Purification and Base Composition Analysis of Phage Lambda Early Promoters

(E. coli/RNA polymerase/sigma factor)

JEAN-YVES LE TALAER AND PH. JEANTEUR

Unité de Biochimie-Enzymologie, Institut Gustave-Roussy, 94, Villejuif, France; and Laboratoire de Biochimie, Centre Regional de Lutte contre le Cancer, Cliniques Saint-Eloi, 34, Montpellier, France

Communicated by André Lwoff, October 4, 1971

ABSTRACT RNA-polymerase of Escherichia coli was allowed to bind to DNA of phage lambda in the absence of precursors. The resulting complex was excised by nuclease digestion and the protected DNA was recovered by phenolextraction and ethanol precipitation. Acrylamide gel electrophoresis of protected DNA fragments reveals the existence of two distinct oligonucleotide peaks corresponding. respectively, to 45-52 and 7-10 nucleotide residues along with species of intermediate sizes. Peak I molecules have two properties: (a) their existence is dependent on the presence of sigma factor during the initial binding step, and (b) they are considerably enriched in A-T (up to 67%). On the contrary, peak II molecules have the same base composition as DNA of phage lambda, whether obtained in the presence or absence of sigma factor. Peak I molecules are thus believed to contain DNA sequences involved in promoter recognition, whether they are the promoters themselves, adjacent, or related sequences.

The very high degree of specificity involved in promoter recognition by RNA-polymerase (EC 2.7.7.6) in the presence of sigma-type transcription factors calls for the existence of unique features of the promoters DNA making them recognizable by the polymerase through the mediation of their cognate sigma factors.

Taking advantage of the DNase resistance of DNA binding sites for RNA-polymerase (1-5), several groups have looked for peculiarties in base composition of these sites, as compared to that of total DNA. Though some enrichment in G-C was observed in Escherichia coli DNA moieties that are bound to the 15S form of RNA-polymerase (4), no differences were detected in E. coli with the 22S polymerase (4) nor with T_7 (2), or λ (5). In this earlier work with λ DNA (5), no attempt was made to limit nonspecific binding sites and the analysis was performed on protected DNA fragments amounting to 9-18 sites per genome. At variance with these results, we reported (6) conditions under which a significant enrichment in A-T of the protected DNA fragments could be observed when limited amounts of polymerase were bound, corresponding to 1 or 2 binding sites per lambda genome. At higher levels of protection, no enrichment in A-T could be detected. This discrepancy is easily explained if one considers that the number of 1-2binding sites we analysed is well within the range of 2-4 initiation sites that are liable to be active under these conditions. Indeed, besides the two genetically defined early

promoters under direct control of the λ repressor (7), there may be one or two additional promoters, active in repressed lysogens (8) and probably involved in the synthesis of the repressor itself. By measuring the number of rifampicinresistant initiation sites, Bautz and Bautz (9) found three promoters on λ , recognizable by the *E. coli* sigma factor. From kinetic data, Naono and Tokuyama (10) inferred four initiation sites. This latter figure, thus, seems to be a maximum.

Though the experimental conditions required for the formation of a stable promoter-polymerase complex, namely, presence of sigma and high temperature $(37^{\circ}C)$ binding (9, 11), were already met in our earlier work, the results merely suggested that the population of binding sites that were analysed might contain the specific initiation sites for DNA transcription. To establish conclusively that the A-T enrichment observed was related to the presence of such sites, it was necessary to show that the A-T enrichment is sigma-dependent. We now report that such is the case.

Polyacrylamide gel electrophoresis of the protected DNA (6) had shown, in addition to a sharp peak corresponding to pieces about 50 nucleotides long, a significant amount of smaller size material likely to result from secondary binding of excess free polymerase molecules to the very initial hydrolysis products, during the DNase digestion step. Further purification of the DNA fragments on acrylamide gels was thus performed and led to further enrichment in A-T.

