Volume 6 Number 4 April 1979 Nucleic Acids Research

Operators and promoters in the O_R region of phage 434

Vincenzo Pirrotta *

Biozentrum der Universität Basel, Klingelbergstrasse 70, CH - 4056 Basel, Switzerland

Received 6 December 1978

ABSTRACT

The O_R operator region of phage 434 contains three 14 bp blocks with sequence ACAAGA-A--TTGT which are presumed to be the 434 repressor recognition sites. Operator constitutive mutations are located in two of these blocks, while a mutation affecting repressor levels in the lysogenic state is located in the third. Two transcripts obtained in vitro, one leftwards and one rightwards, are tentatively identified as the P_{rm} and P_R transcription starts. The arrangement of the 434 operator region appears to be very similar to that of the O_R of phage λ.

INTRODUCTION

The early genes of phage λ are regulated by a repressor protein interacting with two remarkable structures on the phage DNA: the O_L and O_R operators, controlling respectively the leftward and the rightward early genes. Each operator consists of three 17 bp recognition sequences separated by 3-7 bp arranged in order of their affinity for repressor (1,2). The promoters P_L and P_R are located within the operators and overlap with the first and second repressor binding sites, O_L1-O_L2 and O_R1-O_R2 (3). As a consequence, when repressor binds to these sites, it blocks the access of RNA polymerase to the corresponding promoters and thereby represses transcription initiation.

While in O_L the role of the third repressor binding site, if any, is unclear, in the right hand operator, O_R3 serves another important regulatory function. O_R3 and part of O_R2 overlap with another promoter, P_{rm}, which is responsible for the transcription of the repressor gene, cI, in the lysogenic state. P_R and P_{rm} are symmetrically disposed at the ends of O_R and are oriented in opposite directions (4,5).

When repressor binds to O_R1, it not only represses P_R, but at the same time it activates P_{rm}. The mechanism of this activation is not clear but might be analogous to the activation of the lac promoter by CAP factor or it might simply result from blocking the competing P_R promoter and therefore allowing the polymerase to bind to P_{rm}.

Because of the lower affinity of repressor for O_R3, this site is not occupied until higher repressor concentrations are achieved in the cell. When this happens, repressor bound to O_R3 blocks the access of polymerase to P_{rm} and turns off further transcription of the repressor gene. Repressor synthesis is therefore autoregulated, resulting in the maintenance of a low but steady level of repressor in the lysogenic cell.

Phage 434 is a lambdoid phage so closely related to λ that many of its genes can recombine genetically with those of λ. In 1957 Kaiser and Jacob (6) exploited this fact to construct a hybrid phage, λ imm^{434}, in which the immunity region of λ is replaced by that of phage 434.

λ imm^{434} contains the cI gene, the cro gene and the two operators and promoters of 434 and therefore responds to control by the 434 repressor but is insensitive to the λ repressor. Conversely, λ is insensitive to the 434 repressor. In vivo experiments also indicate that, as in λ, the expression of the repressor gene in a 434 lysogen depends on the presence of active 434 repressor presumably acting on O_R to activate a P_{rm} promoter (7).

In the work reported here, I have determined the sequence of the 434 O_R region and of mutations affecting the operator. An analysis of these sequences and of transcription initiating from this region, indicates that the 434 O_R is also a triple structure with a functional organisation similar to that of the λ O_R. However, the details of the relationship between promoters and repressor binding sites differ from that of λ and the repressor recognition sequence is altogether unrelated.

METHODS

Purification of 434 repressor

10 g of E. coli C600 (λ imm^{434}T) were resuspended and sonicated in

20 ml of 50 mM Tris pH 7.5, 0.2 M KCl and 10 mM $MgCl_2$, then centrifuged at 35 K rpm for 1 hour at 4°. The supernatant was fractionated with ammonium sulfate. The fraction precipitating between 35 and 50% saturation was resuspended and dialysed against 20 mM potassium phosphate pH 7.5, then applied to a 1.5 x 20 cm phosphocellulose column equilibrated with the same buffer. The column was developed with a gradient 0.02-0.15 M potassium phosphate and fractions were assayed for preferential binding to ^{32}P-λ imm^{434}DNA (8,9). The peak fractions, still containing considerable nonspecific binding activity, were dialysed against 50% glycerol, 20 mM phosphate pH 7.5 and stored at -20°.

