Journal of Chromatography, 131 (1977) 169–177

CHROM. 9497

SEQUENCE ANALYSIS OF SYNTHETIC OLIGONUCLEOTIDES BY HIGH-PERFORMANCE LIQUID ANION-EXCHANGE CHROMATOGRAPHY

J. H. VAN BOOM and J. F. M. DE ROOY
Gorlaeus Laboratories, Leiden University, P.O. Box 75, Leiden (The Netherlands)
(Received June 4th, 1976)

SUMMARY

A simple procedure for the sequential analysis of small oligonucleotides is reported. The method is based on the simultaneous identification and quantitation of monomers released by venom phosphodiesterase digestion of oligonucleotides using high-performance anion-exchange chromatography on Permaphase AAX at room temperature and by applying isocratic elution conditions. In this way, the correct sequence of five oligomers, *e.g.*, r-ACCUCC, r-CUGUU, r-AGGA, d-ATTACC and d-GGTAAT, could easily be established unambiguously.

INTRODUCTION

During an extensive programme on the synthesis of nucleic acids[1–3], there was a need to develop a technique for the sequence analysis of synthetic oligonucleotides using high-performance liquid chromatography (HPLC). The procedure described in this paper is based on a method, originally introduced by Holley *et al.*[4], which consists in partial digestion of an oligonucleotide with venom phosphodiesterase (PDE) followed by isolation and further analysis of the enzymatically obtained fragments by conventional techniques.

Our method differs from Holley *et al.*'s approach in that the identities and amounts of the monomers, as released in time by the action of PDE on oligonucleotides, are determined simultaneously by means of HPLC.

HPLC has been used successfully by several workers[5–12] for the identification and quantitation of nucleotides and nucleosides, but no attempt has ever been made, to our knowledge, to use HPLC for the sequential analysis of oligonucleotides. We report here a simple procedure that enabled us to derive the correct sequence of five synthetic oligomers.

MATERIALS AND METHODS

Chemicals and enzymes

The oligomers used were prepared according to the phosphotriester approach[1–3]. Nucleotides and nucleosides were purchased from Waldhof (Mannheim,

(G.F.R.). A standard solution (1 mg/ml) of venom phosphodiesterase (E.C. 3.1.4.1) from Boehringer (Mannheim, G.F.R.) was used, without purification, for the enzymatic digestions.

Apparatus

UV spectra were measured with a Cary C15 recording spectrometer and absorbances with a Zeiss PMQ 11 spectrophotometer. The chromatographic system employed was a Hupe-Busch/Hewlett-Packard 1010A liquid chromatograph equipped with a gradient mixing system and a UV absorption detector (254 nm). The photometer output (10 mV) was displayed on a flat-back recorder (Leeds & Northrup) and the peak areas were recorded by a Becker 7021 integrator. High-performance anion-exchange chromatography was performed with the strong anion-exchange resin Permaphase AAX[19] (DuPont, Wilmington, Del., U.S.A.) dry-packed into a stainless-steel column (1 m × 2.1 mm).

Isocratic elution of the monomers (nucleosides and nucleotides) was effected by buffer A (0.005 *M* KH_2PO_4, pH 4.5). Buffer B was composed of 0.05 *M* KH_2PO_4, 0.5 *M* KCl, pH 4.5. The buffers were prepared from a saturated stock solution (1 *M*) of purified[7] KH_2PO_4 by dilution with glass-distilled water to the required molarities. The pH was adjusted to 4.5 with H_3PO_4 (Baker, analyzed reagent). KCl was an analyzed reagent from Baker (Deventer, The Netherlands).

Gradient elution was performed by building up a linear gradient starting with buffer A and applying 3% of buffer B per minute. A flow-rate of 1 ml/min at a pressure of 70 kP/cm² was standard, resulting in a total elution time of 12 min.

RESULTS AND DISCUSSION

A practical sequence analysis of oligonucleotides based on the action of the enzyme venom PDE has to fulfil the following demands: (1) the enzyme must have no base specificity; (2) the enzyme must be readily available and not be contaminated with other enzyme activities; (3) inactivation of the enzyme must not involve time-consuming and complicated procedures; (4) the characterization of the monomers as released, in a step-wise process, from the oligomers by the enzyme must be rapid and reproducible; (5) quantitation of the monomers must be rapid, accurate and reliable.

