# SEQUENCE DETERMINATION OF Gp-RICH OLIGONUCLEOTIDES BY MEANS OF THE KETHOXAL MODIFICATION

W. MIN JOU and W. FIERS

*Laboratory of Molecular Biology, University of Ghent Ledeganckstraat, 35, B-9000 Ghent, Belgium*

## 1. Introduction

The first steps in RNA sequence analysis are usually degradation of the RNA with the base specific ribonuclease $T_1$ which hydrolyzes only Gp-N diester bonds, and with pancreatic ribonuclease which cleaves after pyrimidine nucleotides. The sets of $T_1$ and P-oligonucleotides are separated and the sequence of each product is determined. For this purpose, the oligonucleotide is degraded with the complementary enzyme (pancreatic ribonuclease for $T_1$-oligonucleotides and vice versa). Further characterization was initially almost totally dependent on partial degradation with venom and spleen exonuclease [1–3]. This method poses several problems of its own: partial digestion conditions are critical, the radioactivity is distributed over several intermediates, products are often also produced by non-specific endonucleolytic cleavage.

Now, RNA sequencing methodology has become much more versatile, especially for the structure determination of $T_1$-oligonucleotides. $U_2$ ribonuclease, a purine-specific enzyme, will create $(Pyp)_n$ Pup sequences [4]. A ribonuclease from *Physarum polycephalum* gives rise to $(Cp)_n$ Np pieces [5]. Although they lack a precise base specificity, spleen acid ribonuclease ([6], our unpublished results) and silkworm endonuclease [7,8] also give useful products. Uracil and guanine residues can be modified by a

*Abbreviations*: Pup: purine nucleotide; Pyp: pyrimidine nucleotide; $T_1$-oligonucleotide: product with the structure $(Np)_n$Gp released by ribonuclease $T_1$ digestion; P-oligonucleotide: product released by pancreatic ribonuclease with the structure $(Pup)_n$Pyp.

water-soluble carbodiimide derivative, thus making pancreatic RNase C-specific [9].

Sequencing of P-oligonucleotides, however, is still dependent on the use of exonucleases whenever $T_1$ RNase digestion does not give the complete answer, although a simpler and more sensitive method has been presented for the study of spleen exonuclease digests [10].

Recently we have drawn attention to the fact that Gp-rich sequences, especially with the G-residues forming a consecutive run, are obtained in variable and submolar quantity due to aggregation [11]. On columns, some products were completely overlooked [12,13] due to their low recovery. Such polypurine tracts are scarcely broken down by exonucleases for the same reason, even after heating [13]. Thus one ends up with the situation that G-rich P-products are not only hard to obtain in molar yield but also extremely difficult to sequence. Hitherto it was impossible to solve conclusively the structure of the product GpGpGpGpApGpUp [13].

Kethoxal, a reagent specific for G-residues [14], has been used for the selective modification of exposed guanine sites in RNA molecules [15,16] and in the 30S ribosome [17]. We have used kethoxal modification followed by ribonuclease $U_2$ digestion to obtain information of polypurine sequences. After reaction of the $(Pup)n$ Pyp tract with kethoxal, the modified G-residues are resistant during the subsequent ribonuclease $U_2$ treatment. The $U_2$ enzyme thus gives rise to $(Gp)_n$ Ap sequences and the 3'-terminus $(Gp)_n$ Pyp which are easy to characterize. The usefulness of the method is demonstrated here with some G-rich P-oligonucleotides derived from ^{32}P-labeled MS2 RNA which posed

problems upon exonuclease digestion. This method, however, may also provide useful information when applied to less G-containing polypurine sequences. A somewhat similar method, but using the $T_2$ enzyme instead of $U_2$ ribonuclease, has been described recently [18].

