A NEW APPROACH TO THE SEQUENCE ANALYSIS OF DNA

Wolfgang OERTEL* and Heinz SCHALLER
Max-Planck-Institut für Virusforschung, BRD 7400 Tübingen, Spemannstrasse 35, German Federal Republic

Received 1 September 1972

1. Introduction

For the analysis of structure and function of DNA it would be of advantage, if short defined segments of the macromolecule could be copied *in vitro.* DNA polymerase I is able to initiate DNA synthesis at internal sites on a single-stranded DNA template, if oligonucleotides are offered as primers [1, 2]. Recent experiments indicated that the initiation reaction may be sequence specific, since a high template specificity was found when oligonucleotides were used as primers which were long enough to anneal stably fo a template DNA [3, 4].

In the present communication we show that DNA polymerase I can initiate DNA synthesis at a single site on a phage DNA template, if a unique base complementary oligonucleotide is offered as a primer. Fd DNA minus strands were used as template, the polypyrimidine tract C_9T_{11} from the fd plus strand [5] as primer. Pulse-labelled reaction products of various size were characterized by fingerprinting techniques, which allowed to deduce the sequential arrangement of the polypyrimidine tracts in the nucleotide sequence following the 3′ terminus of the primer nucleotide.

2. Materials and methods

Density labelled fd minus DNA was obtained from double stranded covalently closed fd RF, which had been synthesized *in vitro* [6, 7] by Dr. H. Köster from fd plus DNA [8] in the presence of bromodeoxyuridine triphosphate. The fd RF was fragmented by incubation at 75° at pH 7.5 for 3 hr. This treatment produced mainly 3′ phosphate termini which are not primers for DNA polymerase. The DNA strands were separated in an alkaline density gradient (ρ 1.54) containing the DNA solution (28 ml), 1 M NaOH (2 ml), 60% Cs_2SO_4 (16.4 ml), and CsCl (16 g). The isolated fd minus DNA had an average chainlength of about 1500 nucleotides, as judged from sedimentation analysis. Its molar equivalents are expressed using the chainlength of its complete nucleotide sequence (6000 nucleotides) to allow a direct correlation of the concentrations of template and primer DNA sequences.

A mixture of pentadecanucleotides from fd DNA was prepared by ion exchange chromatography in 7 M urea from a pancreatic DNAase digest of fd DNA [3]. C_9T_{11}-oligonucleotide was prepared from ^{32}P-labelled fd DNA (55 μmole P, 7×10^6 cpm) by the depurinization method of Burton [9], and purified by repeated gel filtration on Sephadex G-50 in 0.1 M $(NH_4)_2CO_3$, pH 8.6 [10]. Terminal phosphate was removed by treatment with *E. coli* alkaline phosphatase. The nucleotide obtained (15 000 cpm ^{32}P, 0.11 μmoles DNA (P), 5.5 nmoles DNA chains) migrated homogeneously with an apparent chainlength of 20 nucleotides during electrophoresis on standardized polyacrylamide gels [11]. Annealing of oligonucleotides and fd minus DNA was carried out in 0.1 M Tris-HCl pH 7.5, 0.5 M NaCl. Samples were heated in sealed capillary tubes to 95° (1 min), and incubated at 20° (15 hr). For DNA synthesis these solutions were diluted 10-fold, in analytical experiments 5-fold. The reaction mixtures contained 0.05 M Tris-HCl pH 7.5, 0.1 M NaCl, 0.01 M $MgCl_2$, 100 μM dATP, 100 μM dGTP,

* Present address: University of California, San Diego, Department of Medicine, School of Medicine, La Jolla, Calif. 92037, USA.