Selection of fragments containing operators

^{32}P-λ imm^{434}DNA was cut with appropriate restriction enzymes. 100000 cpm aliquots were incubated with 434 repressor in 0.3 ml BW buffer (10 mM Tris pH 7.0, 20 mM KCL, 5% dimethyl sulfoxide and 5 mM EDTA) in the presence of 10 µg sonicated calf thymus DNA to absorb nonspecifically binding proteins. EDTA is necessary as the presence of Mg^{++} inhibits 434 repressor binding (8). After 10 min at room temperature, the mixture was filtered slowly through a Sartorius BA/85 membrane filter and washed twice with 10 drops of BW buffer. The fragments retained on the filter were extracted by shaking with 0.2 ml BW buffer saturated with 1-2 drops of phenol. The extracted material was precipitated with four volumes of ethanol and then analysed by electrophoresis through a 2.2% acrylamide - 0.7% agarose mixed gel in 90 mM Tris-borate pH 8.3. The gel was then dried and autoradiographed.

End labeling and DNA sequencing

HpaII 1350 or Eco RI 1280 fragments were labeled at the 5' ends with polynucleotide kinase and γ-^{32}P-ATP (10). The labeled fragment was cleaved with either Eco RI or HpaII as appropriate and the 210 bp fragment labeled at one end only was isolated by gel electrophoresis. The purified fragment was sequenced by the Maxam-Gilbert method, using the G>A, A>G, C and C+T reactions (10). The products were displayed by electrophoresis through a 20% acrylamide gel containing 7 M urea. The sequence was confirmed using the MboII 120 fragment whose strands were separated after kinase labeling.

Transcription

Operator fragments, either Eco RI-HpaII 210 or MboII 120 (2-5 p moles) were incubated in 20 µl transcription buffer (20 mM Tris pH 7.9, 40 mM KCL, 10 mM $MgCl_2$, 0.1 mM EDTA and 0.1 mM dithiothreitol, with 2-5µmoles RNA polymerase holoenzyme in the presence of 20 µM ATP and GTP. After 10 min at 37°, 4 µg heparin were added, followed two minutes later by UTP and CTP to a final concentration of 10 µM. One or more of the NTPs were labeled with ^{32}P at the α-position (100-300 mCi/µmole). After 10 min at 37°, the reaction was stopped by adding 20 µg tRNA and 5 volumes of ethanol. The precipitated product was resuspended and analysed by electrophoresis through a 12% acrylamide gel containing 7M urea.

Analysis of the transcription product

Bands were eluted from the gel of the transcription product, and analysed as previously described (1). Fingerprints were obtained by separating the products of RNase T1 digestion with electrophoresis on cellulose acetate pH 3.5, followed by homochromatography in the second dimension on DEAE-cellulose thin layers (12) using homomix VI (13). After autoradiography, oligonucleotides were eluted from the thin layer, digested with pancreatic RNase and analysed by electrophoresis on DEAE paper, pH 3.5. Sequences could be deduced for the most part by comparing the analyses of corresponding oligonucleotides labeled separately with each NTP.

RESULTS

Restriction mapping of the operators

When repressor binds to DNA fragments containing the operators, the resulting complex is selectively retained on nitrocellulose membrane filters. Fragments thus isolated can be eluted from the filter and displayed by gel electrophoresis. Some typical experiments are shown in Fig.1 while Fig. 2 summarizes the mapping results obtained for fragments containing O_R. 434 repressor selects two fragments out of an Eco RI digest of λ imm^{434}DNA. This finding is consistent with the fact that λ imm^{434}DNA contains one Eco RI site within but very close to the right end of the 434 immunity region (14).