Enzyme

The enzyme PDE has been widely studied and extensively purified[13–18]. From these studies it is apparent that PDE is exonucleolitic (the digestion of oligonucleotides starts at the righthand-end and proceeds towards the left producing the mononucleoside 5'-phosphates) and devoid of base specificity.

Inactivation of PDE could be accomplished by decreasing the pH to 5 followed by storage at −20°. The efficacy of this process was demonstrated by the following experiment.

The pentaribonucleotide r-CUGUU was subjected, under the conditions described under *Sequence analysis*, to PDE digestion. After 13 min at 20°, a sample (20 μl) was withdrawn from the digest and diluted with KH_2PO_4 (80 μl, 0.2 *M*, pH 3.5). The diluted mixture (pH 5.5) was used for three experiments: (1) a sample (5 μl) was withdrawn from the mixture and immediately analyzed by HPLC; (2) part of the

TABLE I

COMPOSITION OF PDE DIGEST OF r-CUGUU AFTER INACTIVATION OF THE ENZYME

Experiment No.	*Molar fractions of nucleotides and nucleoside*		
	pU	*pG*	*C*
1	1.95	0.64	0.06
2	1.95	0.62	0.08
3	2.04	0.80	0.10

mixture was frozen (liquid nitrogen) and kept at −20° for 24 h; (3) the remaining mixture was left at 20° for 24 h. The results obtained after HPLC analysis, using guanosine 3′-phosphate as internal standard, are given in Table I.

The results of experiments 1 and 2, which are within the limits of the experimental error, show clearly that a combination of decreasing the pH and freezing of the digestion mixture is essential for effective inactivation of PDE.

The presence of other enzyme activity in the PDE batch was corroborated as follows. Adenosine 5′-phosphate (3 absorbance units) in Tris-HCl buffer (0.3 ml, 0.1 *M*, 0.01 *M* $MgCl_2$, pH 8.9) was treated with PDE (6 μl of standard solution) for 24 h at 37°. Analysis of the solution by HPLC revealed the presence of adenosine (5.4%); this finding being indicative of the presence of alkaline phosphatase and/or 5′-nucleotidase activity. A control experiment revealed that the substrate was stable under the reaction conditions.

The absence of endonuclease activity was proved by the exclusive formation after PDE digestion of ribo- and deoxyoligonucleotides, of nucleoside 5′-phosphates and nucleosides.

Characterization

The identity of the four common d-nucleoside 5′-phosphates, obtained after digestion of d-oligonucleotides with PDE, can be easily and unambiguously derived from their retention times (Table II). The same procedure is also applicable to the four corresponding ribonucleotides. The nucleotides are rapidly and well separated under the isocratic elution conditions used with buffer A (Fig. 1).

However, under the above conditions, no separation is effected between the different nucleosides (Table II). In this instance, the identity of the common nucleosides was derived from their UV spectra.

A nucleoside formed after complete digestion of an oligonucleotide with PDE was separated by HPLC from its accompanying nucleotides, the peak corresponding with the nucleoside was collected, and finally its UV spectrum was recorded.

The identity of the unknown nucleoside can easily be deduced by comparing its UV spectrum with those of the common nucleosides recorded under exactly the same conditions.

Quantitation

The UV spectra of commercially available nucleotides and nucleosides were recorded in the same buffer A as applied for the isocratic elution of these compounds by HPLC. From these data, the ε values for the different nucleotides and

TABLE II

RETENTION TIMES* AND ε_{254} (pH 4.5) VALUES OF THE COMMON NUCLEOTIDES AND NUCLEOSIDES

Compounds	*Retention time (min)*	ε_{254} *(l/mmole·cm)*
rA; dA	0.0	13.0
rC; dC	0.0	6.6
rG; dG	0.0	13.5
rU	0.0	8.9
dT	0.0	7.4
prA; pdA	2.6	14.0
prC; pdC	1.2	5.3
prG; pdG	3.7	13.6
prU	1.7	9.1
pdT	1.7	6.8
rGp	7.3	13.6

* Relative to injection peak.

nucleosides can easily be calculated. The ε_{254} (pH 4.5) values are listed in Table II.

