Characterization of in vitro transcription initiation and termination sites in Col E1 DNA

Roger K. Patient¹

University of Wisconsin, Department of Biochemistry, College of Agricultural and Life Sciences,
420 Henry Mall, Madison, WI 53706, USA

Received 25 April 1979

ABSTRACT

Overlapping restriction fragments from the region between the single Eco RI site and the origin of replication of the plasmid, Col E1, have been utilised as templates in an in vitro transcription assay using *E. coli* RNA polymerase. Transcription towards the single Eco RI site is initiated at a point 415 bp to the origin side of that site. In vivo, transcription starting at this point probably produces the mRNA for the colicin immunity protein. Transcription away from the Eco RI site is initiated at a point 140 bp to the origin side of that site and terminated 30 bp further on. This terminator is probably the point at which transcription of the colicin gene is terminated in vivo. DNA sequence analysis in both these regions demonstrated several similarities to other prokaryotic regulatory regions. 50% homology between the putative immunity promoter and other prokaryotic promoters is apparent, so are similarities in AT-content. Upstream of the ATG start codon the sequence PuPuTTTPuPu and a termination codon (TAA) appear; both are typical of prokaryotic ribosome binding sites. The colicin terminator demonstrated similarities to other rho-independent prokaryotic terminators: a GC-rich region with termination in an adjacent AT-rich region containing T clusters on the non-coding strand. The possible role of initiation upstream from the colicin terminator is discussed.

INTRODUCTION

The plasmid, Col E1, has been used extensively as a cloning vehicle (2-5) and as a model for DNA replication (6-8). The genetics of this extra-chromosomal element have been examined with the aid of deletion and Tn mutants (9-11) and several structural genes have been localized, including the colicin, immunity and mobility genes (see Fig. 1). However, the start points and directions of transcription of these genes are not clear. For instance, the expression of genes inserted in the Eco RI site of Col E1 and the location of the strongest RNA polymerase binding site in the region of the colicin gene indicated transcription away from the origin of replication towards the single Eco RI site (i.e., anticlockwise in Fig. 1) (5). However, the sizes of the proteins produced in mini-cells by the parental plasmid and by a derivative containing an insertion at the Eco RI site suggested that

Figure 1. The genetic map of Col E1 (10) showing the colicin, immunity, and mobility genes as well as the origin of replication (Ori). The colicin and immunity genes exhibit overlap. The heavy line indicates the portion of Col E1 retained in pVH51 (12). The single Eco RI site is included as a reference point.

transcription was in the opposite direction (4).

We have previously reported (13) that the two most AT-rich fragments in pVH51, a Col E1 derivative containing about half the Col E1 genome (12, see Fig. 1), mapped in the colicin and immunity regions and bound *E. coli* RNA polymerase very tightly. This paper reports *in vitro* transcription and DNA sequencing studies on these two regions. The DNA sequence is correlated with the transcription results and the genetic evidence.

MATERIALS AND METHODS

Preparation of (dT₉)_n markers. These were prepared essentially as described by Raae *et al.* (14) with the dT₉ first of all being end-labelled. [γ -³²P] ATP was prepared as described (15) and 100 μ l (1.4 mCi, $\sim$ 70% in ATP) was evaporated to dryness and dissolved in a solution (99 μ l) containing 0.15 mM dT₉ (gift of E. Selsing), 40 mM Tris HCl pH 7.5, 2 mM dithiothreitol, 10 mM MgCl₂ and 0.1 mM cold ATP. T4 polynucleotide kinase (P.L. Biochemicals, 1 μ l, 10 units) was added and the mixture was incubated at 37°C for 30 min. The solution was extracted with equilibrated phenol and then with ether, the

last remnants of ether being removed under vacuum. The remaining aqueous solution was made 50 mM in Tris HCl pH 7.5, 7 mM in dithiothreitol and 1 mM in ATP. dA_n (gift of J. Larson, 12 μ l, 0.14 OD₂₆₀) and DNA ligase (gift of E. Selsing, 2 μ l, 0.4 units (ATP-PP_i exchange assay, 16)) were added and the mixture was left at 4°C for 16 h. After the addition of dT_n (67.5 μ l, 0.76 OD₂₆₀), the mixture was heated at 100°C for 3 min and then quick-chilled in ice/water. The marker mixture was stored at -20°C and used directly.

Restriction fragment purification. Restriction fragments derived from the appropriate nonradioactive or *in vivo* [³H]-labelled plasmid DNAs were cut out of ethidium bromide-stained 5% polyacrylamide, 25% glycerol slab gels and eluted as described (15). The eluate was extracted with equilibrated phenol and then with ether, and the DNA was precipitated with ethanol, collected by centrifugation, washed with ethanol, recentrifuged and dissolved in water. The amount of each fragment was quantitated for the transcription assays by UV spectroscopy or scintillation counting (in the case of the [³H]-labelled DNAs). Two of the fragments used in the transcription assays (575 and 425^{III}) were isolated by RPC-5 chromatography of a Hae III digest of pRZ2 DNA (17). In fact, the superscript in 425^{III} denotes elution at a higher salt concentration than two other fragments of the same size (17).