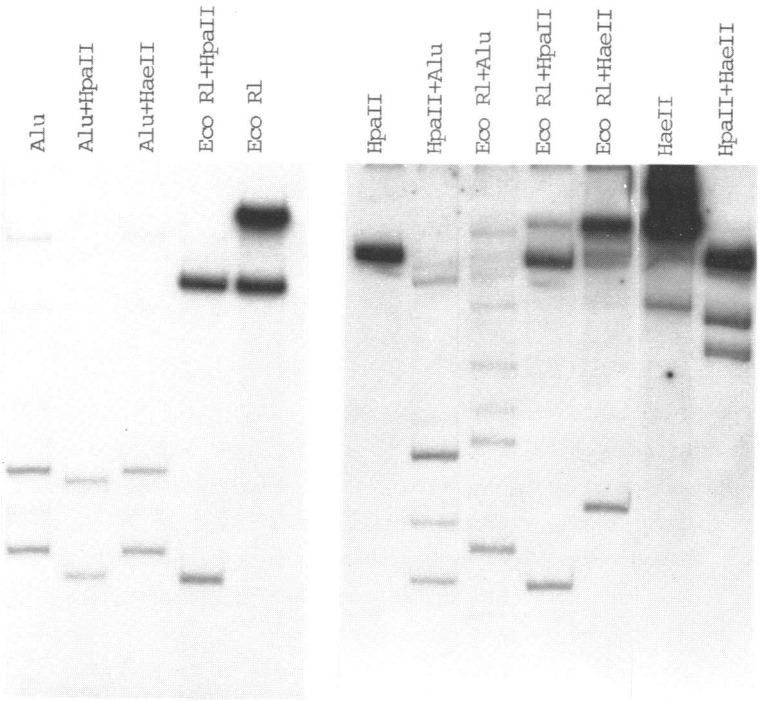

Figure 1. Gel electrophoresis of repressor binding DNA fragments. ^{32}P-λ imm^{434} DNA digested with various restriction enzymes was incubated with 434 repressor. Fragments retained on a membrane filter were then electrophoresed through a 4% acrylamide - 0.7% agarose gel -.

The next Eco RI site to the left of the immunity region is several thousand base pairs away, while to the right the next site is located in the O gene, only a thousand base pairs from the edge of the immunity region (15). The large fragment therefore contains O_L, while the small fragment, 1280 bp long, contains O_R. The Eco RI site then must be about

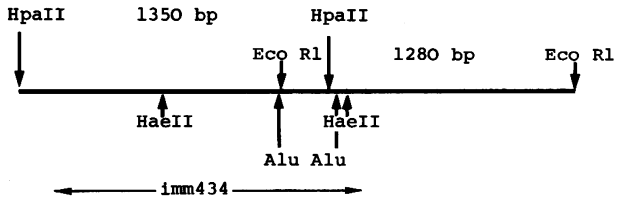

Figure 2. Partial restriction map of the λimm^{434} immunity region.

300 bp from the right edge of the immunity region and very close to O_R.

Repressor binds to only one fragment, 1350 bp long, out of a HpaII digest of λ imm^{434} DNA. This fragment contains both O_R and O_L as shown by the fact that it is cut by Eco RI into one 1140 bp and a 210 bp fragment both of which bind to 434 repressor. Since the two operators are separated by the cI gene which must be at least 600 bp long, the larger fragment must contain O_L and most of the cI gene while the shorter fragment contains O_R (Fig. 2). The Eco RI-HpaII 210 fragment was therefore used for sequence studies.

Sequence of the 434 O_R region

The Eco RI-HpaII 210 fragment was labeled with polynucleotide kinase either at the Eco RI or at the HpaII 5' end and sequenced by the Maxam-Gilbert chemical degradation method. Inspecting the sequence shown in Fig. 3 for triply repeated sequence features which, by analogy with the λ O_R, might correspond to repressor binding sites, we find that there are in fact three blocks of 14 bp each with the sequence ACAAGA-A--(T)TGT. Out of 14 base pairs only three (or four) are non-homologous. Like the λ recognition sequences, these 14 bp blocks contain a center of partial symmetry, located in this case between the seventh and the eighth nucleotide pair. By analogy with λ, these