MATERIALS AND METHODS

(a) Enzymes and DNA. RNA-polymerase was extracted from E. coli A 19 (RNase)⁻ according to Babinet (12) and further purified by two cycles of glycerol gradient centrifugation (13). Stock solutions of the enzyme were kept at -20° C in Burgess' storage buffer (pH 7.9) (13). Acrylamide gel electrophoresis in 0.1% SDS (14) revealed the presence, in addition to the regular α , β , β' , and σ subunits, of τ band and traces of μ band (15). This preparation will be referred to as complete enzyme. Pure core enzyme was obtained by further treatment of complete enzyme, through two successive phosphocellulose columns, according to Burgess (14), in order to ensure exhaustive removal of σ band. This preparation, referred to as PC enzyme, contained only α , β , and β' bands with no detectable trace of σ band in overloaded gels. Storage conditions were as for the complete enzyme. Pancreatic DNase was obtained from Worthington. Venom phosphodiesterase (EC 3.1.4.1) with no 5'-nucleotidase activity and Pronase were purchased from Calbiochem.

Abbrevations: SDS, sodium dodecyl sulfate; EDTA, ethylene diamine tetra acetate (Na salt); PC enzyme, pure core enzyme (without σ factor).



FIG. 1. Radioactivity profiles on acrylamide gels of DNA protected fragments of phase lambda DNA. Fragments obtained in the presence (a, b, and c) or absence (d) of sigma factor, were prepared as described in *Methods* b. The binding mixtures contained per ml: (a) 81.6 μ g DNA and 18.6 μ g complete enzyme (+ σ); (b) 70 μ g DNA and 84 μ g complete enzyme (+ σ); (c) 35 μ g DNA and 436 μ g complete enzyme (+ σ); (d) 35 μ g DNA and 32.4 μ g core enzyme (- σ). *Arrows* indicate the marker dye position.

The labeled lambda phages were obtained by thermal induction of *E. coli* C 600 (λ C₁ 857)/ λ grown in modified Kaiser's medium (16), containing 5% casamino acids and 1 mM K₂HPO₄. After growth at 37°C to 0.7–0.8 absorbance at 650 nm, the culture was quickly centrifuged at 4000 rpm in a GSA Servall rotor and resuspended in half volume of the same fresh medium containing 20 μ Ci/ml H₃³²PO₄ (CEA, Saclay, France) and previously equilibrated to 41°C. After 15 min at 41°C, the culture was cooled to 38°C and further incubated for 1 hr with vigorous stirring. A few drops of chloroform were added with continuous shaking; the phages were purified from the lysate by phase partition and banding in cesium chloride (17). After dialysis, DNA was extracted with phenol, with gentle shaking as described (6) and exhaustively dialyzed against 0.04 M Tris HCl (pH 8.0)–1 mM EDTA.

(b) Binding of Polymerase and Isolation of Protected DNA Pieces. Binding mixtures, as described (6), were in 0.04 M Tris. HCl (pH 8.0)-0.01 M MgCl₂-1 mM CaCl₂ and contained various amounts of [³²P]DNA and polymerase. After 15 min of incubation at 37°C, pancreatic DNase (EC 3.1.4.5.) (Worthington DP grade) and venom phosphodiesterase (Calbiochem) were added to 200 μ g/ml and 7.5 U/ml, respectively, and the incubation was continued for an additional 0.5 hr. Pronase digestion, SDS-phenol extraction, and ethanol precipitation were as described (6) except that EDTA concentration was raised to 20 mM and that only one such precipitation was performed. The pellet was finally dissolved in 0.2 ml of 1 mM Tris·HCl (pH 8.0)-1 mM EDTA.

(c) Gel Electrophoresis of Protected DNA Fragments. 10% polyacrylamide gels containing 5% glycerol and no SDS were prepared according to the method of Peacock and Dingman (18) in 0.8 \times 15 cm plastic tubes and electrophoresed for 30 min at 10 mA/tube. A few sucrose crystals and $4 \,\mu l$ of 0.2%bromophenol blue were added to the 0.2-ml samples obtained in (b). After dissolution, the samples were layered on top of gels and electrophoresed at 8 mA/tube until the dye had migrated 10 cm (about 3.5 hr). Gels were cut into 4-mm slices and counted in a Tri-Carb scintillator by the use of Cerenkov radiation. DNA fragments were recovered from the gels by the following procedure. Slices corresponding to peak fractions were cut in quarters, placed on a plastic tube, and immersed in a polyacrylamide mixture identical to that used for the first run. After polymerization had occurred, the tubes were fitted at their lower end by a dialysis bag that was closed so as to contain about 1 ml of the electrophoresis buffer without air bubbles. Electrophoresis was continued for 1.5 times the duration necessary for the marker dye to reach inside the dialysis bags. The bags were then removed from the tubes,