In vitro transcription assays in the presence of [α -³²P] UTP. These were carried out essentially as described by Majors (18) and Maquat and Reznikoff (19). 0.1 μ g of the 192 bp fragment or the molar equivalent of a different sized fragment were preincubated for 3 min at 37°C in 14 μ l of a solution containing 30 mM Tris HCl (pH 7.9), 0.1 mM EDTA, 3 mM MgCl₂, 100 mM KCl and 0.1 mM dithiothreitol. RNA polymerase (gift of L. Maquat, 1 μ g in 1 μ l) was added at a 2.7:1 molar ratio of enzyme to DNA and the incubation was continued for 10 min when heparin (Sigma, 1 μ l, 2 μ g) was added. One minute later, the reaction was made 200 μ M in ATP and GTP, 10 μ M in CTP and 5-10 μ M in [α -³²P] UTP (Amersham Searle or New England Nuclear, 350 Ci/mmol). 10 min later ice-cold transcription buffer (as above, 100 μ l) containing tRNA (100 μ g) was added to terminate synthesis. The reaction mixture was extracted with equilibrated phenol and made 0.3 M with respect to sodium acetate. After precipitation with ethanol, the RNA-containing pellet was washed with 90% aqueous ethanol, dried *in vacuo* and redissolved in 0.1 x Tris-borate-EDTA electrophoresis buffer (20) (16 μ l). A solution (4 μ l) containing 50% glycerol, 0.25% bromophenol blue and 0.25% xylene cyanol was

added to each sample before electrophoresis on 10% polyacrylamide-7 M urea (Research Plus Laboratories, Inc.) slab gels (21). Gels were covered with Saranwrap and exposed, without drying, to X-ray film (Kodak) at -70°C .

In vitro transcription of the 192 bp fragment in the presence of $[\gamma\text{-}^{32}\text{P}]$ ATP or $[\gamma\text{-}^{32}\text{P}]$ GTP. 0.85 μg of the 192 bp fragment was preincubated for 3 min at 37°C in 119 μl of transcription buffer (as above). RNA polymerase (8.5 μl , 8.5 μg) was added and after incubation at 37°C for 10 min, heparin (8.5 μl , 17 μg) was added. 1 min later, the reaction was made 240 μM in UTP, CTP and either ATP or GTP, and also 240 μM in either $[\gamma\text{-}^{32}\text{P}]$ GTP or $[\gamma\text{-}^{32}\text{P}]$ ATP (both were gifts of R. Burgess, 10 Ci/mmol). Control reactions were: synthesis from the 192 bp fragment in the presence of $[\alpha\text{-}^{32}\text{P}]$ UTP as described above; and synthesis for 30 min at 37°C from λplac5 DNA (4 μg) with RNA polymerase (4 μg) in transcription buffer (40 μl) which was 240 μM in UTP, CTP and GTP and 60 μM in $[\gamma\text{-}^{32}\text{P}]$ ATP (10 Ci/mmol). Reaction mixtures were all treated and fractionated as described above.

Partial T1 ribonuclease digestion of the transcripts from the 192 bp fragment

The two major transcripts from the 192 bp fragment, synthesized in the presence of $[\gamma\text{-}^{32}\text{P}]$ GTP, along with the 6S transcript from λplac5 , synthesized in the presence of $[\gamma\text{-}^{32}\text{P}]$ ATP, were cut out from gels and eluted in the manner described above for DNA. The final dried pellets were re-dissolved in 60 μl of a solution containing 20 mM sodium citrate pH 5, 1 mM EDTA, 7 M urea, 0.025% xylene cyanol, 0.025% bromophenol blue and tRNA (75 μg). Each solution was then divided into three 20 μl aliquots. Ribonuclease T1 (Calbiochem (gift of E. Calva and R. Burgess), 1 μl , 0.5 unit) was then added to one of the aliquots, 2 μl of the resulting mix was added to one of the other aliquots and then 2 μl of this resulting mix was added to the last. All three solutions in each case were incubated at 50°C for 15 min and then fractionated directly on a 20% polyacrylamide "sequencing" gel (15), modified as described below. Autoradiography detected the labelled bands. The gel was calibrated using the sizes of the products of partial T1 ribonuclease digestion of λ 6S RNA (22). Interestingly, the $(\text{dTg})_n$ products migrated about 1 cm further than predicted from this calibration curve.