Peak areas, as recorded by the integrator, are normalized by dividing the corresponding areas by the ε_{254} (pH 4.5) values of the nucleotides or nucleosides measured.

The fidelity of the ε values was proved by the correct calculation of the composition, after separation by HPLC, of test mixtures containing known and variable amounts of nucleotides and nucleosides.

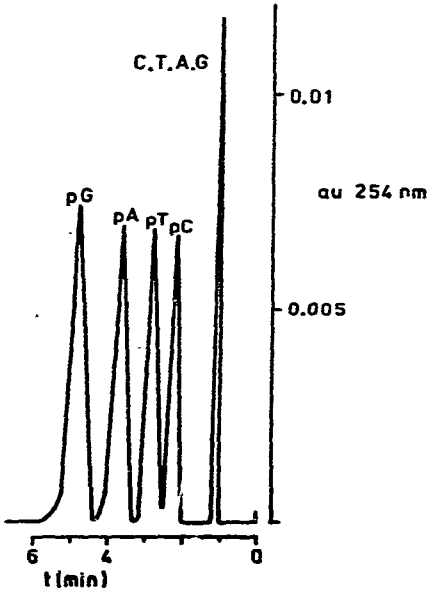


Fig. 1. Separation of 4 deoxyribonucleoside 5'-phosphates on a 1 m × 2.1 mm Permaphase AAX column using 0.005 *M* KH_2PO_4, pH 4.50, at 1 ml/min.

TABLE III

CONDITIONS APPLIED IN THE SEQUENTIAL ANALYSIS OF OLIGONUCLEOTIDES

The digestions for the sequence analysis were carried out at 20°.

Oligomer	*Absorbance units*	*Buffer (ml)*	*Internal standard, rGp (mg)*	*Enzyme*** *(μl)*
r-A–(C_4U)	19	0.2*	0.3	4
r-A–(AG_2)	9	0.3*	0.2	4
r-C–(U_3G)	1.1	0.1**	0.05	1
d-A–(AT_2C_2)	84	0.4*	0.7	6
d-G–(GA_2T_2)	81	0.3*	0.7	6

* Buffer: 0.1 *M* Tris-HCl, 0.01 *M* $MgCl_2$, pH 8.9.
** Buffer: 0.05 *M* Tris-HCl, 0.01 *M* $MgCl_2$, pH 8.9.
*** Enzyme: commercial PDE (1 mg/ml).

Further, the detector–integrator response is linear over a wide range on the detector scale, and when plotted as normalized areas *versus* nanomoles of nucleotides or nucleosides straight lines are obtained.

Sequence analysis

The sequential analysis of the synthetic oligonucleotides involves a two-stage process: (1) determination of base composition; (2) quantitation and identification of monomers as evolved in time by the action of PDE.

For this purpose, five stock solutions containing known amounts of each oligonucleotide, Tris-HCl buffer and internal standard, but no enzyme, were prepared according to the data in Table III.

The base composition of a synthetic oligonucleotide was then determined as follows. To a sample (10 μl) withdrawn from the corresponding stock solution was added PDE (2 μl of standard solution). After incubation for 2 h at 37°, a sample (5 μl) was analyzed, under isocratic conditions, by HPLC and the peak areas were divided by their corresponding ε_{254} (pH 4.5) values to give the molar fractions of the released monomers from which the base composition of the oligonucleotide could easily be calculated (Table IV).

The results in Table IV show that the calculated base composition is in excellent agreement with the expected composition of the synthetic oligomers. This finding is also indicative of the purity of these synthetic compounds.

The high degree of purity of the synthetic oligonucleotides was also independently ascertained by HPLC using gradient elution; the results of this study will be reported elsewhere.

It is also clear that, by the mode of action of PDE, the identity of the base at the left-hand terminus of the oligomer is also determined.

The stepwise degradation of oligonucleotides by PDE and the simultaneous recording of the amounts of nucleotides and nucleosides evolved was performed as follows.