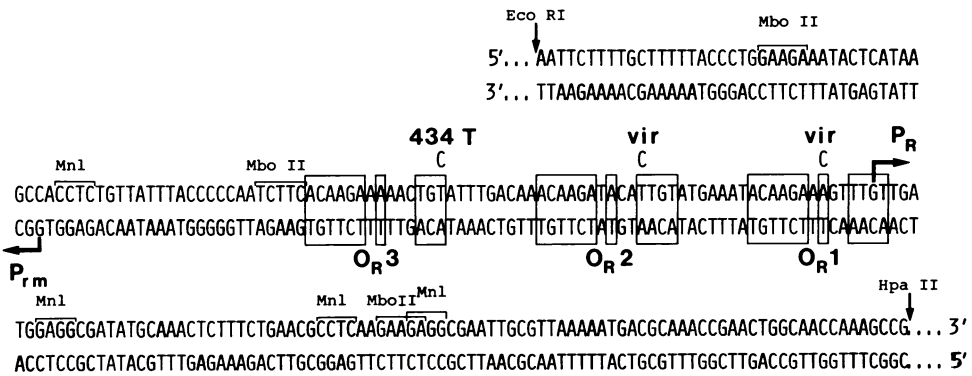

Figure 3: Sequence of the Eco RI-HpaII 210 fragment. The boxes indicate homologous nucleotides in the three presumed recognition sites. The 434T and virulent mutations are shown as well as the startpoints for right- and leftwards transcription. The sequence is very rich in recognition sites for the Mbo II and Mnl restriction enzymes which cut 8 bp in the 3' direction from GAAGA and CCTC respectively.

could correspond to O_R1, O_R2 and O_R3. The site furthest to the left, as expected for O_R3, is less symmetric than the other two and the extent of the homology is decreased by one nucleotide pair.

Mutations affecting O_R

To identify functional elements in this region, I have sequenced the same DNA fragment isolated from a λ imm^{434} virulent mutant. This phage carries true operator mutations which in vitro decrease the affinity of O_R for 434 repressor and in vivo result in the constitutive expression of the right hand operon (Pirrotta, unpublished experiments). Compared to the wild type sequence, the virulent mutant contains two nucleotide changes, one in the O_R1 site and one in O_R2. In both cases the nucleotides affected are part of the homology (Fig. 4).

Another mutation which might affect the O_R region is imm^{434}T. This mutation results in the production of 5-10 times more repressor in a lysogen of λ imm^{434}T than in a lysogen of wild type λ imm^{434} (8). Genetic crosses locate this mutation very close to but to the left of the virulent mutations described above (Pirrotta & Segrè, unpublished experiments). The sequence of the O_R region from λ imm^{434}T shows that it contains one nucleotide change in the region corresponding to O_R3.

P_R promoter

RNA polymerase binds to the Eco RI-HpaII 210 fragment as well as to a smaller fragment of 120 bp produced from it by cleavage with MboII. From this binding site the polymerase can initiate transcription. Figure 5 shows an RNase T1 fingerprint of the transcript labeled with α-^{32}P-GTP. Analysis of the oligonucleotides from this or from fingerprints labeled with all four nucleotides confirms that this is a rightward transcript. All the oligonucleotides can be fitted in the DNA sequence, beginning with the UUG corresponding to the right end of O_R1. The only spot in the fingerprint which remains unassigned migrates like pGp upon electrophoresis on DEAE paper pH 3.5. I conclude that transcription initiates with G as indicated in Fig. 3 from within the O_R1 site.

The P_{rm} promoter

In some transcription experiments using the Eco RI-HpaII 210 fragment, a new shorter transcript was obtained. In the experiment shown in Fig. 6,