North-Holland Publishing Company – Amsterdam

100 μM dTTP, 10 μM [^{3}H]dCTP (9.5 Ci/mmole). For preparation of pulse-labelled DNA the unlabelled dTTP was replaced by 10 μM [^{32}P]dTTP (3 Ci/mmole). Samples of 310 μl containing 7.5 pmole of fd minus DNA/C_9T_{11} duplex (10-fold excess of C_9T_{11}) were incubated at 20° for 3 min. DNA synthesis was initiated by adding DNA polymerase (35 units, 35 pmole), and terminated by adding 1 M EDTA (10 μl). The DNA was precipitated with 2 vol of ethanol. The pellet was washed with 70% ethanol and dissolved in H_2O containing 7 M urea and 20% sucrose (150 μl). The samples of bromophenol blue (final volume about 150 μl). The samples were applied onto 12% polyacrylamide gels in 0.04 M acetate pH 5, 7 M urea. Electrophoresis was carried out as described [11]. Gels were fractionated in a Savant gel cutting device using 2 mM Tris base, 0.2 M LiCl as elution buffer. ^{32}P-labelled DNA was located by measuring Cherenkow radiation in a liquid scintillation counter. From fractions of interest the gel fragments were removed by centrifugation, and the DNA was desalted by ethanol precipitation or by gel filtration on Sephadex G-50 in H_2O. For chainlength determination gel electrophoresis was carried out in the presence of RNA fragments of defined chainlength (125, 73, 55 and 27 nucleotides, respectively [11]). Fingerprints of the depurination products [9] were obtained according to Southern and Mitchell [12]. The base compositions of the oligonucleotides were derived from their position on the plate [12] using fingerprints from ^{32}P-labelled fd DNA as a reference [11]. These base compositions were confirmed for all nucleotides (including the material in tails) by the ^{3}H-C/^{32}P-T ratios, and also by their migration in a different two-dimensional thin-layer system [5, 13].

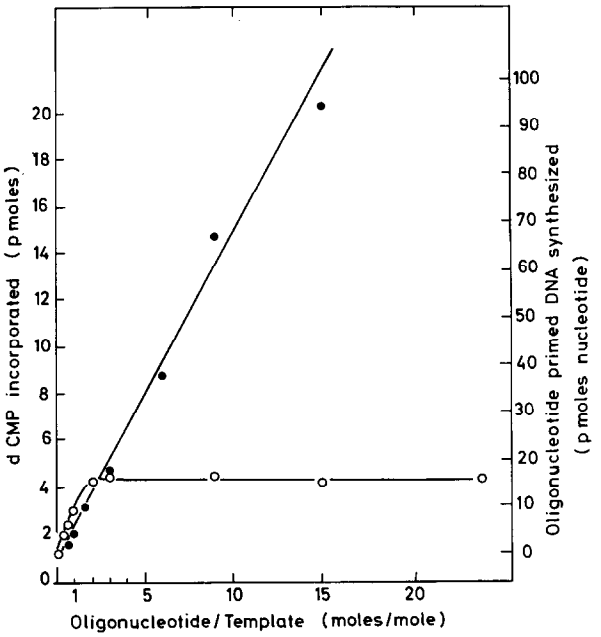

Fig. 1. Saturation of fd minus DNA with C_9T_{11} oligonucleotide. Fd minus DNA (0.33 pmoles) and various amount of oligonucleotide were incubated in annealing buffer (20 μl) as described under Methods. Nucleotide incorporation was carried out at 20° for 2 min in 100 μl of incorporation mixture containing 3.5 units of DNA polymerase I. DNA synthesis in presence of C_9T_{11} oligomer (o—o—o), in presence of random pentadecanucleotides from fd DNA (•—•—•).

3. Results and discussion

As found previously with other long oligonucleotides [3] the stimulation of DNA synthesis by the C_9T_{11}-oligomer depends on its stable annealing to the fd minus DNA template. Since the annealing reaction required several hours for completion [3], a 15 hr preincubation was used routinely to insure maximum duplex formation.

To estimate the number of possible initiation sites on the template, the initial rate of C_9T_{11}-primed DNA synthesis was measured at various primer/template ratios. Fig. 1 shows that the DNA template is saturated at a low amount of C_9T_{11} oligomer. Large excess of the primer nucleotide does not lead to a further stimulation of nucleotide incorporation. This is in contrast to a chain initiation primed by a mixture of random pentadeca-nucleotides from fd plus DNA which as shown in fig. 1 is proportional over a wide range to the amount of primer added. These results suggest that on the fd minus DNA there are very few sites, to which the C_9T_{11}-oligomer binds stably, and at which DNA synthesis is started.

More direct evidence that C_9T_{11}-primed DNA chains are initiated at a single site on the template comes from the analysis of pulse labelled reaction products. DNA was synthesized in the presence of [^{3}H]dCTP and [^{32}P]dTTP for short time intervals to

Table 1
Sequential synthesis of pyrimidine nulceotides in C_9T_{11}-primed DNA chains.