DNA sequencing. This was carried out essentially as described by Maxam and Gilbert (15). The appropriate plasmid DNA (50 μg) or the Hae III fragment 425^{III} (17) (5 μg) were digested with suitable restriction enzymes and the

fragments to be sequenced were isolated from 5% polyacrylamide gels. The 5'-terminal phosphate groups were removed using bacterial alkaline phosphatase (Sigma). [γ - ^{32}P] ATP was prepared as described and ~ 1.5 mCi of the final reaction mixture was used, with T4 polynucleotide kinase (Biolabs or P.L. Biochemicals) to end-label each fragment. After secondary restriction cleavage, the fragments labelled at one end were isolated from a 5% polyacrylamide slab gel. The four reactions performed on each fragment were A > G, G, C and C + T. Reaction times were generally 25, 10, 25 and 20 minutes respectively. The gels were either made up according to Maxam and Gilbert (15) or the amount of bis-acrylamide was increased (ratio of acrylamide to bis 19:1). This modification seemed to increase the spacing of the larger cleavage products. Both types of gels were generally electrophoresed at between 350 and 450V in 1 x Tris-borate-EDTA electrophoresis buffer (20). Gels were covered with Saranwrap and exposed without drying to Kodak No-Screen X-ray film at -70°C .

RESULTS

All the Col E1 derivatives used in this study retain the region between the Eco R1 site and the origin of replication, i.e., the colicin immunity gene and one end of the colicin gene.

In a previous communication (13), we demonstrated RNA polymerase binding to the two Hae III fragments which mapped between the Eco R1 site and the origin of replication of Col E1. This observation was interesting because these two fragments mapped at one end of the colicin gene (365) and in the region of the colicin immunity gene (425^{III}), see Fig. 2. This RNA polymerase binding data was reinforced by preliminary experiments which showed that the Eco R1-Taq I 192 bp fragment and 425^{III} both acted as templates for in vitro transcription. To determine the direction and start point of this transcription in each case, a number of overlapping restriction fragments from these regions were subjected to in vitro transcription analysis and DNA sequencing.

Transcription in the immunity region

Fig. 2 (bottom) shows four overlapping fragments in the region of 425^{III} . They are 425^{III} itself (a Hae III fragment), the 333 bp Hae III - Hpa II fragment, the 282 bp Hae III - Hha I fragment and the 455 bp Taq I - Hha I fragment. Fig. 3 shows the results of an in vitro transcription assay utilizing these fragments as templates. The major transcript synthesized on

Figure 2. The restriction map and genetic loci of a region of pRZ2 DNA depicting the fragments subjected to *in vitro* transcription analysis and the transcripts produced from them. The plasmid pRZ2 (23) is a derivative of pVH51 in which a 789 bp Hind II fragment containing the lac control elements was inserted into the single Hind II site of the vector (23). pVH51 was generated by spontaneous deletion of a Col E1 derivative containing an Eco RI fragment of ϕ 80pt190 inserted in the Eco RI site of the vector (12). The top of the figure contains the restriction map for six enzymes (23); the sizes of the fragments are in bp. Above this, the origins of the various parts of this region of DNA are delineated and the colicin and immunity gene locations are indicated. The bottom of the figure contains seven double-stranded restriction fragments whose sizes are given in base pairs. In addition, the direction and start points of transcription of these fragments are indicated along with the sizes of the major transcripts in nucleotides.

425^{III} was 50 nucleotides in length. Since an RNA of the same size was the major product from 282 and 333, and these share the same left end as 425^{III} (Fig. 2), this transcript probably results from transcription initiating 50 nucleotides from the left end. This conclusion is supported by transcription on the 455 bp template which resulted in the production of a 220 nucleotide major transcript and no 50 nucleotide product. The 455 bp template extends

170 bp beyond the left end of 425^{III} permitting the addition of 170 nucleotides to the 50 nucleotide RNA before the polymerase reached the end of the fragment. Thus, we can conclude that a start point for transcription is located 50 bp inside the left end of 425^{III} and that transcription proceeds leftwards at least as far as the Taq I site at the end of 455 (see Fig. 2).

In addition to the major transcripts (50 and 220 nucleotides) mentioned above, two minor products (160 and 140 nucleotides) were detected (Fig. 3). The 160 nucleotide band in the 455 channel may be the result of transcription in the same direction as for the 220 nucleotide RNA but shortened by stuttering initiation or premature termination (25). The 140 nucleotide transcript produced from 425^{III} could be the product of leftward transcription starting in the region of the Hpa II recognition site, and terminating within 425^{III}. This would explain why this RNA was not synthesized by 282 and 333. Alternatively, the 140 nucleotide RNA could be the result of transcription terminating at the right end of 425^{III}. If it were, 333 should give rise to a 45 (± 4) nucleotide RNA. This cannot be completely ruled out from Fig. 3, since it could be comigrating with the 50 nucleotide major transcript.

DNA sequence in the immunity region

Scherer *et al.* (26) have described a region of homology in prokaryotic promoters stretching from -45 to +17 (where +1 represents the 5'-terminus of the message). The sequence of such a region surrounding the major (50 nucleotide transcript) start point in 425^{III} was determined. The autoradiogram of one of the sequencing gels is shown in Fig. 4, loads a, b and c. The sequence of 120 bp rightwards from an Alu I site 37 bp inside the left end of 425^{III} is presented in Fig. 5. Since the sequencing reactions were performed only on a fragment labelled at the left end in Fig. 5, gel separation limitations mean that more confidence can be placed on the first 80 residues than the last 40.