## 2. Materials and methods

The P-oligonucleotides used in this study were derived from $^{32}$P-labeled Bacteriophage MS2 RNA. They were isolated from mini-fingerprints of pancreatic ribonuclease digests [11], either from total MS2 RNA or from pure fragments derived from the end of the A-protein gene. Isolation of these fragments by partial digestion of the total RNA followed by successive polyacrylamide gel electrophoreses has been described previously [19].

Pancreatic ribonuclease was bought from Sigma Chemical Co., St. Louis, Mo. $U_2$ ribonuclease was a generous gift from Sankyo Co., Tokyo, Japan, and a kethoxal sample was kindly sent to us by Dr G. Underwood (Upjohn Company, Kalamazoo, Mich.) as a 60% aqueous solution. The P-oligonucleotides used for sequencing had been characterized by $T_1$ ribonuclease (Sankyo).

Spots, always containing carrier RNA from the homochromatographic step in the fingerprinting procedure, were eluted with triethylamine carbonate, pH 10.0, as described previously [11]. The material was dried in a plastic tube (1.5 ml Eppendorf microcentrifuge tube) in the 37°C room. After one or two washings with water the oligonucleotide was ready for further sequencing.

Modification of material obtained from one spot on the homochromatogram was carried out in a 30 μl solution containing 0.10 M sodium acetate buffer, pH 5.0, 0.001 M EDTA and 0.15 M kethoxal and incubation was for 1 h at 37°C. 0.1 volume of 20% potassium acetate and 2 volumes ethanol were then added to the solution which was kept at −20°C for 4 h. The RNA precipitate was collected by centrifugation, washed with ice-cold 70% ethanol and dried. Digestion with $U_2$ ribonuclease was done in 5 μl solution containing 0.1 M sodium acetate buffer, pH 4.5, 0.001 M EDTA, 1 μg/μl gelatine, and 0.08 units ribonuclease $U_2$, for 8 or 15 h at

37°C. The digest was separated on a minifingerprint [11]. The spots were visualized by autoradiography and their composition was determined after alkaline hydrolysis. G-residues are demodified during this treatment since the modification reaction is reversed at slightly alkaline pH [14].

## 3. Results and discussion

A spot with composition (Gp₃,ApGp) Up, isolated from a pancreatic ribonuclease minifingerprint of complete MS2 RNA (component No. 25 in fig.2 from [11]), was treated with kethoxal and $U_2$ ribonuclease as described in Materials and methods (8 h incubation with $U_2$ enzyme). The reaction mixture consisted mainly of three products (fig.1): GpGpAp (spot No. 1), GpGpUp (spot No. 2) and undegraded material (spots No. 3 and 4; most likely one of these is incompletely modified). Base analyses of the spots were within 10% accuracy. Some minor components were also present. However, on a molar basis they represented less than 10% of the material found in the products 1 and 2. They were most likely derived from contaminating material, which is to be expected due to streaking in that G-rich part of the complex fingerprint. From the nature of the degradation products we conclude that the structure of the P-oligonucleotide is GpGpApGpGpUp. This is the only sequence having this composition known to be present in the MS2 RNA molecule [21], and previously it had taken us many trials before its structure could be established by means of spleen exonuclease.

A heptanucleotide having the composition (Gp₄,ApGp)Up was obtained from minifingerprints of pancreatic RNase digests of fragments derived from the end of the A-protein gene [19]. The reaction (15 h incubation with $U_2$ ribonuclease) gave the following results (fig.2): spot No. 1: GpAp; spot No. 2: GpGpGpGpUp; spot No. 3: undigested material. This leads unambiguously to the structure GpApGpGpGpGpUp. This sequence is part of the total structure of the A-protein gene, as recently published [20].

