Proc. Nat. Acad. Sci. USA
Vol. 71, No. 8, pp. 3091–3095, August 1974

Escherichia coli RNA-Polymerase Binding Sites on DNA Are Only 14 Base Pairs Long and Are Located between Sequences That Are Very Rich in A+T

(regulation/transcription/protein-nucleic acid interaction)

PAOLO U. GIACOMONI, JEAN YVES LE TALAER*, AND JEAN BERNARD LE PECQ

Laboratoire de Pharmacologie Moléculaire n° 147 associé au CNRS, Institut Gustave-Roussy, 94800-Villejuif, France

Communicated by Jacques Monod, May 28, 1974

ABSTRACT ***E. coli* DNA-dependent RNA-polymerase binding sites on DNAs of T5, T7, and lambda coliphages have been isolated according to three different methods in order to analyze the binding sites themselves as well as the nearest neighboring regions. It is shown that the binding sites are regions that are rather rich in G+C, are about 14 base pairs long and are located between DNA sequences highly enriched in A+T. The biological implications of this result are discussed.**

Le Talaer and Jeanteur have recently developed a technique for the isolation of the DNA binding sites of *Escherichia coli* DNA-dependent RNA-polymerase holoenzyme (nucleoside-triphosphate:DNA deoxynucleotidyltransferase, EC 2.7.7.6). They have isolated DNA segments which are rich in A+T, σ specific, and which represent the specific binding sites of RNA polymerase holoenzyme on DNA. These segments are about 50 nucleotides long, as deduced from gel electrophoresis analysis, and are partially resistant to *Neurospora crassa* endonuclease which specifically attacks single-stranded DNA. They concluded that these segments are double-stranded (1, 2).

We attempted to characterize these binding sites isolated according to their procedure in more detail by means of physico-chemical analysis. In the current such study, it was observed that protruding single strands were left on these DNA segments. This finding was expected since in the purification procedure a single exonuclease, working in only one direction (from the 3'OH termini) was used. It was, therefore, apparent that it was possible to analyze the binding sites themselves as well as the adjacent sequences. We were led to modify the preparation procedure in order to try to eliminate the protruding single-stranded segments. In this paper we describe a methodology for isolating the binding sites of *E. coli* DNA-dependent RNA-polymerase holoenzyme on DNAs from phages T5, T7, and lambda, with and without the protruding single strands.

We have observed that the segments without the protruding parts are 14 base pairs long, homogeneous in size and rather rich in G+C, whereas, the segments with the protruding parts are A+T rich, as previously observed, and 38 nucleotides long. A quantitative analysis of the results leads to the conclusion that RNA-polymerase binding sites on DNA are rather rich in G+C and that they are located between DNA regions constituted of sequences that are very rich in A+T.

Abbreviation: EDTA, ethylenediaminetetraacetate.

* Present address: Service de Biologie, Centre Régional François Baclesse. 14018-Caen-France.

MATERIALS AND METHODS

Enzymes. DNA-dependent RNA polymerase holoenzyme (EC 2.7.7.6) was purified from *E. coli* A 19 according to Burgess and Travers (3); the activity peak from the high-salt agarose column yielded an active (500 units/mg) and almost pure enzyme (up to 95%), fully saturated with σ subunit, as deduced from quantitative analysis by electrophoresis on acrylamide gel.

Pancreatic DNase (EC 3.1.4.5.), viper's venom phosphodiesterase (EC 3.1.4.1) and alkaline phosphatase from calf intestine (EC 3.1.3.1) were purchased from Worthington, Calbiochem, and Boehringer, respectively.

Micrococcal nuclease (EC 3.1.4.7), obtained from Boehringer, was tested according to Heins, Taniuchi, and Afinsen (4).

Phages and DNAs. ^{32}P-Labeled T7, T5, and λ were grown and purified as previously described (1, 2). DNAs were purified by three phenol extractions, followed by three ether extractions, and exhaustively dialyzed against 0.01 M Tris·HCl buffer at pH 7.9, 0.01 M NaCl. They were monitored spectrophotometrically, according to Felsenfeld and Hirschman (5).