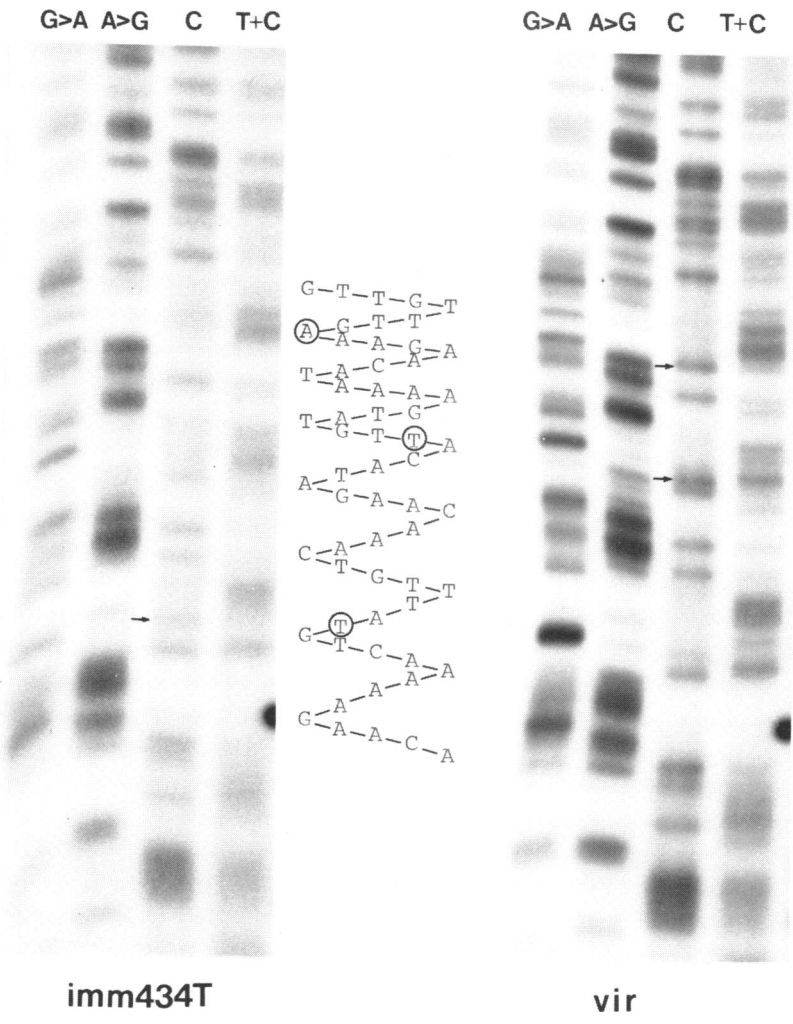

Figure 4. Sequence gels of the virulent and 434T mutants. In the middle is shown the wild type sequence. Nucleotides affected by the mutations are circled and bands corresponding to the altered nucleotides are indicated by arrows.

the transcript was separately labeled with each α-^{32}P-NTP. The sequence which was deduced from the analysis of the fingerprints could only come from the leftward DNA strand, close to the Eco RI end and initiating from a site to the left of O_R3.