Previous attempts to sequence the latter heptanucleotide by partial spleen exonuclease treatment had all been unsuccessful. However, repeated partial spleen and venom exonuclease
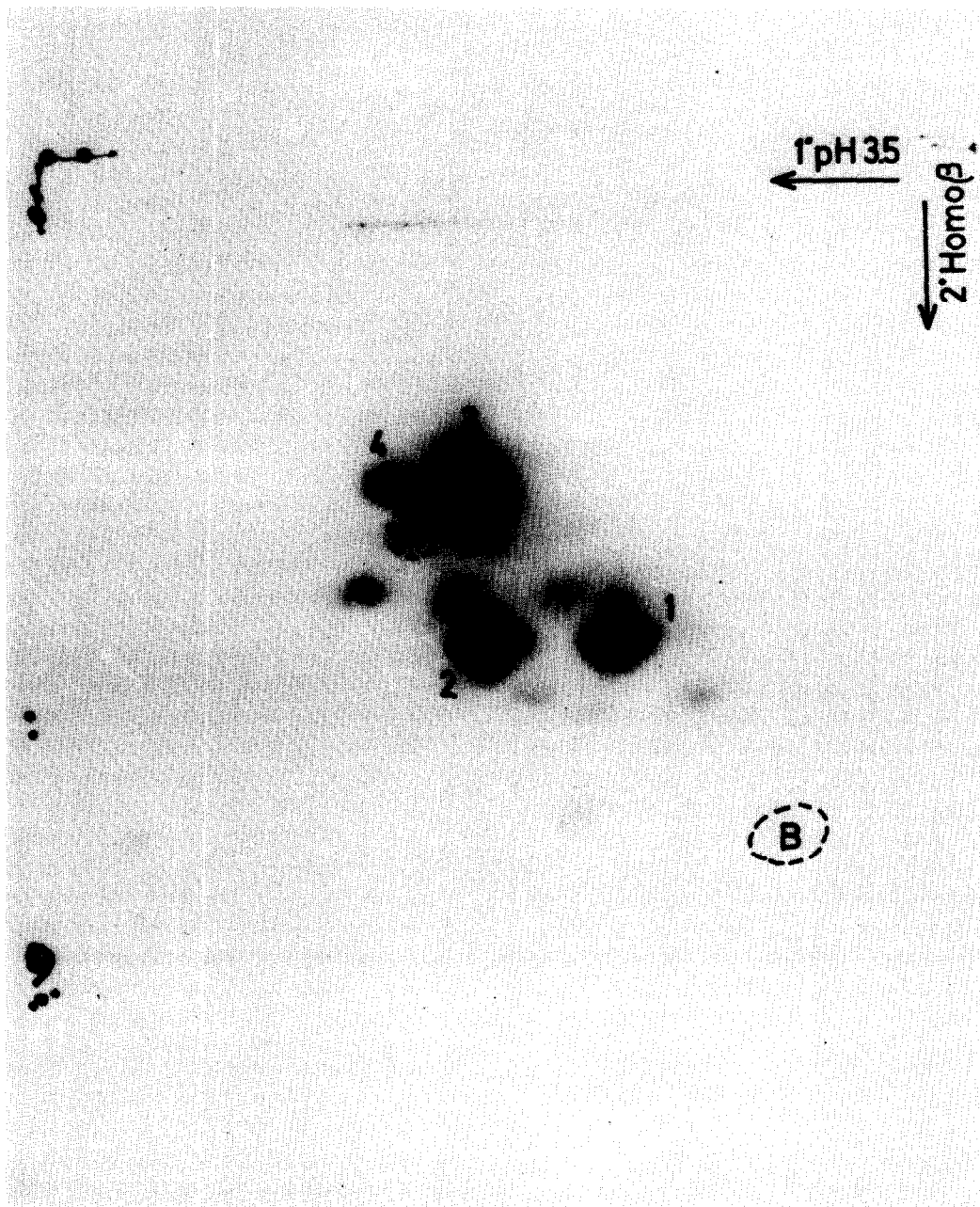
Fig.1. Two-dimensional separation of the reaction products