Isolation and Determination of Base Composition of RNA Polymerase Binding Sites. RNA polymerase holoenzyme is incubated with DNA at 37° for 15 min in 0.04 M Tris·HCl at pH 7.9, 0.01 M NaCl, 0.01 M $MgCl_2$, 0.01 M $CaCl_2$. The solution contains 100 μg/ml of sterile bovine-serum albumin for protecting the polymerase.

The complex DNA·RNA polymerase is then submitted to nucleolytic attack according to one of the following methods.

Method 1. Pancreatic DNase 1 is added to the incubation medium to a final concentration of 200 μg/ml for 30 min, at 37°; then viper's venom phosphodiesterase is added to a final concentration of 7.5 units/ml and the incubation is continued for an additional 30 min, the reaction is then stopped by addition of ethylenediaminetetraacetate (Na Salt) (EDTA) to a final concentration of 0.02 M.

Method 1 is the degradation procedure of Le Talaer and Jeanteur (1) in which the pronase step has been omitted.

Method 2. Micrococcal nuclease is added to the incubation medium at a final concentration of 10 units/ml for 30 min at 37°, the incubation is continued for an hour in the presence of alkaline phosphatase (3.5 units/ml) and viper's venom phosphodiesterase (7.5 units/ml) until EDTA is added to stop the degradation.

Method 3. The digestion is first carried out by pancreatic DNase and viper's venom phosphodiesterase as in *Method 1;*

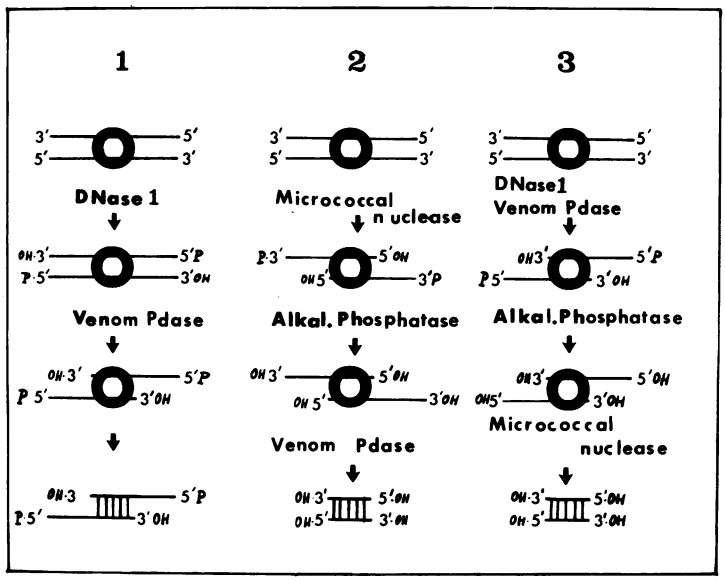

FIG. 1. The three different methods used for the purification of the RNA polymerase binding sites. In *Method 1* the pancreatic DNase leaves 3'OH and 5'P termini. The exonucleolytic attack is carried out by viper's venom phosphodiesterase alone, which digests from the 3'OH termini; it is, therefore, expected that single-stranded segments terminated by 5'P are left undegraded. *Method 2* has been developed to eliminate the single-stranded parts, micrococcal nuclease acts as an endonuclease, leaving 3'P and 5'OH, and as an exonuclease digesting from 5'OH termini (see ref. 10). The alkaline phosphatase is added to yield free 3'OH termini, thus allowing the viper's venom phosphodiesterase to act. *Method 3* has been developed to demonstrate the possibility of cutting out the protruding segments left in *Method 1*, thus establishing the relation between the segments prepared by *Methods 1* and *2*.

the incubation is then continued for an hour in the presence of alkaline phosphatase (3.5 units/ml) and micrococcal nuclease (10 units/ml); EDTA is again added to stop the degradation.

The extent of the nucleolytic attack is monitored by determining the percentage of radioactivity insoluble in 7% trichloroacetic acid and adsorbable on norit (organic phosphates).

Immediately after the addition of EDTA, the segments of DNA that are protected against the nucleolytic attack by RNA-polymerase, are purified by three phenol extractions, alcohol precipitation, and centrifugation at 15,000 rpm in an AJ21 Beckmann rotor for 30 min.