FIG. 2. Calibration curve of electrophoretic mobility versus molecular weight for several RNA markers. This curve was used for estimation of the size of molecules in peaks I and II. Molecular weights, expressed in number of nucleotide residues, are plotted on semilog coordinates versus relative mobilities. In addition to 4S and 5S RNA from KB cells, the following oligonucleotides were used: no. 51 (5 nucleotides), no. 52 and 53/56 (6 nucleotides), no. 54 (7 nucleotides), and no. 55 (13 nucleotides). For nomenclature and sequences, see reference (19). Between *arrows* are shown the ranges of relative mobilities observed for peaks I and II from Fig. 1.

Expt. No.	Binding conditions	Protec- tion	Unfractionated					Peak I					Peak II				
			С	A	Т	G	A-T	С	A	Т	G	A-T	С	A	Т	G	A-T
1	Total DNA		24.8 (12)	26.0	24.5	24.7	50.5							_			
2	$+\sigma$, R = 0.23	0.09	21.1 (4)	28.5	28.4	22.0	56.9	16.8 (2)	32.0	34.6	16.6	66.8			—		
3	$+\sigma, R = 1.2$	0.79	21.2(1)	28.5	24.6	25.7	53.1	18.0(2)	32.3	33.1	16.6	65.4					_
4	$+\sigma, R = 9.6$	4.3						19.2(2)	30.4	29.4	21.0	59.8					
5	$+\sigma$, R = 20		24.7(1)	26.6	25.4	23.4	52.0	_ `					25.0(1)	24.9	24.3	25.8	49.2
6	$-\sigma, R = 0.93$	0.75											23.6(1)	25.3	24.0	27.1	49.3
7	$-\sigma$, R = 15.4		-		—								26.5 (1)	25.4	24.2	23.9	49.6

TABLE 1. Nucleotide composition of protected DNA fragments

Binding conditions were as described in *Methods* b with the following amounts of [³²P]DNA of phage lambda and polymerase per 1 ml incubation mixture: (*Expt. 2*) 81.6 μ g DNA; 18.6 μ g complete enzyme. (*Expt. 3*) 70 μ g DNA; 84 μ g complete enzyme. (*Expt. 4*) 35 μ g DNA; 436 μ g complete enzyme. (*Expt. 5*) 28 μ g DNA; 560 μ g complete enzyme. (*Expt. 6*) 35 μ g DNA; 32.4 μ g core enzyme. (*Expt. 7*) 28 μ g DNA; 432 μ g core enzyme. *R* refers to the weight ratio between polymerase and DNA. Unfractionated material was prepared according to *Methods* b except that the ethanol precipitation was performed twice as described (6). Peaks I and II were obtained by gel electrophoresis and correspond, respectively, to Fig. 1a, b, c, and d for Expts. 2, 3, 4, and 6. The % protection was measured as described (6). The figures in parentheses indicate the number of determinations.

closed, and dialyzed against $10^{-2}~\rm Tris\cdot HCl~(pH~8.0)-10^{-2}~M~MgCl_{2}-1~mM~CaCl_{2}.$

(d) Nucleoide-Composition Analyses. Analyses were performed on Dowex 1×8 columns as described (6).

(e) Size Determination of Protected DNA. In an attempt to estimate the size of protected DNA fragments, gels were calibrated by the use of the following ³²P-labeled RNA markers: 4S and 5S RNA from KB cells and oligoribonucleotides of known length and sequence obtained by T_1 RNase digestion of the oligoribonucleotides (19). All these RNA species were kindly supplied by Dr. C. J. Larsen. Gels were subjected to electrophoresis as described above and the relative mobilities with respect to that of the marker dye were determined for the different oligonucleotides. The log of sizes, expressed in number of nucleotide residues, were then plotted against relative mobilities (20). Protected DNA, obtained as in (b) was run comparatively with or without previous heating for 5 min at 100°C, followed by quenching in ice.