obtained after kethoxal modification and U$_2$ ribonuclease degradation of component (Gp$_3$,ApGp)Up.
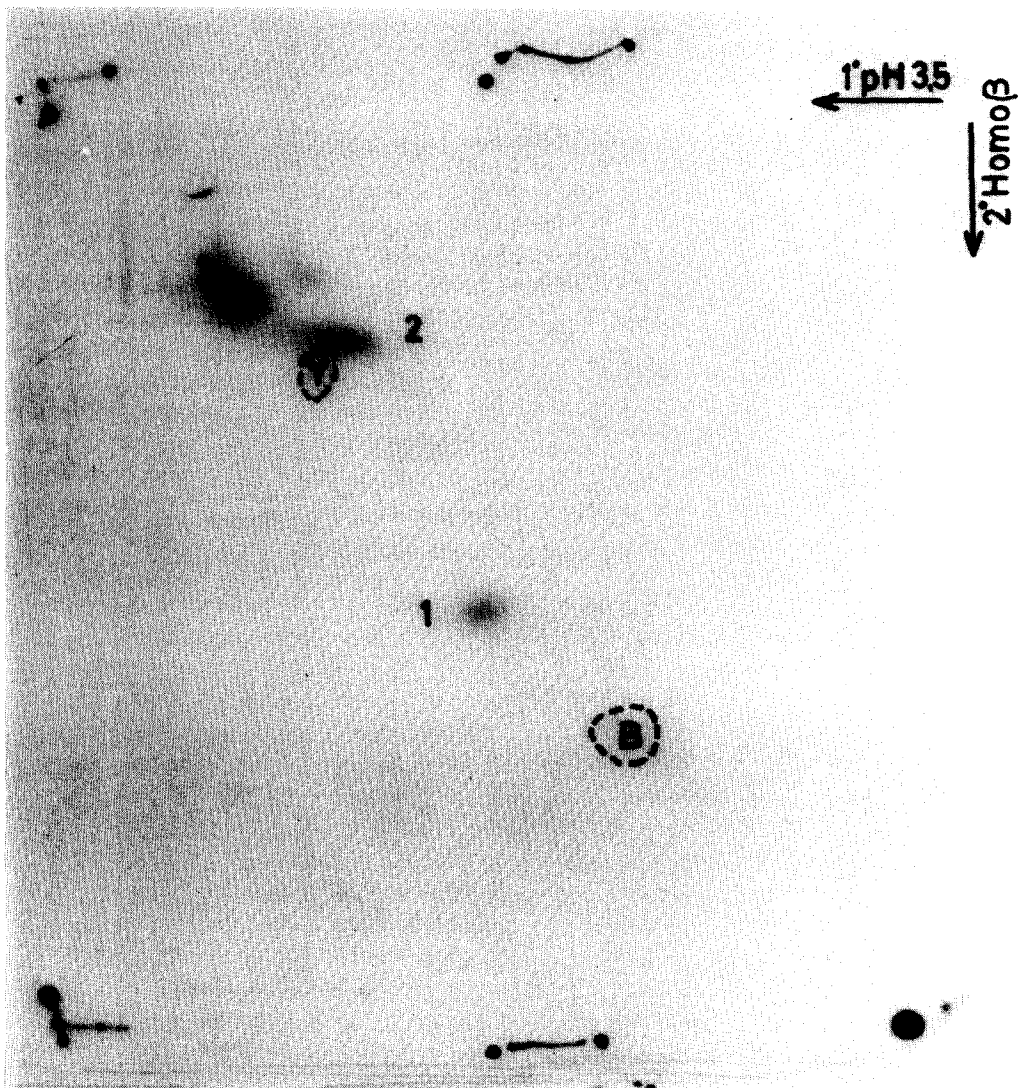
Fig.2. Two-dimensional separation of the reaction products obtained from component $(Gp_4,ApGp)Up$ derived from the end of the A-protein gene.

