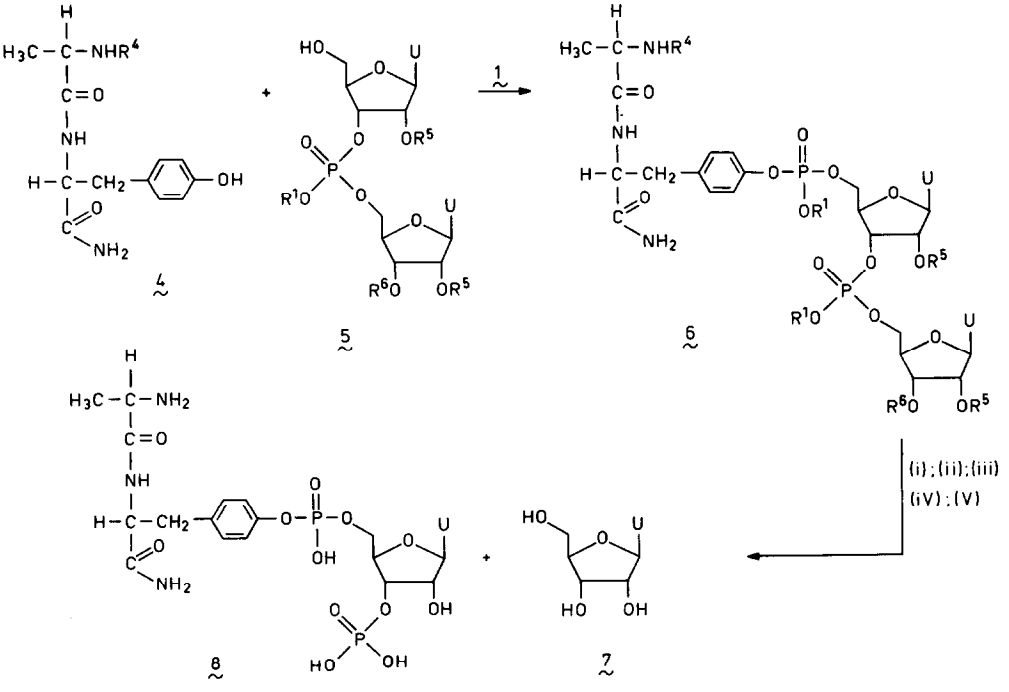
Nucleoproteins are naturally occurring biopolymers which consist of proteins covalently linked to nucleic acids (DNA or RNA). So far, the results obtained from degradation studies of nucleoproteins revealed that, in most cases, the 5'-OH of the nucleic acids component is linked by a phosphodiester bond to either the hydroxyl group of L-tyrosine or L-serine in the protein moiety. For instance, L-serine in protein p3 is linked[1] through a phosphodiester bond to the 5'-OH of the DNA of phage ϕ29. It has also been established[2] that virion RNA of polio virus type 1 is similarly linked to L-tyrosine in the genome-protein VPg: the 5'-end of polio genome RNA can now be presented as VPg(Tyr-O)-p-UUAAAACAG. Recently[3], we showed that the bond in the bacteriophage ϕX 174 gene A protein-DNA complex is also a tyrosyl-5'-phosphate ester.

As part of a programme to study the (bio)chemical properties and function of nucleoproteins, we report on the synthesis of a small nucleopeptide fragment (see compound 6: R^1=R^4=R^5=R^6=H) of polio genome RNA[2] in which the dipeptide H-Ala-Tyr-NH_2 is linked, via a phosphodiester bond, to the 5'-end of the RNA dimer UpU.

In principle several routes can be devised for the synthesis of the nucleopeptide fragment 6. A route, however, in which a phosphotriester linkage is introduced between a properly-protected tyrosine containing dipeptide (e.g. 4) and the 5'-end of a properly-protected RNA dimer (e.g. 5) would be, taken into account the present state-of-the-art in peptide and nucleic acids chemistry, the most economical one. Nonetheless, a successful synthesis of 6, according to the above direct phosphotriester approach, demands not only the availability of an effective bifunctional phosphorylating agent but also the selective removal, at the final stage of the synthesis, of all protecting groups. With respect to the phosphorylating agent, we amply demonstrated[4] that the agent 1 (R^1=2-chlorophenyl) could be applied successfully for the formation of phosphodiester linkages. However, in this particular case, the use of agent 1 will afford an intermediate triester carrying two aryl functions with a difference of approximately two between their respective pK values. In order to find out whether or not the 2-chlorophenyl group (2-chlorophenol; pK 8.5) could be removed selectively in the presence of the tyrosine moiety, we prepared firstly the model compound 2b (R^1=2-chlorophenyl; R^2=4-methylphenyl). Derivative 2b was obtained[4b] by phosphorylation of 3'-O-acetyl-thymidine with 1 to afford intermediate 2a. Treatment of 2a with 4-methylphenol, in the presence of N-methylimidazole, gave, after work-up and purification, 2b [^{31}P-NMR data ($CDCl_3$); δ -11.70 and -11.77 ppm] in 80% yield. Compound 2b was now treated with N^1,N^1,N^3,N^3-tetramethylguanidinium syn-4-nitrobenzaldoximate (oximate)[5] to give, after work-up followed by complete removal of the acetyl group with aqueous ammonia, a homogeneous 5'-O-(aryl)phosphate derivative of thymidine.

