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 S_1 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 Δ*cya* strain was also examined in order to

0022–2836/83/321049–05 $03.00/0

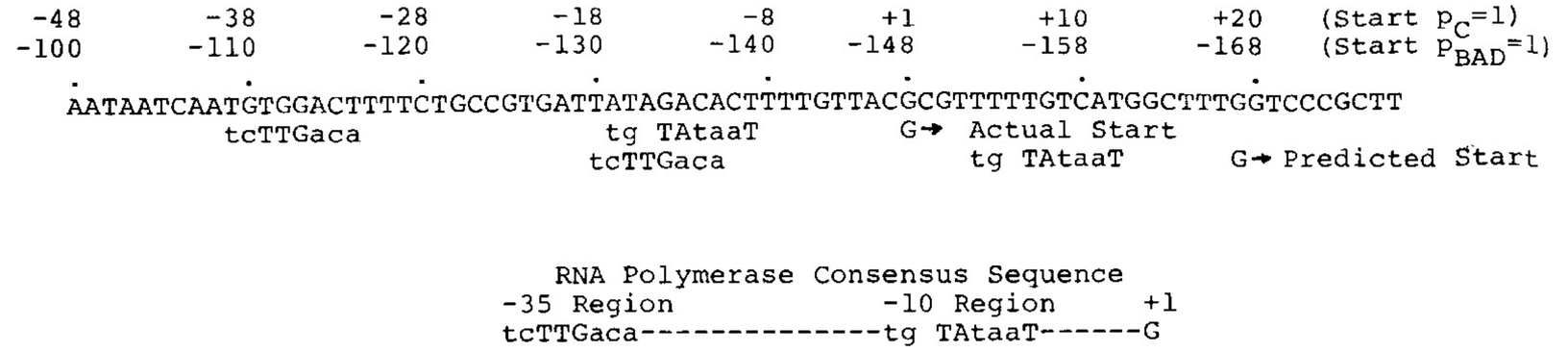

FIG. 1. Nucleotide sequence of the p_C promoter region. Two numberings of the region are shown, one in which +1 is the first nucleotide of the p_C transcript and the other in which +1 is the first nucleotide of the p_{BAD} transcript. The consensus RNA polymerase binding sequence aligned with the predicted and actual transcription start points of p_C is also shown. Hyphens have been omitted from the sequence for clarity.

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 (Aiba *et al.*, 1981).

An end-labeled *araC* DNA restriction fragment, which includes the transcription start site of the p_C 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 Δ*cya* strain deleted of its adenyl cyclase gene. The RNA–DNA hybrid was digested with S_1 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 S_1-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).

FIG. 2. S_1 nuclease mapping DNA–RNA hybrids of the *araC* message. Total cellular RNA was isolated from cells according to the procedure described by Salser *et al.* (1967) with modifications described by Stoner & Schleif (1983). S_1 nuclease mapping of DNA–RNA hybrids was performed according to the procedure described by Berk & Sharp (1977), Barry *et al.* (1980), and Fukumaki *et al.* (1982); with modifications as described by Stoner & Schleif (1983). The optimum hybridization temperature was 52°C, optimum incubation time for hybrid formation was 3 to 6 h, and the optimum S_1 nuclease digestion conditions were 4 units S_1 per 350-μl reaction at 45°C for 30 min. The samples were electrophoresed on thin (0·40 mm) 8% polyacrylamide/urea denaturing gels as described by Sanger & Coulson (1978). A "G" sequencing reaction of the same fragment, prepared according to Maxam & Gilbert (1980) was run concurrently as a size standard. The gels were autoradiographed with preflashed Kodak X-Omat AR film and an intensifying screen at −70°C. Lane 1, G sequencing reaction size standard. Lane 2, RNA isolated from Δ*araC* cells. Lane 3, RNA from cells with *araC* hyperproducing plasmid pRFS13 grown in the absence of arabinose. Lane 4, RNA from Ara(CBAD)$^+$ cells grown in absence of arabinose. Lane 5, RNA from Ara(CBAD)$^+$ cells grown in absence of arabinose but as a control arabinose was present in the ice/sodium azide/chloramphenicol mixture in the first step of RNA isolation procedure. Lanes 6 to 10, RNA from Ara(CBAD)$^+$ cells, grown in the presence of arabinose for 2, 6, 18, 50 and 150 min. Lane 11, RNA from ΔCya Ara(CBAD)$^+$ cells, −arabinose. Lane 12, RNA from ΔCya Ara(CBAD)$^+$ cells, and induced with arabinose for 6 min. Lane 13, RNA from Ara(CBAD)$^+$ cells, and incubated with fucose for 6 min. Lane 14, RNA from a different isolation from Ara(CBAD)$^+$ cells, and induced with arabinose for 5 min. Lane 15, RNA alone control. Lane 16, DNA alone control.

