Transcription Start Site and Induction Kinetics of the 
araC Regulatory Gene in Escherichia coli K-12

The in vivo transcription start site of the araC message was determined by S1 nuclease mapping of hybrids formed between labeled DNA, and RNA extracted from cells grown under a variety of physiological conditions, including the interval of transient derepression following arabinose addition. Under all conditions tested, transcription initiated from the same nucleotide position at -148.

The araC protein of Escherichia coli plays a central role in regulating expression of the metabolic operon, araBAD, and the two arabinose transport operons, araE and araFG (Englesberg et al., 1965; Schleif, 1969; Greenblatt & Schleif, 1971; Brown & Hogg, 1972; Lee et al., 1974; Kolodrubetz & Schleif, 1981). The synthesis of araC protein itself is also regulated. Transcription of the araC gene is reduced in the absence of cAMP receptor protein (CRP) or cAMP, and is derepressed about fivefold in the absence of active araC protein (Casadaban, 1976). Furthermore, the addition of arabinose to cells induces transcription of araC about fivefold for about 20 minutes, after which it falls back to its preinduction rate (Haggerty, 1977; Ogden et al., 1980; Hahn & Schleif, 1983). One of the objectives of the work presented here was to characterize more fully this unexpected and unexplained transient induction.

A second reason for this investigation is derived from questions about the actual initiation site or sites utilized for transcription of araC. Based on the DNA sequence of the araCBAD regulatory region (Smith & Schleif, 1978; Greenfield et al., 1978), in vitro transcription and electron microscopy (Hirsh & Schleif, 1977), the location of the transcription start site for the araC message was predicted to lie at position -167 as shown in Figure 1 (numbering from the +1 araB transcription start site) (Smith & Schleif, 1978). However, RNA sequencing of the E. coli B/r araC message synthesized either in vivo, apparently under steady-state conditions in the absence of arabinose, or in vitro (Wallace et al., 1981) showed that the araC transcription starts at position -148 instead of at the predicted -167 position. The -148 initiation site has four of the six highly conserved bases found in a consensus of RNA polymerase initiation sites (Hawley & McClure, 1983) plus an additional four of the ten less highly conserved bases. In contrast, the postulated -167 initiation site possesses four of the six highly conserved and six of the ten less-conserved bases (Fig. 1). Therefore, we examined the possibility that some or all araC transcription originates from the -167 site under different physiological conditions, such as during the transient derepression interval after arabinose addition to cells, or in cells grown in the presence of anti-inducer D-fucose. The start site in a Δcyα strain was also examined in order to
look for a second promoter adjacent to the first, one which might be more easily observed in the absence of cAMP receptor protein, analogous to the gal operon (Alba et al., 1981).

An end-labeled araC DNA restriction fragment, which includes the transcription start site of the pC promoter, was denatured and hybridized to unlabeled total cellular RNA isolated from cells grown in either the absence or the presence of arabinose for varying lengths of time, or grown in the presence of anti-inducer D-fucose. DNA was also hybridized to RNA isolated from a ΔcyA strain deleted of its adenyl cyclase gene. The RNA–DNA hybrid was digested with S1 nuclease to remove unpaired regions before denaturing and electrophoresis on DNA sequencing gels. Based on the mobility of the RNA-protected DNA fragment relative to DNA from a “G” sequencing reaction and correcting for the 1.5-base size difference between the S1-digested product versus the sequencing product (Hentschel et al., 1980), all detectable transcription of the araC gene, >95%, initiated from -148±3 base-pairs, under all cell growth conditions (Fig. 2).
The $S_1$ nuclease mapping technology permits quantitation of relative messenger levels. Therefore, the induction kinetics of $araC$ message could easily be measured following addition of arabinose. $araC$ message is induced within two minutes of the addition of arabinose, increases to four times its basal level by six minutes, and then declines back to its basal level (Figs 2 and 3). The basal level of $araC$ transcription is repressed by the addition of anti-inducer D-fucose. This is
consistent with the result of Haggerty (1977) who found that pC is 2.5-fold more strongly repressed in vivo when D-fucose is added to cells. The basal level of transcription is fourfold lower in the ΔCya strain than in the Cya+ strain and it only induces very slightly if at all in the first six minutes after arabinose addition.

The controls for these experiments were: (1) denatured araC DNA incubated with RNA isolated from a ΔAraC strain yielded no hybrid; (2) DNA hybridized to RNA isolated from a strain containing an araC hyperproducing plasmid (Schleif & Favreau, 1982), grown in the absence of arabinose, yielded a higher basal level of araC message; (3) araC message was not induced in cells grown in the absence of arabinose and harvested by pouring them into a bottle containing arabinose in addition to the ice, sodium azide and chloramphenicol, which shows that transcription is terminated immediately in the first step of the RNA preparation; (4) the RNA alone control showed no hybrid formation; and (5) the DNA alone control showed only the undenatured or renatured UKA probe.

In this letter we report determination of the in vivo transcription start point of the araC messenger in E. coli cells growing under a variety of physiological conditions. All detectable transcription initiated from the same nucleotide position at -148. This is the position originally determined under one physiological condition, by Wallace et al. (1981), and not the position predicted on the basis of the DNA sequence in the region (Smith & Schleif, 1978). Even though we find that the addition of arabinose transiently derepresses the pC promoter at least fourfold, all this additional transcription also initiates from the same site. Since the -148 site utilizes a promoter sequence with poorer homology to the -35 and -10 RNA polymerase consensus sequences than the unused -167 site,
it is clear that homology to these sequences alone does not provide a reliable indication of promoter activity.

This work was supported by United States Public Health Service research grant 18277, and training grant GM212 from the National Institutes of Health.

Department of Biochemistry  
Brandeis University  
Waltham, Mass. 02254, U.S.A.

Received 3 May 1983

REFERENCES


Edited by S. Brenner