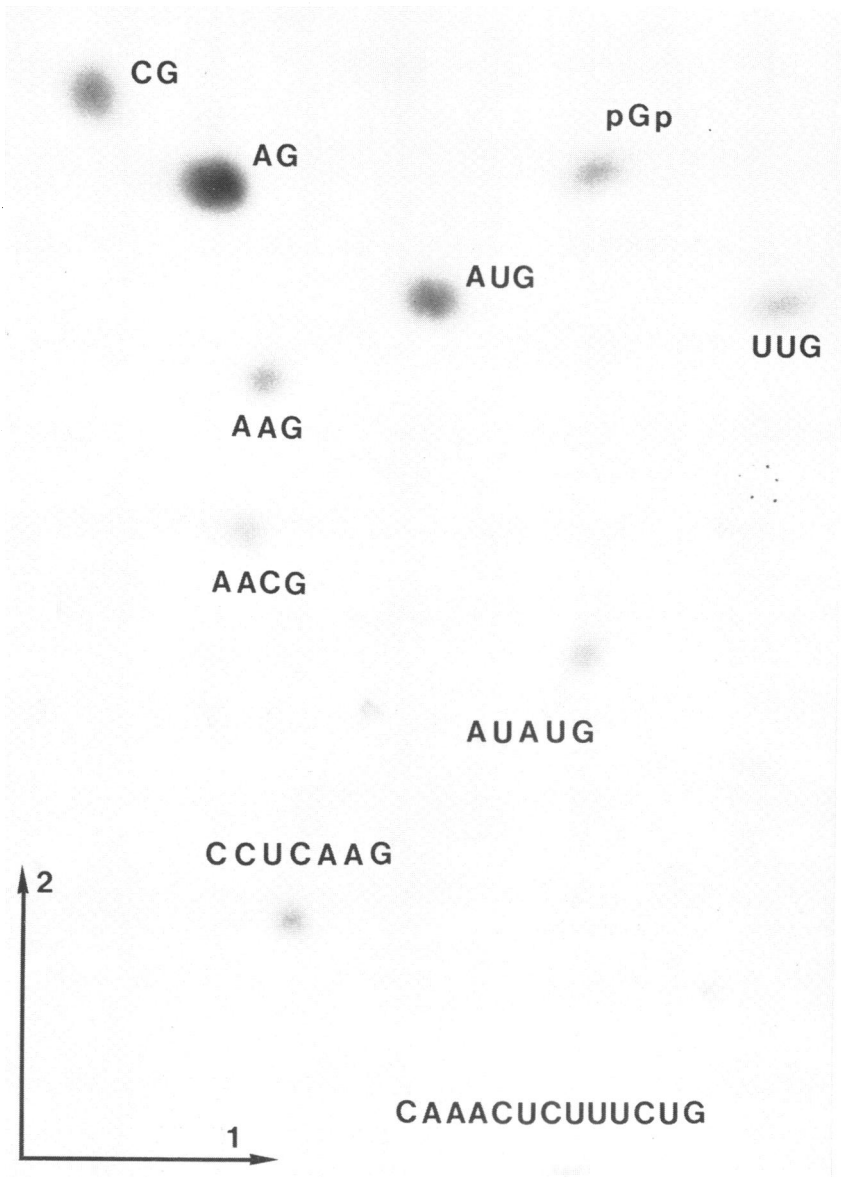

Figure 5. Fingerprint of the rightwards transcript. The MboII 120 fragment was used as template in a transcription reaction labeled with α-^{32}P-GTP. The major transcription product was treated with RNase T1 and fingerprinted by electrophoresis on cellulose acetate pH 3.5 in the first dimension and homochromatography on DEAE-cellulose thin layer in the second dimension.

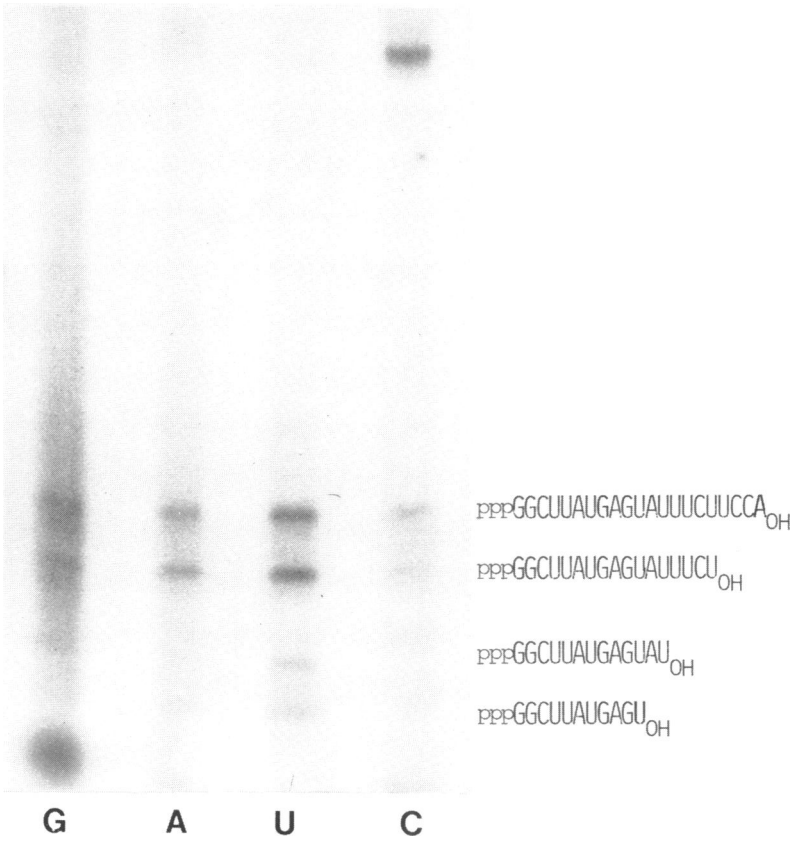

Figure 6. Gel electrophoresis of the leftwards transcript. The EcoRI-HpaII 210 fragment was transcribed with a 5-10 fold molar excess of RNA polymerase, in the presence of 12% glycerol. The new short transcripts observed were fingerprinted and analysed. The sequences were deduced by comparing the results of the four parallel reactions.