^{31}P-NMR spectroscopy[6] of this product revealed the presence of only one resonance with a δ-value of -4.16 ppm. On the other hand, ammonolysis of 2 (R^1=2-chlorophenyl)[7] in which R^2 is replaced by a cyanoethyl group will afford the 5'-O-(2-chlorophenyl)phosphate derivative of thymidine. Indeed, ^{31}P-NMR spectroscopy[6] of the product showed the presence of one resonance at -4.40 ppm.





(i): Oximate. (ii): $N(Et)_3$/MeOH) H_2O. (iii): MeOH/HCl (pH 2.5). (iV) 0.01 NHCl (pH 2). (V): RNase.

From these results we may conclude that the removal of the 2-chlorophenyl (R^1) group from 2b proceeds with a high degree of selectivity. We now treated the uridine derivative 3[8] [R^1=2-chlorophenyl; R^4=2-nitrophenylsulfenyl (NPS)] with oximate, followed by removal of the NPS and 2',3'-O-methoxymethylidene groups with acid (see later). T.l.c.-analysis of the crude product, obtained after oximate treatment, indicated the absence of the dipeptide 4 (R^4=NPS). Furthermore, ^{31}P-NMR analysis[6] of the completely deblocked and purified (Sephadex A-25) product 3 showed the presence of one resonance at -4.23 ppm. The above results indicate that also in this case the removal of the 2-chlorophenyl group is highly selective.

The synthesis of the fully-protected nucleopeptide fragment 6 was realized by phosphorylation of dimer 5[4e] [R^1=2-chlorophenyl; R^5=tetrahydropyranyl (THP); R^6=acetyl] in pyridine with agent 1. After 30 min at 20°C, a slight deficiency of dipeptide 4[9] (R^4=NPS) together with an excess of N-methylimidazole were added. Work-up of the mixture, after 1 h at 20°C, followed by purification of 6 by silicagel column chromatography gave homogeneous 6 in 80% yield from 5. Synthesis of 6 could also be accomplished by adding 1 to dipeptide 4 followed by the addition of dimer 5 (yield 78% from 4). ^{31}P-NMR analysis[6] of 6 revealed two sets of resonances at 7.1-7.2 and 12.0-12.2 ppm. Fully-protected 6 was completely deblocked by the following four steps procedure. Firstly, the 2-chlorophenyl (R^1) groups were deblocked by oximate. Complete removal of the acetyl (R^6) group was effected by treating 6 (R^1=H) with triethylamine-methanol-water[10]. Finally the NPS (R^4) and the THP (R^5) groups were removed by acidic hydrolysis: first, hydrochloric acid in methanol (pH 2.5) for 3 h at 20°C, followed by aqueous hydrochloric acid (pH 2) for 14 h at 20°C. Crude 6 was purified (Sephadex A-25) to afford 6, after lyophilization, as a colourless fluffy solid. ^{1}H- (300 MHz)[11] and ^{31}P-NMR data of 6 (R^1=R^4=R^5=R^6=H) were in accordance with the proposed structure. The identity of 6 was further corroborated by monitoring the enzymatic digestion of 6 with the endonucleolytic enzyme ribonuclease using ^{31}P-NMR spectroscopy. It can be seen (see under a and c in Fig. 1) that the nucleopeptide 6 is completely digested by RNase. T.l.c.-analysis of the digest showed that one of the products had the same R_f-value as uridine (i.e., 7). ^{1}H- and ^{31}P-NMR spectroscopy of the other product, which was obtained after purification (Sephadex A-25), indicated that its structure was in agreement with the expected product 8. It is interesting to note that we observed the precursor of 8 (i.e., the 2',3'-cyclic phosphate derivative of 8: see resonance 4 under b in Fig. 1), the presence, however, of an intermediate resonance at 19.5 ppm (see peak 1 under b in Fig. 1) cannot be explained for the time being.