Samples (5 μl) were taken, at appropriate intervals, from the corresponding stock solutions to which were added the appropriate amounts of enzyme (Table III) and, after inactivation (5 μl of 0.2 *M* KH_2PO_4, pH 3.5), analyzed by HPLC under

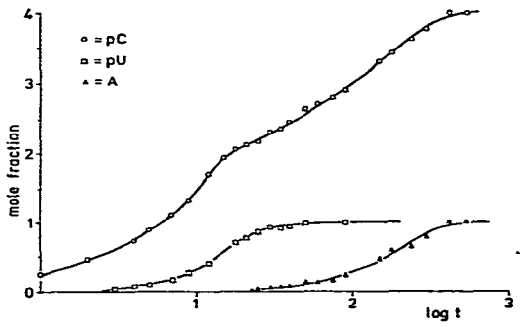
TABLE IV

BASE COMPOSITION OF SYNTHETIC OLIGONUCLEOTIDES AS DETERMINED BY HPLC ANALYSIS OF THEIR PDE DIGESTION PRODUCTS

Expected composition of synthetic oligonucleotide	*Molar fractions found*	*Calculated composition of synthetic oligonucleotide*
r-(AC_4U)	rA, 0.98 prC, 4.00 prU, 1.00	r-A-(C_4U)
r-(A_2G_2)	rA, 1.00 prG, 1.95 prA, 0.95	r-A-(AG_2)
r-(CU_3G)	rC, 1.00 prU, 2.98 prG, 1.02	r-C-(U_3G)
d-($A_2T_2C_2$)	dA, 1.00 pdT, 2.01 pdA, 1.00 pdC, 2.00	d-A-(AT_2C_2)
d-($G_2A_2T_2$)	dG, 1.00 pdG, 1.02 pdA, 1.98 pdT, 2.02	d-G-(GA_2T_2)

isocratic elution conditions. In this way, 20 samples could be analyzed continuously without intermittent gradient elution of the column. Peak areas of the corresponding monomers were divided by the peak areas of the internal standard to give normalized areas.

Plotting of molar fractions of released monomers, obtained by dividing normalized areas of nucleotides/nucleoside during the course of the digestion by the corresponding normalized areas at complete digestion, against time results in curves that relate the concentration of each nucleotide/nucleoside evolved to the time of PDE action. The results thus obtained for the five oligonucleotides are depicted in Figs. 2–6.



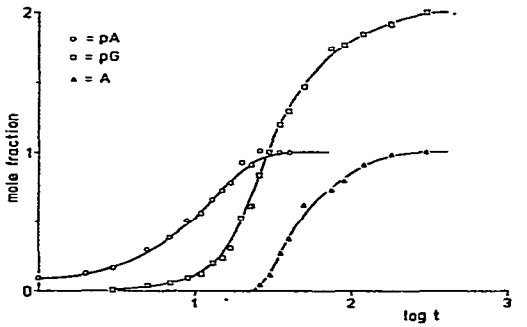
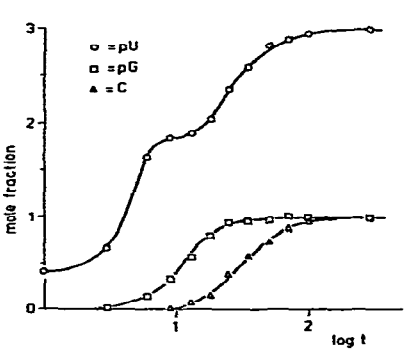


Fig. 2. Curves of the time of appearance of nucleotides (pC and pU) and nucleoside (A) from PDE action. Synthetic ribo-6-mer = r-ApCpCpUpCpC.

Fig. 3. Curves of the time of appearance of nucleotides (pA and pG) and nucleoside (A) from PDE action. Synthetic ribo-4-mer = r-ApGpGpA.



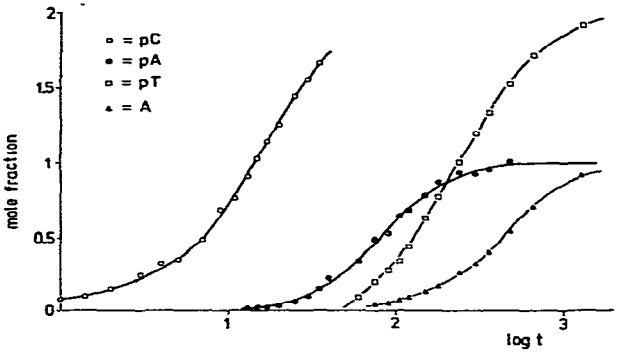


Fig. 4. Curves of the time of appearance of nucleotides (pU and pG) and nucleoside (C) from PDE action. Synthetic ribo-5-mer = r-CpUpGpUpU.