The start point should be 13 (± 2) bp from the Alu I cut (i.e., 15 (± 2) bp from the left end of the sequence in Fig. 5), based on the measured size of the major transcript produced by 425^{III} (50 bp) minus 37 bp and the estimated error. This would make one of the nucleotides from the sequence GGUUU the starting nucleotide, the more frequent occurrence of A or G in this position (26) would favour one of the Gs. Furthermore, if the sequence is lined up with the prototypic sequence of Scherer *et al.* taking each of these 5 positions as +1 in turn, then the G marked in Fig. 5 gives the greatest

Figure 3. The size of transcripts synthesized on the DNA fragments mapping on the right of the restriction map in Fig. 2. The fragments are identified by their size in bp. Lanes 1 and 7 contain the same preparation of the $(dT_9)_n$ markers as appear in lane 6 of Fig. 6. Lanes 3 and 4 contain the transcription products from two different preparations of 425III: one was isolated from a polyacrylamide slab gel, the other from RPC-5 column chromatography (17). Transcription in the presence of $[\alpha\text{-}^{32}\text{P}]$ UTP was carried out and the products were fractionated as described in Methods. The numbers on the right indicate the sizes of the marker bands in nucleotides. The numbers on the left indicate the sizes in nucleotides of the major transcripts as deduced from a semi-log plot of the sizes of the marker bands versus their positions of migration. The marker bands used were $(dT_9)_n$ and a Hae III digest of $\phi\text{x}174(+)$. The validity of using oligodeoxynucleotide markers to size oligoribonucleotide products was demonstrated by Maniatis et al. (24). In addition, the electrophoretic position of tRNA^{Arg} (76 nucleotides) agreed, within 4%, with the position predicted from the standard curve. The positions of migration of xylene cyanol (XC) and bromophenol blue (BPB) are marked.

degree of homology. This is true whether one considers only the more strongly preserved residues (52% homology) or both the strong and less strong ones (46% homology).

One of the two possible locations of the template for the minor (140 nucleotide) transcript could just encroach on the sequenced region in Fig. 5.

Figure 4. Sample DNA sequencing data for the immunity (loads a, b and c) and colicin (loads d and e) regions of Col E1. The actual sequence is indicated on the left and right for each region. The reactions loaded in each slot are noted above and reactions loaded at the same time are grouped by a bracket at the bottom and labelled with a lower case letter. Both gels were the modified type (see Methods). The reactions shown here for the immunity region were performed on a fragment labelled at the Alu I site 37 bp to the right of the leftward Hae III site of 425^{III} and extending to the right in Fig. 2. The reactions shown here for the colicin region were performed on a fragment labelled at the Eco RI site and extending to the right in Fig. 2. The experimental procedure is described in Methods.

Transcription starting 45 bp to the left of the Hpa II site (the maximum allowed for cutting with Hpa II to affect transcription (26) would place termination at position -96±4 in Fig. 5 and indeed the terminal residues show some similarity to the prototypic transcription terminator (27). The other possible source of the 140 nucleotide transcript, transcription terminating at the right end of 425^{III}, would place the limit of the promoter (residue -45) outside the sequence shown in Fig. 5.

Transcription in the colicin region

The region between 425^{III} and the Eco RI site contains one end of the

Figure 5. DNA sequence from the immunity region in Col E1. This sequence was deduced from the data in Fig. 4 and a repeat of that using four loads instead of three. The numbers denote residues relative to the start point for transcription at +1 (marked with an arrow).

colicin gene (9,10). However, as discussed in the Introduction, it is not clear whether it is the beginning or the end of the colicin gene (4,5). Fig. 2 (middle) shows three fragments which overlap this region: the 192 bp Eco RI - Taq I fragment, the 365 bp Eco RI - Hae III fragment and the 575 bp Hae III fragment. Fig. 6 shows the results of an *in vitro* transcription assay utilizing these fragments as templates.

The two major transcripts synthesized on 192 were 31 and 37 nucleotides in length. The presence of these same two major RNAs among the products of synthesis on the 365 and 575 templates implied that they were not the result of transcription terminating at either end of the 192 bp fragment, but rather the result of internal starting and stopping.

The minor (228, 195 and 100 nucleotide) RNAs transcribed from 575 and 365 are the same for both templates and therefore could not result from transcription terminating at the left ends. Their probable identity as read-through products starting in the same region as the major (31 and 37 nucleotide) transcripts is discussed below in the light of DNA and RNA sequence information.

DNA and partial RNA sequence in the colicin region

DNA sequencing coupled with partial T1 ribonuclease digestion of the end-labelled transcripts was performed to determine the major transcription initiation and termination points in this region (i.e., those for the 31 and 37 nucleotide RNAs).