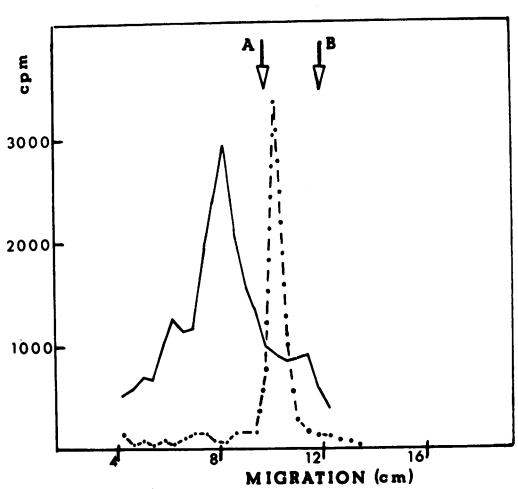

FIG. 2. Profile of migration of T5 DNA·RNA polymerase binding sites prepared according to *Methods 1* and *2* and subjected to electrophoresis on 10% acrylamide gels. The *arrows* indicate (*A*) the distance migrated by bromophenol blue, and (*B*) the distance migrated by the peak II of Le Talaer and Jeanteur (refs. 1 and 2). ——, migration profile of end products of *Method 1*. ·–·–·, migration profile of end products of *Method 2*.

The pelleted DNA segments are resuspended in 0.01 M Tris·HCl buffer at pH 7.9, 0.01 M NaCl, 0.001 M EDTA. After characterization by gel electrophoresis, they are then extracted from the gel, hydrolyzed to nucleotides, and analyzed by thin-layer chromatography for determining their base composition, as previously described (2).

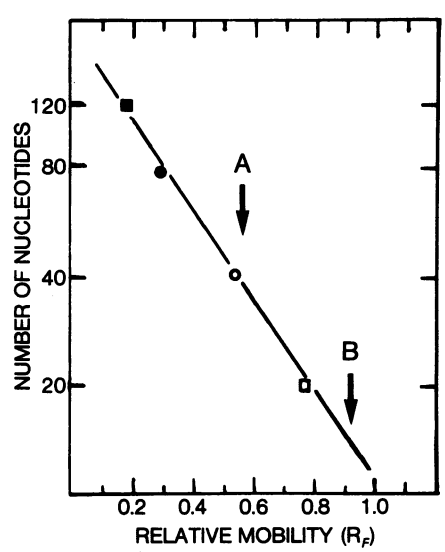

FIG. 3. Length determination of denaturated binding sites by electrophoresis on 15% acrylamide–7 M urea gels which were standardized with ^{32}P-labeled 5S RNA (■), tRNA (●), and partially degraded 5S RNA yielding polyribonucleotides 40 (○) and 20 (□) nucleotides long. The values of R_F are determined relatively to the migration of bromophenol blue. The number of nucleotides are on a log scale. The *arrows* show (*A*) the binding sites prepared according to *Method 1* and (*B*) the binding sites prepared according to *Methods 2* and *3*.

TABLE 1. *Properties of RNA polymerase binding sites prepared by the three different methods*

Method		1			2			3		
DNA	R	% G+C	L	R_F	% G+C	L	R_F	% G+C	L	R_F
T5	0.15	39	38	0.8	70	14	1.04	—	14	1.04
T7	0.1	35	38	0.8	50	14	1.04	50	14	1.04
λ	0.2	43	38	0.8	64	14	1.03	54	14	1.03

R is the ratio of the weight RNA polymerase to that of DNA; L is the length of protected DNA segments determined by electrophoresis on 15% acrylamide–7 M urea gels. R_F indicates the mobility of isolated binding sites (relative to bromophenol blue) on 10% acrylamide gel electrophoresis. The ratios of the weight enzyme to that of DNA have been chosen in order to select only the specific binding sites (see refs. 1 and 2) and thus avoid contamination with secondary binding sites.

Size Determination of RNA Polymerase Binding Sites. Size of the binding sites was determined by heating to 100° for 3 min in the presence of 7 M urea, followed by chilling in ice and electrophoresis on 15% acrylamide–7 M urea gels, as described by Gilbert and Maxam (6). The standardization was performed using ^{32}P-labeled intact or partially degraded 5S RNA and total tRNA from *E. coli*. 5S RNA submitted to partial attack of T1 RNase yields polyribonucleotides that are 20 and 40 bases long (7).