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

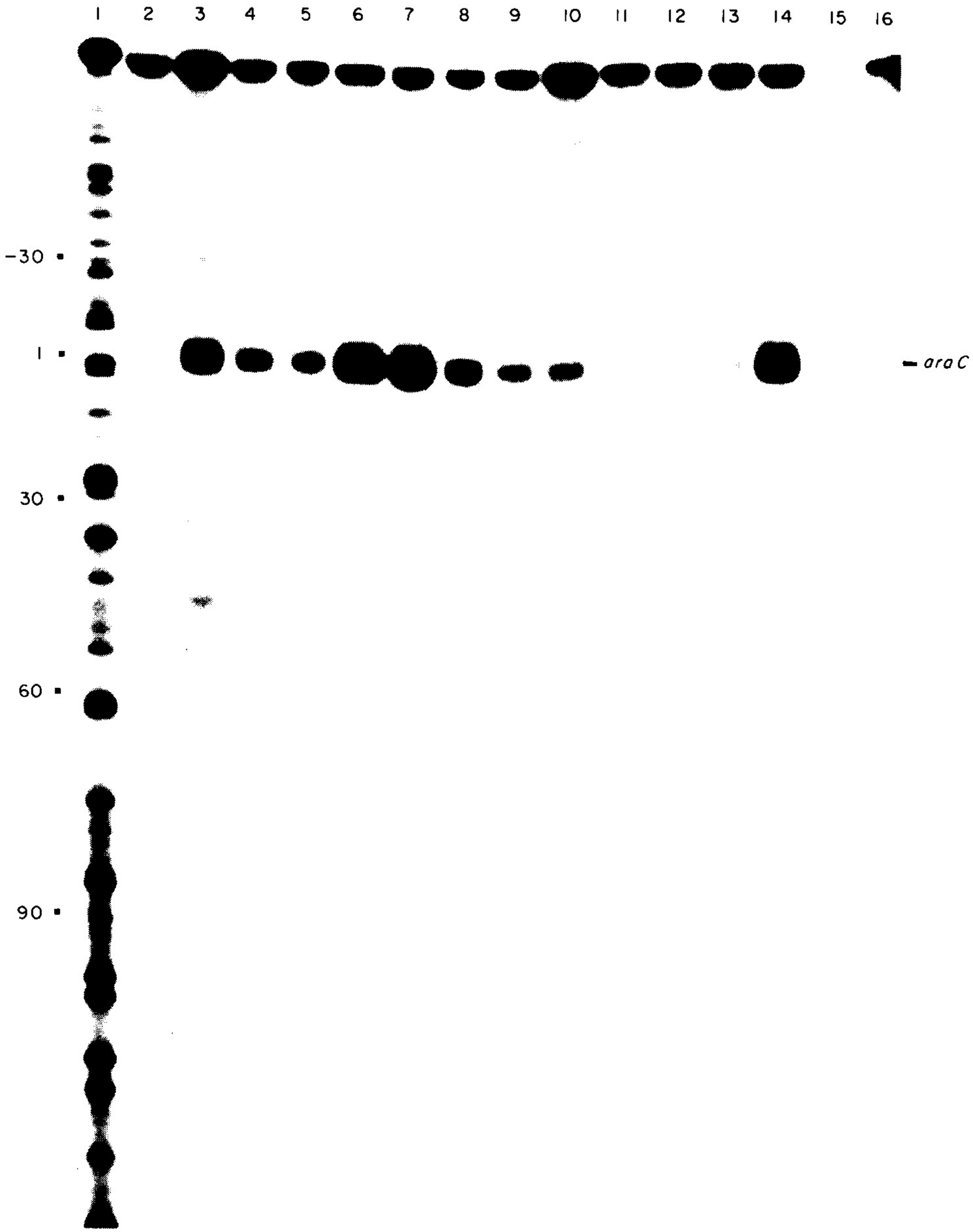

FIG. 2

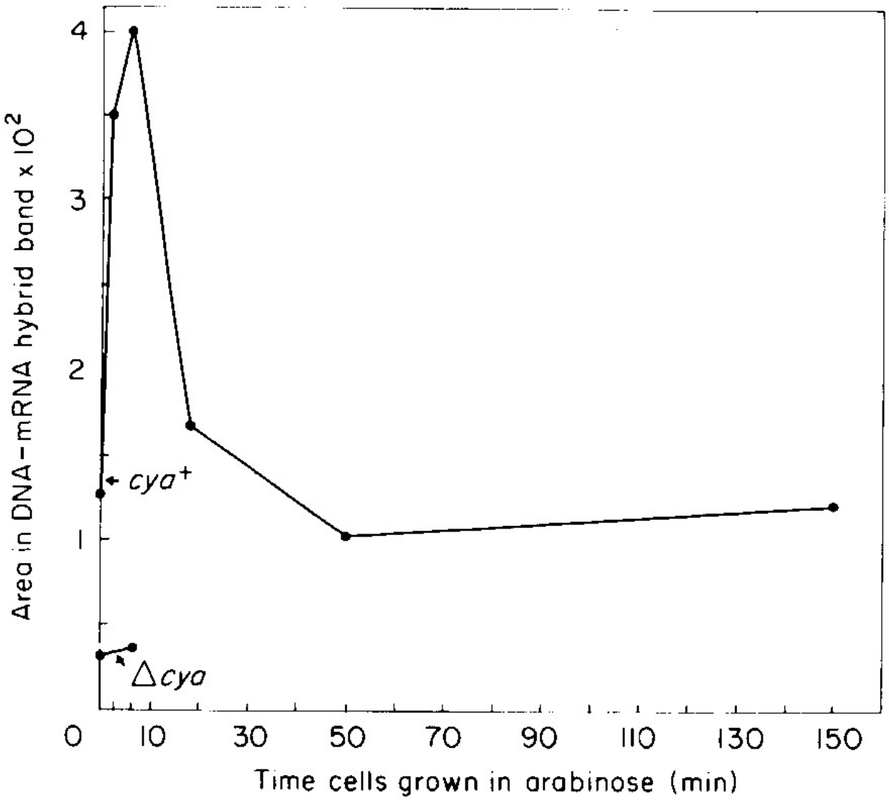

FIG. 3. Kinetics of p_C promoter activity. Cells were harvested and RNA was extracted at various times after arabinose addition. RNA–DNA hybrids were formed, digested with S_1 nuclease, denatured, and separated by electrophoresis. Amount of radioactivity protected in the *araC* [^{32}P]DNA–RNA hybrid was determined as the densitometry of an autoradiogram.

consistent with the result of Haggerty (1977) who found that p_C 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 DNA 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 p_C 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.

Carol M. Stoner
Robert F. Schleif

Received 3 May 1983

REFERENCES

Aiba, H., Adhya, S. & deCrombrugghe, B. (1981). *J. Biol. Chem.* **256**, 11905–11910.