Figure 6. The size of transcripts synthesized on the DNA fragments mapping on the left of the restriction map in Fig. 2. The fragments are identified by their size in base pairs. Lanes 1 and 6 contain two different preparations of the $(dT_9)_n$ markers. Lanes 4 and 5 contain the transcription products from two different preparations of the 192 bp fragment. Transcription in the presence of $[\alpha\text{-}^{32}\text{P}]$ UTP was carried out and the products were fractionated as described in Methods. The numbers on the left indicate the sizes of the marker bands in nucleotides; the positions of migration of xylene cyanol (XC) and bromophenol blue (BPB) are also marked. The numbers on the right indicate the sizes in nucleotides of the major transcripts deduced as described in the legend to Fig. 3.

The DNA sequence between the Eco RI site and the Taq I site 192 bp to the right of it (Fig. 2) was determined. A sample gel is shown in Fig. 4 loads d and e, and the sequence is presented in Fig. 7. Since this region was sequenced from both ends with overlap of approximately 30 bp in the middle, greater confidence in the sequence of the DNA distal to the two labelled ends is possible.

Obviously, the positions of initiation and termination of transcription could not be deduced from the DNA sequence alone, since these are both

internal. Therefore, *in vitro* transcription was carried out on the 192 bp fragment in the presence of [γ - 32 P] ATP and [γ - 32 P] GTP separately. The two major transcripts (37 and 31 nucleotides) were labelled in the presence of [γ - 32 P] GTP but not [γ - 32 P] ATP (data not shown). This result suggested that the starting nucleotide was G in each case.

The two major products of the *in vitro* transcription on 192 in the presence of [γ - 32 P] GTP were isolated and partially digested with ribonuclease T1. Electrophoretic separation of the products (data not shown) revealed a run of G residues at positions 22, 23, 24, 25 and 26 for the 37 nucleotide transcript and 18, 19, 20, 21 and 22 for the 31 nucleotide product.

This distinctive run of five Gs occurs only once in the DNA sequence shown in Fig. 7, at position 22-26, and demonstrates that the direction of transcription must be from left to right. The different distances of these five Gs from the labelled ends of the RNAs suggests that synthesis occurs from two separate start points (+1 and +5), 4 residues apart. The absence

Figure 7. DNA sequence from the colicin region in Col E1. This sequence was deduced from the data in Fig. 4, two repeats of that using three and four loads instead of two and two sets of reactions (two and three loads) performed on a fragment labelled at the *Taq* I site to the right of the *Eco* R1 site extending to the left in Fig. 2. The numbers at the tick marks denote residues relative to the start point for transcription at +1. Two transcription initiation points (at +1 and +5) are marked with horizontal arrows labelled according to the size of the transcript produced (37 and 31 nucleotides respectively). The two transcription termination points are marked with vertical arrows distinguished by their corresponding transcript size. Triplets marked with a line represent termination codons in all three phases (+, o and x) which would be contained within RNAs transcribed in both directions from this region.

of a G residue in position +1 (Fig. 7) is rather disconcerting since the 37 nucleotide transcript was shown to initiate with [γ - 32 P] GTP. The DNA sequencing data for this region was scrutinized very closely for a possible C $\rightarrow$ T confusion (the transcribed strand was sequenced) but none was apparent. The nearest G residue in the sequence is at position -4, which is outside the experimental error for reading the gel of the partial T1 digested RNAs. A possible failure of the transcription assay using [γ - 32 P] ATP is ruled out by its success with the λ plac5 control. Thus, I have no explanation for this discrepancy.

Transcription initiating at the two start points, +1 and +5 (Fig. 7), places the transcription termination points at positions +35 for the 31 nucleotide transcript and +37 for the 37 nucleotide RNA. Transcription initiating at the same two start points and terminating at the common right end of 365 and 575 would result in the 228 nucleotide product we see in Fig. 6. The 195 nucleotide product could be the result of premature termination (25) of such a read-through transcription, as might the 100 nucleotide RNA, although there is insufficient data to confirm this.

DISCUSSION

The transcription initiation site shown in Fig. 5 probably represents the beginning of the colicin immunity gene. Leftward transcription initiating at that point (415 bp to the origin side of the Eco RI site) extends beyond the Taq I site on the right of the sequence in Fig. 7. An ATG codon is found at positions +10 to +12 (Fig. 5) and termination codons in all three phases are identified in Fig. 7. Depending on the correct phase, the immunity protein would be between about 80 and 105 amino acid residues in length. Such small molecules are typical of immunity proteins (28). The position of these translation termination codons places the end of the immunity gene 90 to 170 bp to the origin side of the Eco RI site. This location of the immunity gene (90-170 to 415 bp to the origin side of the Eco RI site, 90-170 to 460 bp including promoter requirements (26)) agrees well with the genetic evidence of Heffron *et al.* (11) which shows that the gene extends from 100 to 500 bp to the origin side of the Eco RI site.