We made the assumption that there is no difference in mobility between single-stranded polyribonucleotides and single-stranded polydeoxyribonucleotides of the same length.

The values of R_F were determined in relation to the mobility of bromophenol blue, used as internal marker.

Thermal Denaturation. *Tm* and the hyperchromic effect upon thermal denaturation were determined by means of a Hellma water jacketed microcuvette with a 1-cm optical path which fitted the sample chamber of a Beckmann Acta 3 spectrophotometer connected with a thermoregulated bath. The sample was in 0.15 M NaCl, 0.015 M Na citrate. The total volume was 0.25 ml. The temperature was determined with a calibrated microthermistance fitting the cuvette. Evaporation was avoided by layering a drop of paraffin oil on the sample.

Ethidium Bromide Binding. The measurement was performed by spectrofluorometry as already described (8) in 0.2 M Tris·HCl at pH 8, 0.2 M NaCl at 20°. Because of the small quantity of available material, the determination was done on 0.2 ml at a sample concentration of about 3 μg/ml using a single photon counting spectrofluorometer built in this laboratory (9).

RESULTS

The expected structures of the DNA segments isolated according to the three methods are illustrated in Fig. 1.

In Fig. 2 we report the migration profile on 10% acrylamide gel electrophoresis of the RNA polymerase binding sites isolated from T5 DNA according to *Methods 1* and *2*. The migration profiles of binding sites from T7 and λDNA are identical. The R_F values of the binding sites purified according to *Method 3* and *2* are exactly equal.

According to Le Talaer and Jeanteur (1), for 10% acrylamide gels an R_F of 0.8 corresponds to a length of about 50 base pairs, and an R_F of 1.05 corresponds to about 15 base pairs.

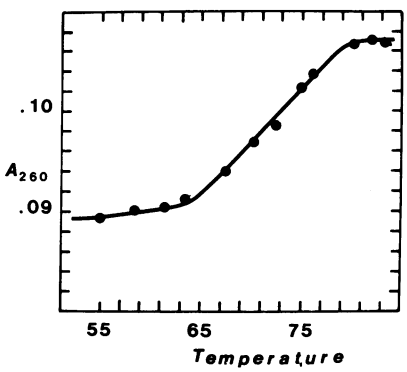

FIG. 4. Thermal denaturation profile of RNA polymerase binding sites from T7 DNA, prepared according to Le Talaer, Kermici, and Jeanteur (2).

Since their standardization was relative to single-stranded polyribonucleotides, and their DNA pieces were at least partially double-stranded, it is impossible to relate precisely the migrated distance to a definite number of nucleotides. We determined, therefore, the size of the heat denatured binding sites by electrophoresis on 15% acrylamide, 7 M urea gels. In Fig. 3 we show the results of this determination: the binding sites isolated by *Method 1* are 38 ± 2 nucleotides long, whereas, the binding sites isolated by *Methods 2* and *3* are 14 ± 2 nucleotides long. This last measurement is possibly slightly affected by the absence of a phosphate group at the 5′ end.

The base composition of the RNA polymerase binding sites isolated from the DNAs of phages T5, T7, and lambda, by the three methods described, is reported in Table 1.

The results of thermal denaturation and of ethidium bromide binding to RNA polymerase binding sites from T7 DNA, prepared according to Le Talaer, Kermici, and Jeanteur (2) are reported in Figs. 4 and 5. Thermal denaturation yields a hyperchromic effect of only 20% and the *Tm* value is 73°. The *Tm* value of short, double-stranded polynucleotides depends on length, concentration, and base composition. Let us use the equation

$$\frac{1}{Tm} = \frac{1}{Tc} + \frac{R \ln c}{N\Delta H_1}$$

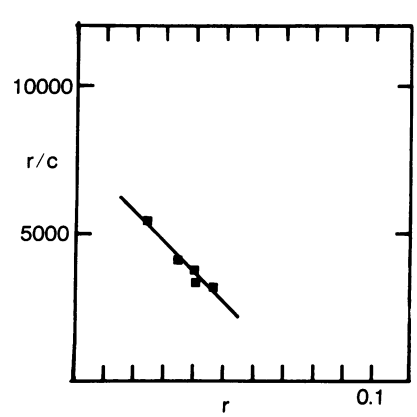

FIG. 5. Scatchard plot of the complex formed by ethidium bromide with RNA polymerase binding sites from T7 DNA, prepared according to Le Talaer, Kermici, and Jeanteur (2). r indicates the number of intercalated dyes per nucleotide and c is the concentration of the free dye.