Sample	Time of DNA synthesis	Chain-length	Relative yields of polypyrimidine nucleotides								
	(sec)	(n)	(moles)								
			C_X	C_3T_3	C_3T_5	CT	C_2T_2	C_3T_4	T_3	T_2	T_1
1	5	28	3.0	0.17							
2	20	31	2.8	1.0	0.07						
3	10	34	2.3	1.0	0.11						
4	20	38	3.6	1.0	0.82	0.2					
5	20	47	2.5	1.0	0.97	1.08					
6	30	53	2.8	1.0	1.16	1.33	1.0				
7	40	83	2.5	1.0	1.06	2.2	1.43	0.8	0.15	<0.01	<0.01

DNA samples from preparative gel electrophoresis (see fig. 2) were analysed for chainlength [11] and polypyrimidine sequences (see fig. 3). The radioactivity of each pyrimidine nucleotide was measured, and the molar yieds normalized relative to that of the nulceotide C_3T_3. Data from sample 1 were normalized relative to 3 moles of ^{3}H-labelled cytosine in position "C_X". Separate experiments showed that "C_X" consisted of an equimolar mixture of C_1 and C_2 nucleotides. The limit of detection by autoradiography of ^{32}P-labelled spots was below 0.01 moles of T, since in sample 7 the nucleotides T_2 and T_1 were visible as weak spots on the X-ray film.

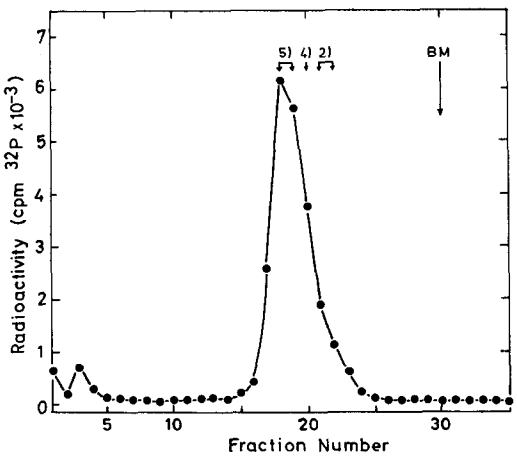

Fig. 2. Separation according to size of pulse labelled DNA. ^{3}H-C, ^{32}P-T labelled DNA was synthesized for 20 sec as described under Methods, and the reaction products subjected to electrophoresis on 12% polyacrylamide gels [11]. Migration was from the left to the right. BM indicates the position of the blue marker. The fractions indicated by the arrows were analysed further for chainlength and polypyrimidine distribution (see table 1).

yield DNA chains ranging from 25–100 nucleotide residues (table 1). They were separated from the template DNA and fractionated according to size by preparative gel electrophoresis under denaturing conditions in 7 M urea (fig. 2). The reaction products showed a rather broad size distribution indicating that the DNA chains did not grow synchronously. DNA pieces of almost homogeneous size were, however, obtained by eluting different regions of the gels. Their chainlength was determined on standardized urea gels [11]. The complexity of the newly synthesized nucleotide sequences was analysed by fingerprinting the depurination products of the labelled DNA. The fingerprints as visualized by autoradiography and quantitatively by liquid scintillation counting (fig. 3, table 1) show that the C_9T_{11}-primed DNA chains contain very few polypyrimidine sequences. The labelled polypyrimidine tracts appear sequentially with increasing chainlength; they approach single molar ratios; in longer chains the nucleotides CT and C_2T_2 approach two molar ratios. In addition to the ^{32}P-T labelled nucleotides, nucleotides containing [^{3}H]cytidine only (about 3 equivalents) were detected on the chromatograms in the position of C_1 and C_2 nucleotides throughout all reaction products. From these data the sequence of the polypyrimidines