Fig. 5. Curves of the time of appearance of nucleotides (pC, pA and pT) and nucleoside (A) from PDE action. Synthetic deoxyribo-6-mer = d-ApTpTpApCpC.

Close inspection of these curves shows that the sequence of all of the oligonucleotides can easily be inferred. For instance, the sequence r-ACCUCC is easily derived from the curves in Fig. 2.

An initial inspection shows that the first monomer o appear ($t = 6$ min) up to a molar fraction of 1 is pC, indicating that it occupies the right-hand terminus. At $t = 18$ min, pC reaches a molar fraction of 2 and pU a molar fraction of 0.7, indicating that the sequence must be r-(AC_2U)–C–C.

The monomer pU reaches its maximum molar fraction of 1 at $t = 50$ min, while at the same time the molar fractions of pC and A are 2.54 and 0.1, respectively. The sequence can therefore be written as r-(AC_2)–U–C–C.

Further inspection of the curves shows that the fourth base is C, because its molar fraction at $t = 100$ min is 3 before any appreciable amount of the base A appears, thus affording the sequence r-(AC)–C–U–C–C.

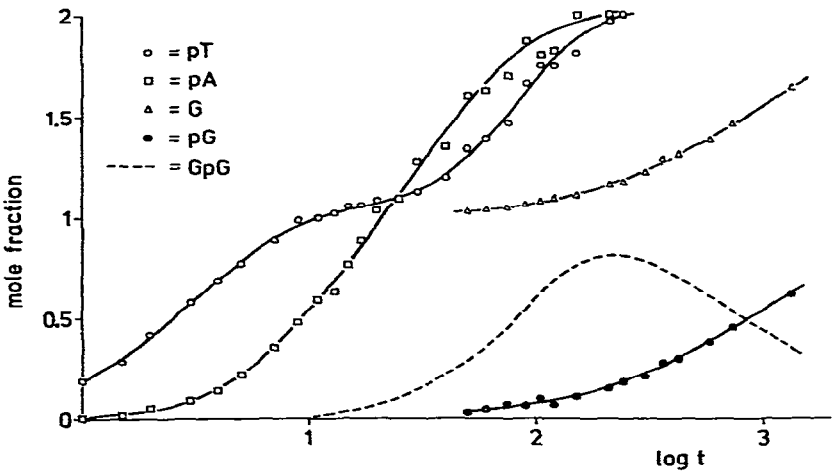


Fig. 6. Curves of the time of appearance of nucleotides (pT, pA and pG) and nucleoside (G) from PDE action. Synthetic deoxyribo-6-mer = d-GpGpTpApApT.

This leaves one A and one C unaccounted for. The only possible sequence for the last two bases at the left-hand terminus must be r-A–C, because the result of the base composition of this oligomer (Table IV) puts A in position six, leaving C only to be positioned as the fifth base. The inferred sequence is therefore r-A–C–C–U–C–C.

By the same reasoning, the correct sequence of all the other oligonucleotides can easily be deduced from the curves in Figs. 3–6.

The curves in Figs. 2–6 show that the rate of digestion of dinucleoside monophosphates is very slow. This effect is clearly illustrated in Fig. 6: at $t = 200$ min, the dimer evolved, dGpG, has a molar fraction of 0.80, while the released monomers, pdG and dG, both have a molar fraction of 0.12.

It should be mentioned that the sluggish action of PDE on several deoxyguanosine oligonucleotides, *e.g.*, dGpG and pdGpG, has also been observed by Ralph *et al.*[19].