studies on the assumed unique product with this composition derived from pancreatic ribonuclease digests of *complete* MS2 RNA had led to a tentative sequence GpGpGpGpApGpUp [13] and this structure was initially adopted for the P-oligo-nucleotide present in the terminal part of the A-protein gene [19]. During further sequencing of MS2 RNA, however, a component with the same $T_1$-composition was isolated from a region

in the RNA polymerase gene [21] and its sequence was shown by spleen exonuclease degradation to correspond to the one derived from pancreatic ribonuclease digests of total MS2 RNA (A. Raeymaekers and G. Volckaert, unpublished). We have applied the kethoxal-$U_2$ ribonuclease method also to the mixture of the two heptanucleotides obtained from a minifingerprint of a pancreatic RNase digest of total MS2 RNA. From the degrada-

tion products it follows (data not shown) that both sequence isomers GpApGpGpGpGpUp and GpGpGpGpApGpUp are indeed present in roughly equimolar amounts. It is not clear in which proportion the two sequences were recovered in earlier experiments by column chromatography, but only pieces corresponding to the sequence GpGpGpGpApGpUp were detected upon exonuclease treatment [13]. Either the other isomer was virtually missing or else it might have been more refractory to both spleen and venom phosphodiesterase treatment and so remained undetected in the mixture, in agreement with other similar observations (R. Contreras, unpublished).

In conclusion, we have shown that the structure of G-rich sequences which have proven extremely difficult to determine with other techniques, can be solved in a very simple way by kethoxal modification followed by $U_2$ ribonuclease degradation. The method can give valuable information for many other products derived from pancreatic RNase digests and make exonuclease treatments superfluous in many cases.

## Acknowledgement

## References

[1] Holley, R. W., Madison, J. T. and Zamir, A. (1964) Biochem. Biophys. Res. Comm. 17, 389–394.

[2] Sanger, F., Brownlee, G. G. and Barrell, B. G. (1965) J. Mol. Biol. 13, 373–398.

[3] Brownlee, G. G. and Sanger, F. (1967) J. Mol. Biol. 23, 337–353.

[4] Arima, T., Uchida, T. and Egami, F. (1968) Biochem. J. 106, 609–613.

[5] Contreras, R. and Fiers, W. (1975) Anal. Biochem. 67, 319–326.

[6] Jeppesen, P. G. N. (1971) Biochem. J. 124, 357–366.

[7] Mukai, J. and Soeta, E. (1967) Biochim. Biophys. Acta 138, 1–9.

[8] Pinkerton, T. C., Paddock, G. and Abelson, J. (1973) J. Biol. Chem. 248, 6348–6365.

[9] Ho, N. W. Y. and Gilham, P. T. (1967) Biochem. 6, 3632–3639.

[10] Volckaert, G. and Fiers, W. (1974) Anal. Biochem. 62, 573–583.

[11] Volckaert, G., Min Jou, W. and Fiers, W. (1976) Anal. Biochem., in the press.

[12] Min Jou, W. and Fiers, W. (1969) J. Mol. Biol. 40, 187–201.

[13] Haegeman, G., Min Jou, W. and Fiers, W. (1971) J. Mol. Biol. 57, 597–613.

[14] Staehelin, M. (1959) Biochim. Biophys. Acta 31, 448–454.

[15] Litt, M. (1969) Biochem. 8, 3249–3253.

[16] Bellemare, G., Jordan, B. R., Rocca-Serra, J. and Monier, R. (1972) Biochimie 54, 1453–1466.

[17] Noller, H. F. (1974) Biochem. 13, 4694–4703.

[18] Willems, R., Avdonina, T., Lund, A. and Kisselev, L. L. (1974) Nucl. Acids Res. 1, 1201–1207.

[19] Contreras, R., Ysebaert, M., Min Jou, W. and Fiers, W. (1973) Nature New Biol. 241, 99–101.

[20] Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymaekers, A., Volckaert, G., Ysebaert, M., Van de Kerckhove, J., Nolf, F. and Van Montagu, M. (1975) Nature 256, 273–278.

[21] Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Vandenberghe, A., Volckaert, G. and Ysebaert, M. (1976) Nature 260, 500–507.