The DNA sequence of the putative immunity promoter, as mentioned in Results, is similar to the DNA sequences of other prokaryotic promoters (26). In addition to specific sequence homology, the region from -45 to +16 is overall AT-rich (69%), which agrees well with the 60-70% found for other promoters (26). Furthermore, the AT-rich flanking regions identified by

Scherer *et al.* (26) are even more AT-rich here: -45 to -56, 92% and -21 to +13, 74%. The AT tails, in positions -32 to -25 and -13 to -7, mentioned by Scherer *et al.* as pointing in the direction of transcription, are also present here.

As mentioned above, the ATG codon occurs at +10 to +12. Some features common to untranslated regions are apparent in the sequence proximal to this initiation codon. The sequence PuPuTTTPuPu and a termination codon (TAA) are both found within a few bases of the initiation codon (29). However, no great degree of secondary structure of the mRNA (30,31) is apparent in this area. If transcription initiated three residues upstream from the position marked as +1 in Fig. 5 (i.e., position -3) then complementarity of the resulting mRNA to the 3' terminus of 16S rRNA (32,33) would extend over 5 residues (GAGGT) instead of the two (GT) apparent in Fig. 5. However, such a location for the transcription initiation point would drastically lower the degree of homology of the consequent promoter sequence when aligned with the prototypic sequence (26) (strongly preserved residues, 52% to 29%; strong and less strong, 46% to 26%). Hence, I have opted for the initiation point consistent with the *in vitro* transcription evidence and promoter homology. It is possible, however, that both homologies could be retained if the non-preserved regions of the prototypic sequence possess flexibility in overall size, as well as the types of residues they contain, thus absorbing the three residue difference mentioned above. Indeed, the method by which the prototypic sequence was derived (26) did permit diminished contributions from sequences which were not completely aligned.

Of the two possible locations in 425^{III} for synthesis of the minor 140 nucleotide transcript (mentioned in Results), leftward transcription would result in the production of about a 45 amino acid protein and rightward transcription would produce a larger one. However, both these regions fall outside the expected genetic locus for immunity (11).

Backman *et al.* (34) have discovered a 100 nucleotide transcript produced from the region of Col E1 spanned by 425^{III}. This RNA was postulated to be the primer for DNA synthesis. Its absence among the products of our transcription assay could be due to the *Hae* III cut at position -30 relative to the reported start point for this primer. This is consistent with the claims (26,35) that specific binding and initiation by RNA polymerase require the presence of specific DNA sequences more than 35 residues upstream from the start point of transcription. Another possible explanation is that this 100 nucleotide RNA is only synthesized on a super-

coiled template. Similar findings have been reported elsewhere (36).

The other region of Col E1 that has been investigated is one end of the colicin gene to the right of the Eco R1 site (Figs. 1 and 2). Most of the genetic evidence suggests that this is probably the end, rather than the beginning, of the gene. The evidence of Meagher *et al.* (4) states that insertion into the Eco R1 site of Col E1 results in the truncation of the protein produced (56000 to 52000). Since the bulk of the colicin gene has been localised (9,10) to the left of the Eco R1 site (Fig. 1), transcription must be in a clockwise direction (left to right towards the origin of replication). Selker *et al.* (3) showed that functional expression of the trp A gene inserted at the Eco R1 site of Col E1 was mitomycin C inducible. Since colicin production is known to be induced by mitomycin C (37), transcription of the trp A insert was thought to be under the control of the colicin promoter. Subsequently the orientation of the trp A insert was determined (28) and this, in agreement with the above protein evidence, indicated that the colicin promoter was situated left of the Eco R1 site, with transcription of the colicin gene proceeding towards the origin of replication. The same direction of transcription of the colicin gene has been suggested by experiments with pCR1 (3). This plasmid contains the kanamycin gene inserted in the Eco R1 site of Col E1, with the subsequent deletion of the Eco R1 site between the kanamycin insert and the larger part of the colicin gene (38). DNA fragments inserted in the remaining Eco R1 site are no longer expressed *in vivo* probably because they are now too far from the colicin promoter(3).

The only evidence against a clockwise direction of transcription of the colicin gene has been presented by Collins and coworkers (5). They found expression of their inserts in a sense consistent with *anticlockwise* transcription through the Eco R1 site of Col E1. This could be explained by read-through from the immunity promoter or by the presence of a promoter within the insert. They also presented RNA polymerase binding data which showed that binding to the right end of the colicin gene was stronger than to the left. However, they did concede that strong polymerase binding is not a sufficient demonstration of a promoter.