The differential rate of PDE on the dimer level would make the above sequence analysis method very tedious. However, this drawback is overcome in our technique. The base composition, as obtained from the same stock solution as is used for the oligomers to be sequenced, not only permits a direct calculation of molar fractions of monomers evolved in time, but also reveals the identity of the nucleoside at the left-hand terminus. These two factors reduce the time necessary for a complete sequence considerably. In fact, the sequence analysis procedure can be stopped when the base at a position $n - 2$ (where n = number of monomers) from the right-hand terminus of the chain reaches a molar fraction of 0.5; analysis beyond this point is irrelevant to the actual sequence analysis.

The latter implies that the sequence of the oligomer r-ACCUCC is determined at $t = 50$ min instead at $t = 100$ min. The same principle is applicable to the other examples in Figs. 3–6.

The optimal conditions (minimum amount of substrate *versus* enzyme) for determining the sequence of oligonucleotides are still under investigation. However, it is worthwhile mentioning that the sequence of the pentamer r-CUGUU could be determined successfully by using 0.2 absorbance unit of substrate (4.63 nmole) and 0.5 μl of enzyme. Under these conditions, the actual time of sequencing was reduced by a factor three.

CONCLUSION

The method described demonstrates that the combined use of HPLC and PDE action promises to be a practical and simple technique for the sequential analysis of small oligonucleotides that have free 3′- and 5′-hydroxy functions. Studies are currently in progress to adapt this method for sequencing oligonucleotides that have phosphate groups at one or both ends.

Further, it may be mentioned that HPLC enabled us to evaluate, for the first time, the theoretical study of Cantor *et al.*[20] of exonuclease kinetics and its application to the determination of base sequences in nucleic acids.

ACKNOWLEDGEMENTS

This research was supported by the Netherlands Foundation for Chemical

Research (S.O.N.) with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

We thank Mrs. G. Wille-Hazeleger, Drs. P. M. J. Burgers and Drs. P. H. van Deursen for their generous efforts in synthesizing and purifying the oligonucleotides.

REFERENCES

1 J. H. van Boom, P. M. J. Burgers, G. R. Owen, C. B. Reese and K. Saffhill, *Chem. Commun.*, (1971) 869.
2 N. J. Cusack, C. B. Reese and J. H. van Boom, *Tetrahedron Lett.*, (1973) 2209.
3 J. H. van Boom, P. M. J. Burgers and P. H. van Deursen, *Tetrahedron Lett.*, (1976) 869.
4 R. W. Holley, J. T. Madison and A. Zamir, *Biochem. Biophys. Res. Commun.*, 17 (1964) 389.
5 T. F. Gabriel and J. Michalewski, *J. Chromatogr.*, 67 (1972) 309.
6 C. A. Burtis, *J. Chromatogr.*, 51 (1970) 183.
7 W. P. Kennedy and J. C. Lee, *J. Chromatogr.*, 51 (1970) 203.
8 R. A. Hartwick and P. R. Brown, *J. Chromatogr.*, 112 (1975) 651.
9 J. J. Kirkland, *J. Chromatogr. Sci.*, 8 (1970) 72.
10 R. A. Henry, J. A. Schmidt and R. C. Williams, *J. Chromatogr. Sci.*, 11 (1973) 358.
11 C. Horvath, B. A. Preiss and S. R. Lipsky, *Anal. Chem.*, 39 (1967) 1422.
12 C. Horvath and S. R. Lipsky, *Anal. Chem.*, 41 (1969) 1227.
13 W. E. Razzell and H. G. Khorana, *J. Biol. Chem.*, 234 (1959) 2105.
14 W. E. Razzell and H. G. Khorana, *J. Biol. Chem.*, 234 (1959) 2114.
15 E. J. Williams, S. C. Sung and Sr. M. Laskowski, *J. Biol. Chem.*, 236 (1961) 1130.
16 T. Nihei and G. L. Cantoni, *J. Biol. Chem.*, 238 (1963) 3991.
17 H. G. Boman and U. Kaletta, *Biochim. Biophys. Acta*, 24 (1957) 619.
18 H. G. Boman, *Ann. N.Y. Acad. Sci.*, 81 (1959) 800.
19 R. K. Ralph, W. J. Connors, H. Schaller and H. G. Khorana, *J. Amer. Chem. Soc.*, 85 (1963) 1983.
20 C. R. Cantor, I. Tinoco and L. Peller, *Biopolymers*, 2 (1964) 51.