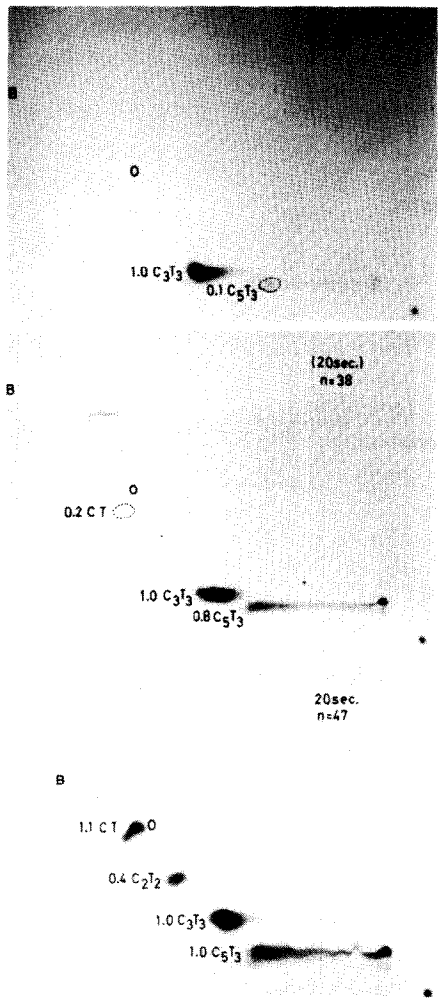

Fig. 3. Radioautographs of two-dimensional fractionation on PEI cellulose [12] of the depurination products of pulse labelled DNA. The first dimension was 2 M pyridine formate pH 3.5 (right to left), the second dimension 1 M LiCl (bottom to top). Fingerprints from DNA samples (2, 4, 5) (table 1) are presented. Base composition and molar ratios of each pyrimidine nucleotide is indicated. B and O denote the positions of dye markers [13].

next to the C_9T_{11} cluster can be assigned tentatively as follows:

$$5'\text{-}C_9T_{11}\text{-}(C_1\text{-}C_2)\text{-}C_3T_3\text{-}C_5T_3\text{-}CT\text{-}C_2T_2\text{-}CT\text{-}$$

$$C_3T_4\text{-}C_2T_2\text{-}T_3\text{-}3'.$$

In all samples analysed these polypyrimidine sequences account for about 50–70% of the material in the newly synthesized portion of a single growing DNA chain leaving 30–50% for interdispersed purine nucleotides. These results strongly support the notion that the reaction products consist of a homogeneous single species of DNA growing from a unique point on the DNA template. The precise nucleotide sequence of this region can now be obtained by using different combinations of labelled and un-labelled nucleoside triphosphates during *in vitro* DNA synthesis [14–16] and by the analysis of the nucleotide sequences within each polypyrimidine cluster [5].

Recently we also observed sequence specific initiation of DNA synthesis on fd plus DNA template by using as primer nucleotides unique fragments from fd minus DNA as well as from fd minus RNA. Thus the method appears to be generally applicable, and provides a means to subdivide a natural DNA into short segments, which can be used in structural and functional studies.

Acknowledgements

We thank Dr. W. Fiers, Gent, for providing RNA markers of defined chainlength, and Dr. K.W. Mundry, Tübingen, for introducing us to fingerprinting techniques. This work was supported by the Deutsche Forschungsgemeinschaft.

References

[1] M, Goulian, Cold Spring Harbor Symp. Quant. Biol. 33 (1968) 11.
[2] M. Goulian, Ann. Rev. Biochem. 40 (1971) 855.
[3] W. Oertel, Thesis Tübingen (1971).
[4] W. Oertel and H. Schaller, submitted for publication (1972).
[5] V. Ling, J. Mol. Biol. 64 (1972) 87.
[6] M. Goulian and A. Kornberg, Proc. Natl. Acad. Sci. U.S. 58 (1967).
[7] B. Heyden, Biochem. Diplomarbeit Tübingen (1970).
[8] H. Schaller, J. Mol. Biol. 44 (1969) 435.
[9] K. Burton, in: Methods in Enzymology, 12 part A, eds. L. Grossmann and K. Moldave (Academic Press, New York, London 1967) p. 222.
[10] Th. Hohn and H. Schaller, Biochim. Biophys. Acta 138 (1967) 466.
[11] B. Heyden, Ch. Nüsslein and H. Schaller, Nature New. Biol. (1972) in press.

[12] E.M. Southern and A.B. Mitchell, Biochem. J. 123 (1971) 623.
[13] G.G. Brownlee and F. Sanger, J. Mol. Biol. 23 (1967) 337.
[14] M.A. Billeter, J.E. Dahlberg, H.M. Goodman, J. Hindley and Ch. Weissmann, Nature 224 (1969) 1083.
[15] W. Salser, K. Fry, C. Brunk and R. Poon, Proc. Natl. Acad. Sci. U.S. 69 (1972) 238.
[16] R. Wu and E. Taylor, J. Mol. Biol. 57 (1971) 491.