The results presented in this paper are consistent with a clockwise direction of transcription of the colicin gene. The transcription termination site(s) indicated in Fig. 7 probably effect termination of transcription of the colicin gene *in vivo*. This location falls within the rightward boundary of the colicin gene (220-275 bp to the right of the Eco R1 site) as deduced

by Dougan *et al.* (10). Translation termination codons in all three phases are indicated in Fig. 7. If the last of the three phases was the correct one, insertion into the Eco RI site of Col E1 could result in the loss of protein of molecular weight around 2600. This agrees quite well (within experimental error for the measurement of protein sizes) with the protein loss observed by Meagher *et al.* (4) (56000 truncated to 52000). The N-terminal sequence of the colicin E1 protein was recently determined, by Watson and Sherratt (39), to be Met Val Leu Ile Val Ala Leu. If transcription of the colicin gene was anticlockwise then some part of the DNA sequence presented in Fig. 7 should code for the above peptide. However, reading from right to left in Fig. 7, only one ATG start codon is present (at positions -1 to -3) and the next six triplets do not code for the above amino acids. Thus, the N-terminus of colicin E1 is not coded for by this region of DNA. Watson and Sherratt also determined the C-terminal residue of colicin E1 to be glycine. However, none of the potential colicin termination codons indicated in Fig. 7 would terminate a protein ending in glycine. The codon terminating in the last of the three phases (i.e., the one marked with a x and the one most favoured by the truncation of protein sizes as discussed above) would result in a protein with glycine as the penultimate residue. The reasons for this discrepancy are not understood.

Several characteristics of prokaryotic transcription terminators are becoming apparent (27). These include a GC-rich region of dyad symmetry with termination in an adjacent AT-rich region containing T clusters on the non-coding strand. Termination as indicated in Fig. 7 complies with all these conditions except the dyad symmetry. Indeed, GC and AT blocks abound in this region, as do T clusters in the non-coding strand. This may ensure fail-safe termination. Sequence heterogeneity at the 3'-OH terminus of mRNA (as indicated in Fig. 7) has also been observed for the trp leader (40) and λ cro (41) RNAs.

The presence of apparent "promoters" up-stream from the colicin terminator (Fig. 7) is somewhat surprising, although a possible function for these may be to slow down the RNA polymerase for termination. A similar situation has been observed in bacteriophages fd (42), ϕ x174 (43) and possibly λ with the cro terminator t_{R1} (27). Some homology with the prototypic promoter sequence (26) is apparent for the two start points indicated in Fig. 7 (38% homology with the strongly preserved residues and 33% with the strong and less strong ones for both "promoters"). Both "promoters" are 66% AT overall and 74% AT between positions -21 and +13, thus conforming to the prototype on

both counts. However, 50% and 58% AT between positions -45 and -56 for the 31 and 37 transcript "promoters" respectively falls below the prototypic value of 68%. So agreement is reasonable but not as good as for the immunity promoter. Further agreement with "promoter" requirements was indicated by cutting with the restriction enzyme, Ava II, in the region -24 to -28 (Fig.7) which seemed to prevent initiation from either of the two start points (data not shown). An alternative but less likely explanation of the "promoters" a short distance upstream from the terminator is, of course, that the DNA in between codes for a leader RNA and that the terminator is in fact an attenuator. Such a high proportion of termination at an attenuator in vitro has precedent in the trp operon (44). These "promoters" and the "attenuator" fall within the genetic locus for immunity and therefore they could function in transcription of the immunity gene. This seems unlikely since that would leave no apparent function for the strong initiation of transcription observed at the other end of the immunity gene.

In conclusion, transcription of the colicin immunity gene probably initiates at a point 415 bp to the right of the Eco RI site and terminates in the region between 90 and 170 bp right of the Eco RI site. Transcription of the colicin gene probably initiates well to the left of the Eco RI site and terminates 170 bp to the right of the Eco RI site. Thus, as suggested by the genetic evidence (9,10), the colicin and immunity genes probably overlap. This could explain the expected and suggested coordinate control of the two (45).

This work was supported by grants from the National Institutes of Health (CA 20279) and the National Science Foundation (PCM 77-15033) to R.D.Wells

REFERENCES

1. Present address and the one to which correspondence should be sent: Imperial Cancer Research Fund, Mill Hill Laboratories, Burtonhole Lane, London, NW7 1AD, U.K.
2. Hersfield, V., Boyer, H.W., Yanofsky, C., Lovett, M.A. and Helinski, D.R. (1974) *Proc. Natl. Acad. Sci. U.S.A.* 71, 3455-3459.
3. Selker, E., Brown, K. and Yanofsky, C. (1977) *J. Bact.* 129, 388-394.
4. Meagher, R.B., Tait, R.C., Betlach, M. and Boyer, H.W. (1977) *Cell* 10, 521-536.
5. Collins, J., Johnsen, M., Jorgensen, P., Valentin-Hansen, P., Karlstrom, H.O., Gautier, F., Lindenmaier, W., Mayer, H. and Sjoberg, B.M. (1978) in *Microbiology*, Schelessinger, D., Ed., pp. 150-153, A.S.M.
6. Helinski, D.R. (1976) *Fed. Proc.* 35, 2026-2030.
7. Tomizawa, J., Ohmori, H. and Bird, R.E. (1977) *Proc. Natl. Acad. Sci., U.S.A.* 74, 1865-1869.