Because of the more heterogeneous incorporation in the G reaction shown in Fig. 6, the 5' triphosphate nucleotide could not be demonstrated unequivocally. However, the presence of a Gp(C) in the RNase T_1 fingerprint, and the absence of a triphosphate labeled by any of the other NTPs indicate that the 5' end is most likely pppGp.

From its position relative to O_R, this transcript is most likely to correspond to the P_{rm} promoted transcription start, although the present evidence is not sufficient to prove it.

The optimal conditions for this transcription start are not clear though they probably require a high ratio of polymerase to DNA and the presence of glycerol. Under similar conditions, however, this transcript was sometimes the major product while other times it was only a minor species accompanying the rightwards transcript.

DISCUSSION

The three 14 bp blocks indicated in Fig. 3 have all the properties expected for repressor binding sites analogous to those in the lambda O_R operator. They are nearly homologous, with O_R3 deviating by only one base pair from the sequence common to the other two. They have a high degree of symmetry, with the axis passing between two nucleotide pairs rather than through one as in λ. One half of the sequence is more conserved than the other half. The blocks are separated by 6-9 base pairs and are somewhat shorter than the 17 base pairs of the λ recognition sites. This may be related to the fact that the 434 repressor is slightly smaller than the repressor and slightly less specific in its binding properties in vitro (8,9).

Mutations which render the operator constitutive are located in the first two of these blocks, demonstrating their involvement in repressor binding. In both cases shown here, the mutations affect bases presumed to be important for recognition because they are conserved in the three blocks.

The 434 T mutation is located within O_R3 and further reduces its homology with the other two sites. Since it increases the level of repressor synthesis in the lysogenic state, 434 T might be supposed to be an up mutation in the P_{rm} promoter. The following argument makes this conclusion unlikely. The limitation to the activity of the P_{rm} is probably set not by the efficiency of the promoter itself but by the affinity of O_R3 for repressor. If O_R3 and P_{rm} overlap as they do in λ, when repressor reaches a concentration sufficient to occupy O_R3, it will also turn off P_{rm}. A more efficient P_{rm} would then cause this limit to be reached faster but would not increase the amount of repressor accumulated. On the other hand, an operator mutation decreasing the affinity of O_R3 for repressor will raise the concentration of repressor attainable before P_{rm} is turned off. To prove directly that 434 T is an operator rather than a promoter mutation will require binding and protection studies with puri-

fied 434 repressor. In the meantime, the fact that 434 T reduces the homology of O_R3 with the other two sites indicates that, whatever its effect on P_{rm}, it most likely reduces the affinity of O_R3 for repressor.

The transcription studies show two preferred sites of transcription initiation, one directed to the right and one to the left. Although I have no direct proof that these sites correspond to P_R and P_{rm}, their location with respect to the operator makes this identification very likely. Under the conditions used, both sites are remarkably inefficient in initiating transcription *in vitro*. When the *Eco* RI 1280 fragment, containing the *oop* promoter in addition to P_R and P_{rm}, was used as a template, the product was predominantly *oop* RNA. The cause of this inefficiency might lie in the sequence of the so called Pribnow box (16), with which RNA polymerase is known to interact near the transcription startpoint.

In the Pribnow box region the 434 P_R promoter has the sequence TACAAGA which constitutes a good Pribnow sequence except that it has a G in the position where nearly all other promoter sequences have a T. In the P_{rm} promoter, the sequence resembling most a Pribnow box is AATAACA.