RESULTS

(a) Analysis of protected DNA on gels

The migration profiles in gels of protected DNA obtained in the presence of sigma factor at three different protection levels are shown in Fig. 1a, b, and c and in the absence of sigma factor in Fig. 1d. When complete enzyme (containing sigma) is used, two distinct peaks are obtained. The slowest one (peak I) is quite sharp and corresponds to an estimated size of 45-52nucleotides (see below). The faster one (peak II) is less sharp and represents hepta to decanucleotides according to the same calibration. Molecules of intermediate sizes were observed in all three profiles. As no radioactivity is present in regions corresponding to oligonucleotides shorter than 7-10 nucleotides, this size is believed to represent the limit of alcohol insolubility. As a matter of fact, when several alcohol precipitations are performed, as was the case in our previous work (6), peak II tends to disappear, though intermediate regions and peak I are not affected.

In striking contrast with the biphasic profile obtained in the presence of sigma is the pattern observed with PC enzyme (lacking sigma factor): peak I and intermediate species disappeared, leaving only a sharp peak II. The molecular weight of peak I molecules (45–52 nucleotides) is in rather good agreement with the polymerase diameter (see *Results*, Section c) and is therefore compatible with the protection resulting from stable binding of the polymerase to native DNA (initial binding, i.e., before nuclease digestion). More precisely, if we assume that 45–52 nucleotide residues are actually covered by the polymerase, the persistence of molecules of such size (peak I) during digestion may be explained in two ways: (*i*) the initial binding complex is irreversible or (*ii*) this complex is reversible but the competition between polymerase and DNAse fragment is still in favor of the polymerase.

If we assume that the promoter-polymerase complex is highly stable when formed in the presence of sigma at 37° C (11), the first possibility would reflect the formation of such a complex, though the alternative would account for the existence of nonspecific binding sites. Of course, both mechanisms may play a role, the second one becomes more significant when the amount of polymerase is increased to such an extent as not to be negligible in comparison to that of the nuclease. According to this scheme, peak I would contain, in addition to the specific irreversible sites, increasing amounts of nonspecific material when the ratio of polymerase to DNA is increased.

On the contrary, molecules of the intermediate and peak II regions are too small to be accounted for by initial binding as defined above. They are likely to arise from secondary binding of polymerase to the products of incomplete hydrolysis by nucleases. This secondary binding would be stable enough to result in protection but would not involve any specific interaction and therefore would be independent of the presence of the sigma factor.

(b) Base composition of peaks from acrylamide gels

Table 1 shows the nucleotide composition of peak I from Fig. 1a, b, and c and peak II from Fig. 1d. The sigma-dependent peak I appears to be strongly enriched in A-T, the extent of this enrichment being more important at low polymerase/DNA ratio. This observation is probably relevant to the fact that peak I can contain nonspecific binding sites according to alternative (i) above, the relative proportion of which being increased with the polymerase concentration. At variance with peak I, the base composition of peak II is quite similar to that of total DNA. However, because of the too low counts that are

recovered from the gels, peak II from Fig. 1*a*, *b*, and *c* could not be analyzed. A similar experiment with complete enzyme but at even higher polymerase/DNA ratio (R = 20) was performed to allow this determination. Results were similar to that of peak II obtained with PC enzyme.

These data strongly confirm our previous base-composition results (6). The further enrichment in A-T (67% compared to 57%) that we observe here is due to the better purification now achieved, which gets rid of protected fragments smaller than about 45 nucleotides.

(c) Size estimation of protected DNA

Fig. 2 shows the calibration curve of 10% acrylamide gels with 4S, 5S RNA, and oligonucleotides as markers. Although not a straight line, as in the case of larger size RNAs (20), this curve can be used for the determination of at least approximately, the size of molecules whose relative mobilities fall between two markers. Molecules in peaks I and II can thus be shown to contain 45-52 and 7-10 nucleotide residues, respectively. However, the validity of such a calibration is limited by two restrictions: (i) it is assumed that the hydroxyl group, present in the RNA markers and absent in the unknown DNA, has no intrinsic effect on mobility, at least at the level of accuracy of the present determination; (ii) the calibration also depends on the secondary structure of the molecules under study. Though large RNA and DNA molecules of the same size behave similarly in gels when both are double-stranded (21), each migrates differently, with respect to the molecular weight, when in single or double-stranded configuration (20-22). As a matter of fact, in case of molecules of a molecular weight above $(0.3-0.4) \times 10^6$, it is possible to infer their secondary structure from the dependence of their electrophoretic mobility on various experimental parameters (21). Unfortunately, this cannot apply to the present DNA fragments, due to their small size.