8. Gellert, M., Mizuuchi, K., O'Dea, M.H., Itoh, T. and Tomizawa, J. (1977) *Proc. Natl. Acad. Sci., U.S.A.* 74, 4772-4776.
9. Inselburg, J. (1977) *J. Bact.* 132, 332-340.
10. Dougan, G., Saul, M., Warren, G. and Sherratt, D. (1978) *Molec. Gen. Genet.* 158, 325-327.
11. Heffron, F., Bedinger, P., Champoux, J.J. and Falkow, S. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 702-706.
12. Hershfield, V., Boyer, H.W., Chow, L. and Helinski, D.R. (1976) *J. Bact.* 126, 447-453.
13. Patient, R.K., Hardies, S.C., Larson, J.E., Inman, R.B., Maquat, L.E. and Wells, R.D. (1979) *J. Biol. Chem.*, in press.
14. Raaij, A.J., Lillehaug, J.R., Kleppe, R.K. and Kleppe, K. (1975) *Nucl. Acids Res.* 2, 423-429.
15. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 560-564.
16. Weiss, B., Jacquemin-Sablon, A., Live, T.R., Fareed, G.C. and Richardson, C.C. (1968) *J. Biol. Chem.* 243, 4543-4555.
17. Patient, R.K., Hardies, S.C. and Wells, R.D. (1979) *J. Biol. Chem.*, in press.
18. Majors, J. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 4394-4398.
19. Maquat, L.E. and Reznikoff, W.S. (1978) *J. Mol. Biol.* 125, 467-490.
20. Peacock, A.C. and Dingman, C.W. (1968) *Biochemistry* 7, 668-674.
21. Gilbert, W. and Maxam, A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* 70, 3581-3584.
22. Lebowitz, P., Weissman, S.M. and Radding, C.M. (1971) *J. Biol. Chem.* 246, 5120-5139.
23. Hardies, S.C., Patient, R.K., Klein, R.D., Ho, F., Reznikoff, W.S. and Wells, R.D. (1979) *J. Biol. Chem.*, in press.
24. Maniatis, T., Jeffrey, A. and van deSande, H. (1975) *Biochemistry* 14, 3787-3794.
25. Maizels, N.M. (1973) *Proc. Natl. Acad. Sci. U.S.A.* 70, 3585-3589.
26. Scherer, G.E.F., Walkinshaw, M.D. and Arnott, S. (1978) *Nucl. Acids Res.* 5, 3759-3773.
27. Adhya, S. and Gottesman, M. (1978) *Ann. Rev. Biochem.* 47, 967-996.
28. Helinski, D.R., personal communication.
29. Robertson, H.D., Barrell, B.G., Weith, H.L. and Donelson, J.E. (1973) *Nature New Biol.* 241, 38-40.
30. Steitz, J. (1973) *Proc. Natl. Acad. Sci. U.S.A.* 70, 2605-2609.
31. Cannistraro, V.J. and Kennell, D. (1979) *Nature* 277, 407-409.
32. Shine, J. and Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* 71, 1342-1346.
33. Steitz, J.A. and Jakes, K. (1975) *Proc. Natl. Acad. Sci. U.S.A.* 72, 4734-4738.
34. Backman, K., Betlach, M., Boyer, H.W. and Yanofsky, S., personal communication.
35. Gilbert, W. (1976) in *RNA polymerase*, Losick, R. and Chamberlin, M., Eds. p. 193, Cold Spring Harbor.
36. Levine, A. and Rupp, W.D. *EMBO Workshop, "Plasmids and other extra-chromosomal genetic elements"*, Max Planck Inst. for Mol. Gen., W.Berlin, April 1-5, 1978.
37. Herschman, H.R. and Helinski, D.R. (1967) *J. Bact.* 94, 691-699.
38. Armstrong, K. and Helinski, D.R. (1977) in *DNA insertion elements, plasmids and episomes*, Bukhari, A.I., Shapiro, J.A. and Adhya, S.L., Eds. p. 681, Cold Spring Harbor.
39. Watson, D.H. and Sherratt, D.J. (1979) *Nature* 278, 362-364.
40. Bertrand, K., Korn, L.J., Lee, F. and Yanofsky, C. (1977) *J. Mol. Biol.* 117, 227-247.

41. Rosenberg, M., Court, D., Shimatake, H., Brady, C. and Wulff, D.L. (1978) *Nature* 272, 414-423.
42. Schaller, H., Beck, E. and Takanami, M. (1978) in *Single Stranded DNA Phages*, Eds. Denhardt, D.T., Dressler, D., and Ray, D.S., Cold Spring Harbor.
43. Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison III, C.A., Slocombe, P.M. and Smith, M. (1977) *Nature* 265, 687-695.
44. Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C.L., Squires, C. and Yanofsky, C. (1975) *Science* 189, 22-26.
45. Inselburg, J. (1977) *J. Bact.* 129, 482-491.