As in λ, the P_R promoter overlaps with O_R1. In 434, however, the P_R startpoint lies within O_R1. Allowing at least 35 bp upstream of the startpoint for the polymerase binding site, the 434 P_R would overlap fully with both O_R1 and O_R2. The position of the leftward transcription startpoint, far to the left of O_R3, also indicates a general leftward shift of both P_R and P_{rm} relative to λ. The distance between the two transcription start points remains the same: 81 bp in λ, 82 bp in 434. According to one model proposed for the transcriptional switch between P_R and P_{rm}, the distance between these two promoters is just shorter than that required for independent binding of polymerase to both promoters. Repressor binding at either O_R1 or O_R2 could then activate P_{rm} by blocking the binding of polymerase to the competing P_R promoter. According to the alternative model, repressor activates P_{rm} by binding adjacent to it and facilitating the binding of the polymerase. In this case, the great distance of O_R1 from P_{rm} suggests that in 434 it might be O_R2, not O_R1, that acts as the primary binding site from which repressor activates P_{rm}.

It is interesting to note that the presumed P_{rm}mRNA, while it does not start with AUG as in λ, it contains an AUG 5 bp from the 5' end. This is the only AUG present in the first 42 nucleotides and it defines a triplet frame which contains no terminator codons within the region sequenced. This AUG is therefore very likely to be the initiator codon

for the 434 cI gene. As in λ, the cI mRNA initiated at P_{rm} would lack a good ribosomal binding site, while, when initiated from the P_{re} promoter farther to the right (17), it would have, near the initiator codon, a GAGGUG Shine-Dalgarno sequence complementary to the 3' end of the 16 S ribosomal RNA (18).

In the P_RmRNA, the first AUG cannot be the initiator codon since its reading frame contains terminator codons. Very probably the beginning of the 434 cro gene corresponds to the next AUG, eight nucleotides downstream. This AUG is preceded by a GGAGG which constitutes a good ribosomal binding site.

ACKNOWLEDGEMENTS

Part of this sequence was also determined by Elisabeth Schwarz whom I thank for communicating her results. I thank Kurt Ineichen and Agathe Stotz for excellent assistance. This work was supported in part by the Swiss National Science Foundation.

* Present address:

European Molecular Biology Laboratory
Postfach 10.2209
D - 6900 Heidelberg
Federal Republic of Germany

REFERENCES

1 Ptashne, M., Backman, K., Humayun, M., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R. (1976) Science 194, 156-161
2 Pirrotta, V. (1976) Current Topics in Microbiology and Immunology 74, 21-54
3 Walz, A. and Pirrotta, V. (1975) Nature 254, 118-127
4 Walz, A., Pirrotta, V. and Ineichen, K. (1976) Nature 262, 665-669
5 Meyer, B., Kleid, D. and Ptashne, M. (1975) Proc. Nat. Acad. Sci. (USA) 72, 4785-4789
6 Kaiser, A.D. and Jacob, F. (1957) Virology 4, 509-521
7 Pastrana, R. and Brammar, W. (1976) Molec. gen. Genet. 146, 191-198
8 Pirrotta, V. and Ptashne, M. (1969) Nature 222, 541-544
9 Pirrotta, V., Ptashne, M., Chadwick, P. and Steinberg, R. (1971) Procedures in Nucleic Acids Research (ed. G. Cantoni and D. Davies) vol. 2 pp. 703-715, Harper and Row.
10 Maxam, A. and Gilbert, W. (1977) Proc. Nat Acad. Sci. (USA) 74, 560-564
11 Pirrotta, V. (1975) Nature 254, 114-117

12 Barrell, B.G. (1971) Procedures in Nucleic Acids Research (ed. G. Cantoni and D. Davies) vol. 2 pp. 751-779
13 Jay, E., Bambara, R. Padmanabhan and Wu, R. (1974) Nucl. Acids Res. 1, 331-353
14 Murray, K. and Murray, N. (1975) J. Mol. Biol. 98, 551-564
15 Scherer, G. (1978) Nucl. Acids Res. 5, 3141-3156
16 Pribnow, D. (1975) Proc. Nat Acad. Sci. (USA) 72, 784-788
17 Reichardt, L. and Kaiser, A.D. (1971) Proc. Nat. Acad. Sci. (USA) 68, 2185-2189
18 Shine, J. and Dalgarno, L. (1974) Proc. Nat. Acad. Sci. (USA) 71, 1342-1346