Therefore, at this point, no conclusions can be drawn from the information on the strandedness of the above fragments. Data from various sources are indeed rather contradictory. Though our DNA fragments, when run in parallel with or without previous heat-denaturation, showed no differences in mobilities of both peaks, suggesting a single-stranded structure, it could be argued that some denaturation might systematically occur at the beginning of the run, due to overheating of the sample in the loading buffer that has a low ionic strength. The length of peak I fragments, agrees nearly as well with the polymerase diameter (10.5-12 nm; 105-120 Å) (23) when in single-stranded (45-52 nucleotides) or doublestranded form (22-26 nucleotide pairs). On the other hand, the very good equality observed between A and T, as well as between G and C, would fit well with a base-paired structure, as would also the data of Matsukage et al. (4) on the basis of adorption to membrane filters. In any case, their protected DNA moieties (38 nucleotide pairs) (4) are larger than ours, as are these of Sentenac et al. (75 nucleotides), although this 75-nucleotide value should be considered as a maximum (3).

DISCUSSION

The present results show unambiguously that among λ DNA fragments that are protected by RNA-polymerase against nuclease digestion, there is a discrete population of molecules containing 45–52 nucleotides residues, the existence of which is strictly dependent on sigma factor. On grounds that promoter recognition by polymerase does involve the sigma fac-

tor (9, 11), we conclude that peak I molecules analyzed here actually contain specific λ DNA sequences related to the early promoters. Whether these sequences are the promoters themselves is difficult to ascertain. They might conceivably be adjacent sequences or extend outside the promoter on one or both side (24), then possibly overlapping or including the operator region. As a matter of fact, the λ repressor and the polymerase, which mutually prevent each other's binding to λ DNA (25, 26), could do so by mere steric hindrance. The possibility remains that the differences between core and complete enzyme binding sites would be due to the protein (15) rather than the sigma subunit, both being present in complete enzyme and absent from core enzyme. However, this possibility seems very unlikely for several reasons: (a) τ has not been assigned any role in the transcription process; (b) on the contrary, the role of sigma in promoter recognition is well established (9, 11); (c) τ does not copurify with the polymerase on DNA columns (27) as σ does.

The second conclusion is that sequences are considerably enriched in A-T. Moreover, the extent of this enrichment correlates well with binding conditions likely to insure greater specificity (i.e., use of limiting polymerase concentration). It thus appears that sequences, which are rich in A-T, are contaminated with molecules of the same size, arising from nonspecific binding and having the base composition of total DNA (50% A-T). The mechanism by which these nonspecific binding sites might still be protected was explained in terms of competition between polymerase and nucleases being more in favor of the polymerases at higher polymerase concentration [see Results, Section (a)]. Such a contamination probably occurs even under our best conditions (lowest polymerase/DNA ratio). Therefore, the 67% A-T content observed here may be a minimum value. Our finding is in good agreement with the observation that RNA-polymerase of E. coli has a strong affinity for poly dAT (28) as well as for the T-rich fragments of phage f1 DNA (29). It would also fit well with the idea of a local melting of DNA during the transcription process.

The involvement of sequences that are rich in A-T DNA may not be limited to the sole RNA-polymerase-promoter recognition process. The *lac* repressor also exhibits a strong affinity for poly dAT (30). The peculiar secondary structure (increased pitch) of A-T rich DNA (67%), recently discovered by Bram (31), has led him to propose that such sequences might be a general feature of DNA segments involved in recognition and control processes. In the present case, it would imply that a double-stranded structure is recognized by the polymerase as it is by both *lac* (32) and λ (33) repressors. However, at that point, no conclusion can be drawn from our data as to the secondary structure of the DNA pieces that are analyzed here.