from Uhlenbeck, Martin, and Doty (11) where T_c is the melting temperature of a DNA with the same G+C content as the polynucleotide, c is the molar concentration of the polynucleotide strands, N is the length in nucleotides of the polynucleotide and ΔH_1 is the enthalpy of formation of one base pair, (7.6 kcal/mole).

If $T_c = 90°$, as given by Marmur and Doty (12) for a DNA of 50% G+C (the same G+C content as of the double-stranded part of T7 DNA·RNA polymerase binding sites), the equation is satisfied for $N = 15$.

The results of the binding of ethidium bromide to RNA-polymerase binding sites show that the value of the association constant is the same (1.02×10^5 M^{-1}) as for the complex formed by ethidium bromide with double helical DNA, but that the apparent number of intercalation sites is three times smaller than expected for a perfectly double-stranded structure. These data can be interpreted to mean that one-third of these DNA segments is double-stranded. Since the total length of each strand is 38 nucleotides, the length of the double-stranded part is, therefore, equal to 13 nucleotides.

DISCUSSION

The procedures for preparing the RNA polymerase binding sites would be expected to result in the structures represented in Fig. 1. Furthermore, from these structures we can predict a series of physical and chemical properties which are indeed experimentally observed.

(*a*) The broadness of the peak observed after subjecting to electrophoresis the binding sites isolated according to *Method 1* is expected; the protruding single strands are heterogeneous in length because pancreatic DNase cuts at random. On the other hand, the binding sites prepared according to *Methods 2* or *3* should represent DNA pieces of definite size and, therefore, a very sharp peak in electrophoresis must be observed; this is indeed the case.

(*b*) The binding sites prepared according to *Methods 2* and *3* are equal in size and shorter than those prepared according to *Method 1*, as expected. Their lengths are respectively 14 ± 2 and 38 ± 2 nucleotides. This implies that binding sites prepared according to *Method 1* consist of a double-stranded part, 14 base pairs in length and of two protruding single-stranded segments, one at each side, about 24 nucleotides in length. In contrast, binding sites prepared by *Methods 2* and *3* consist of double-stranded DNA segments 14 base pairs in length.

(*c*) This length of the double-stranded part of the binding sites prepared according to *Method 1* (14 ± 2 base pairs) is in a remarkable agreement with the values deduced from the ethidium bromide-binding data (13 base pairs) and from the *Tm* of these binding sites (15 base pairs).

The preceding determinations permit us to compute the base composition of the protruding single-stranded regions of binding sites prepared by *Method 1*. The observed base composition of these segments is the weighted mean of the base composition of single-stranded and double-stranded regions.

Let q_s and q_d be the fraction of the DNA pieces in single and double-stranded structure. Let f_s and f_d represent, respectively, the fraction of A+T in single- and double-stranded structures.

The observed total base composition f_T is then given by $f_T = q_s f_s + q_d f_d$. From the results in Table 1 we calculate f_s to be about 0.82 for T5 DNA, about 0.75 for T7 DNA, and about 0.72 for λ DNA.

Thus, for the bindings sites prepared from the three phages according to *Method 1*, the protruding parts are very much enriched in A+T.

Furthermore, the length of the A+T rich sequences on each side of the RNA polymerase binding sites are not exactly known and they need not be exactly 24 nucleotides long. This means that these sequences, if shorter, can only have a higher percent of A+T than we have deduced. For instance, if one assumes that these A+T rich sequences are only 12 nucleotides long and that the remaining protruding single strands contain 50% of each base pair, one would have a 100%, A+T sequence that is 12 base pairs long on each side of the RNA polymerase binding sites.