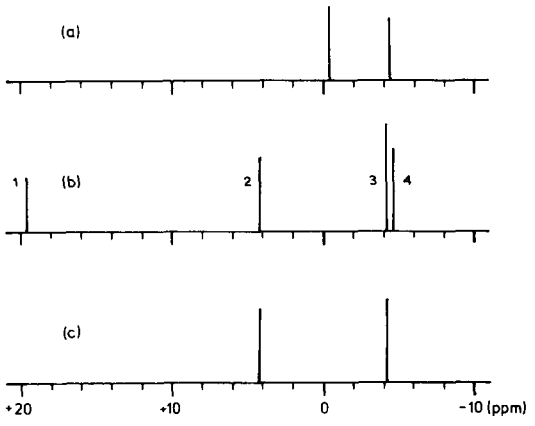


Figure 1. Monitoring by ^{31}P-NMR spectroscopy of the enzymatic digestion of deblocked 6 with RNase. a) At t=0 min. b) at t=10 min. c) at t=90 min.

In conclusion, the data presented here show that phosphorylating agent 1 is very suitable for the introduction of a phosphotriester linkage between properly-protected tyrosine containing peptides and ribonucleic acids. Further, the same strategy can also be followed for the

preparation of nucleopeptides in which a tyrosine containing peptide is similarly linked to a DNA fragment.

References and footnotes

1. R.H. Hermose and M. Salas, Proc. Natl. Acad. Sci. USA, 77, 6425 (1980).
2. V. Ambros and D. Baltimore, J. Biol. Chem. 253, 5263 (1978). P.G. Rothberg, T.J.R. Harris, A. Nomoto and E. Wimmer, Proc. Natl. Acad. Sci. USA, 75, 4868 (1978).
3. A.D.M. van Mansfeld, H.A.A.M. van Teeffelen, P.D. Baas, G.H. Veeneman, J.H. van Boom and H.S. Jansz, FEBS Letters, in press.
4. a) G.A. van der Marel, C.A.A. van Boeckel, G. Wille and J.H. van Boom, Tetrahedron Lett. 3887 (1981); b) C.A.A. van Boeckel, G.A. van der Marel, G. Wille and J.H. van Boom, Chem. Letters, 1752 (1981); c) G.A. van der Marel, C.A.A. van Boeckel, G. Wille, J.H. van Boom and C.F. Hoyng, Chemical and Enzymatic Synthesis of Gene Fragments (Ed. by H.G. Gassen and A. Lang), p. 53, Verlag Chemie, Weinheim (1982); d) G.A. van der Marel, C.A.A. van Boeckel, G. Wille and J.H. van Boom, Nucleic Acids Res., 10, 2337 (1982); e) C.T.J. Wreesmann, A. Fidder, G.A. van der Marel and J.H. van Boom, Nucleic Acids Res., 11, 8389 (1983); f) J.E. Marugg, M. Tromp, P. Ihurani, C.F. Hoyng, G.A. van der Marel and J.H. van Boom, Tetrahedron, 40, 73 (1984).
5. C.B. Reese and L. Zard, Nucleic Acids Res., 9, 4611 (1981).
6. External reference 80% H_3PO_4.
7. This compound was prepared by treating 2a with cyanoethanol in the presence of pyridine. Yield 80% from 2a: ^{31}P-NMR data ($CDCl_3$); δ -6.77 and -7.12 ppm.
8. Compound 3 was prepared by phosphorylation of 2',3'-O-methoxymethylidene-uridine with 1 followed by the addition of dipeptide 4 (R^4=NPS) and 1-methylimidazole. Yield 82%, ^{31}P-NMR data ($CDCl_3$); δ -10.85 and -10.94 ppm.
9. Dipeptide 4 (R^4=NPS) was prepared as follows. Tyrosylamide [M. Brenner et al., Helv. Chim. Acta, 36, 1109 (1953)] in DMF was coupled, according to the HOBT/DCC method [W. König et al., Chem. Ber., 103, 788 (1978)], with N-(2-nitrophenylsulfenyl)-alanine [L. Servas et al., J. Amer. Chem. Soc., 35, 3660 (1963)] to afford 4 in 75% yield. M.p. 136-140°C; $[\alpha]_D^{23}$ -23.1 (c 1, MeOH).
10. The dipeptide 4 (R^4=NPS) did not racemize (optical rotation measurements) under the basic conditions of steps (i) and (ii).
11. ^{1}H-NMR (300 MHz) data (D_2O): δ 7.86 (d, J 8.1 Hz) and 7.73 (d, J 8.1 Hz) 2 x H6 (uridine); 7.19 (d, J 8.6 Hz) and 7.09 (d, J 8.6 Hz) 4H (phenyl, Tyr); 5.97 (d, J 6.0 Hz) and 5.88 (d, J 3.8 Hz) 2 x H1' (uridine); 5.83 (d, J 8.1 Hz) and 5.71 (d, J 8.1 Hz) 2 x H5 (uridine); 3.92 (q, J 7.1 Hz) -CH (Ala); 3.08 (dd, J 6.3 and 14.0 Hz) and 2.95 (dd, J 8.8 and 14.0 Hz) $-CH_2-$ (Tyr); 1.42 (d, J 7.1 Hz) $-CH_3$ (Ala).

(Received in UK 23 July 1984)