This laboratory is part of the Laboratoire de Pharmacologie Moléculaire no. 147, associated with the Centre National de la Recherche Scientifique. We thank Dr C. Paoletti for his support and for the use of his laboratory, Miss S. Colin for hie expert technical assistance, and Mr D. Toutain for his help in growing phage stocks. This work was supported by the Ligue Nationale Française Contre le Cancer and the Institut National de la Santé et de la Recherche Medicale.

- 1. Novak, R. L., Biochim. Biophys. Acta, 149, 593 (1967).
- 2. Blattner, F. R., Ph.D. thesis, 1968.
- Sentenac, A., A. Ruet, and P. Fromageot, FEBS Lett., 2, 53 (1968).

- 4. Matsukage, A., S. Murakami, and T. Kameyama, *Biochim. Biophys. Acta*, 179, 145 (1969).
- 5. Nakano, E., and K. Sakaguchi, J. Biochem. Tokyo, 65, 147 (1969).
- 6. Le Talaër, J. Y., and Ph. Jeanteur, FEBS Lett., 12, 253 (1971).
- Szybalski, W., in Handbook of Biochemistry (Chemical Rubber Co., 1970), p. I-35.
- 8. Green, M. H., W. S. Hayward, and P. Gariglio, Cold Spring Harbor Symp. Quant. Biol., 35, 295 (1970).
- 9. Bautz, E. K. F., and F. A. Bautz, Nature, 226, 1219 (1970).
- 10. Naono, S., and K. Tokuyama, Cold Spring Harbor Symp. Quant. Biol., 35, 375 (1970).
- 11. Hinkle, D. C., and M. Chamberlin, Cold Spring Harbor Symp. Quant. Biol., 35, 65 (1970).
- 12. Babinet, C., Biochem. Biophys. Res. Commun., 26, 639 (1967).
- 13. Burgess, R.R., J. Biol. Chem., 244, 6160 (1969).
- 14. Burgess, R. R., A. A. Travers, J. J. Dunn, and E. K. F. Bautz, *Nature*, 221, 43 (1969).
- Dubert, J. M., B. Hermier, and C. Babinet, *Biochimie*, 53, 185 (1971).
- 16. Kaiser, A. D., J. Mol. Biol., 4, 275 (1962).
- 17. Le Pecq, J. B., and R. L. Baldwin, Cold Spring Harbor Symp. Quant. Biol., 33, 609 (1968).
- 18. Peacock, A. C., and C. W. Dingman, *Biochemistry*, 6, 1818 (1967).

- Forget, B. G., and S. M. Weissman, Science, 158, 1695 (1967).
- Bishop, D. H. L., J. R. Claybrook, and S. Spiegelman, J. Mol. Biol., 26, 373 (1967).
- 21. Fisher, M. P., and C. W. Dingman, *Biochemistry*, 10, 1895 (1971).
- 22. Girard, M., and L. Marty, Bull. Soc. Chim. Biol., 51, 31 (1969).
- 23. Lubin, M., J. Mol. Biol., 39, 219 (1969).
- Zillig, W., K. Zechel, D. Rabussay, M. Schachner, V. S. Sethi, P. Palm, A. Heil, and W. Seifert, Cold Spring Harbor Symp. Quant. Biol., 35, 47 (1970).
- Hayward, W. S., and M. H. Green, Proc. Nat. Acad. Sci. USA, 64, 962 (1969).
- Chadwick, P., V. Pirotta, R. Steinberg, N. Hopkins, and M. Ptashne, Cold Spring Harbor Symp. Quant. Biol., 35, 283 (1970).
- Alberts, B. M., F. J. Amodio, M. Jenkins, E. D. Guttmann, and F. L. Ferris, Cold Spring Harbor Symp. Quant. Biol., 33, 289 (1968).
- 28. Jones, O. W., and P. Berg, J. Mol. Biol., 22, 199 (1966).
- 29. Shishido, K., and Y. Ikeda, J. Biochem. Tokyo, 68, 881 (1970).
- 30. Lin, S. Y., and A. D. Riggs, Nature, 228, 1184 (1970).
- 31. Bram, S., Nature, 232, 174 (1971).
- Riggs, A. D., H. Suzuki, and S. Bourgeois, J. Mol. Biol., 48, 67 (1970).
- 33. Ptashne, M., Nature, 214, 232 (1967).