The enrichment in G+C of RNA polymerase binding sites (T5, λ) suggests the role of pyrimidine clusters in DNAs of various organisms proposed by Szybalski, Kubinski, and Sheldrick (13).

It has recently been observed that more than one polymerase molecule binds per promoter (14, 15) and it has been inferred that storage sites close to a promoter must exist. The DNA fragments prepared in this work would, therefore, represent mostly the storage sites.

The DNA representing the storage sites would be rather rich in G+C. As soon as transcription starts from a promoter, a RNA polymerase molecule, previously in a storage site, would jump to the promoter site and thus be ready for a new transcription.

It is known that a change in DNA conformation, associated with a high activation energy, is necessary for initiation of transcription (16, 17). The alternacy of A+T rich and G+C rich regions could well play a role in such a conformational change which would be necessary in the sequential displacement of RNA polymerase from storage sites to promoter.

Such a mechanism would permit a much faster rate of transcription since the stored RNA polymerases have to diffuse in only one direction to reach the promoter.

The A+T rich regions could also be involved in increasing the stability of the complex DNA·RNA polymerase, in the same manner as that proposed for the complex operator repressor by Gilbert, i.e., "... the presence of arms of DNA around the operator might slow up the decay by permitting the repressor to wander off the operator site and then be recaptured" (18). It is in fact well known (19) that RNA polymerase has a strong affinity for poly(dA, dT).

The authors are indebted to Miss Simone Colin for her expert and skillful technical assistance. The generous gift of ^{32}P-labeled, *E. coli*, pure 5S RNA and total tRNA from Dr. Jean Feunteun, as well as of T_1 RNase, is gratefully acknowledged. This work has been partially supported by a long term EMBO fellowship awarded to one of the authors (P.U.G.) and by a grant from the Délégation Générale à la Recherche Scientifique et Technique.

1. Le Talaer, J. Y. & Jeanteur, Ph. (1971) *Proc. Nat. Acad. Sci. USA* **68,** 3211–3215.
2. Le Talaer, J. Y., Kermici, M. & Jeanteur, Ph. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 2911–2915.
3. Burgess, R. & Travers, A. (1971) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 851–863.
4. Heins, J. N., Taniuchi, H. & Afinsen, C. B. (1966) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 1, pp. 79–84.
5. Felsenfeld, G. & Hirschman, S. Z. (1965) *J. Mol. Biol.* **13,** 407–427.
6. Gilbert, W. & Maxam, A. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 3581–3584.

7. Gray, P. N., Bellemare, G., Monier, R., Garrett, R. A. & Stoffler, G. (1973) *J. Mol. Biol.* **77**, 133–152.
8. Le Pecq, J. B. & Paoletti, C. (1967) *J. Mol. Biol.* **27**, 87–106.
9. Paoletti, J. (1972) Thèse d'Etat, University of Paris.
10. Cuatrecasas, P., Fuchs, S. & Afinsen, C. B. (1967) *J. Biol. Chem.* **242**, 1541–1547.
11. Uhlenbeck, O. C., Martin, F. H. & Doty, P. (1971) *J. Mol. Biol.* **57**, 217–229.
12. Marmur, J. & Doty, P. (1962) *J. Mol. Biol.* **5**, 109–118.
13. Szybalski, W., Kubinski, H. & Sheldrick, P. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 123–127.
14. Schafer, R., Kramer, R., Zillig, W. & Cudny, H. (1973) *Eur. J. Biochem.* **40**, 367–373.
15. Willmund, R. & Kneser, H. (1973) *Mol. Gen. Genet.* **126**, 165–175.
16. Saucier, J. M. & Wang, J. C. (1972) *Nature New Biol.* **239**, 167–170.
17. Travers, A., Baillie, D. L. & Pedersen, S. (1973) *Nature New Biol.* **243**, 161–163.
18. Gilbert, W. (1972) *Polymerization in Biological Systems*, Ciba Foundation Symposium 7 (new series), eds. Wolstenholme, G.E.W. & O'Connor, M. (Elsevier, Amsterdam), pp. 245–255 .
19. Jones, O. W. & Berg, P. (1966) *J. Mol. Biol.* **22**, 199–209.