Barry, G., Squires, C. & Squires, C. L. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3331–3335.
Berk, A. J. & Sharp, P. A. (1977). *Cell*, **12**, 721–732.
Brown, C. E. & Hogg, R. W. (1972). *J. Bacteriol.* **111**, 606–613.
Casadaban, M. (1976). *J. Mol. Biol.* **104**, 557–566.
Englesberg, E., In, J., Power, J. & Lee, N. (1965). *J. Bacteriol.* **90**, 946–957.
Fukumaki, Y., Ghosh, P. K., Benz, Jr., Reddy, V. B., Lebowitz, P., Forget, B. G. & Weissman, S. M. (1982). *Cell*, **28**, 585–591.
Greenblatt, J. & Schleif, R. (1971). *Nature New Biol.* **233**, 166–170.
Greenfield, L., Boone, T. & Wilcox, G. (1978). *Proc. Nat. Acad. Sci., U.S.A.* **75**, 4724–4728.
Haggerty, D. (1977). Ph.D. thesis, Brandeis University.
Hahn, S. & Schleif, R. (1983). *J. Bacteriol.* **155**, 593–600.
Hawley, D. & McClure, W. (1983). *Nucl. Acids Res.* in the press.
Hentschel, C., Irminger, J., Bucher, P. & Birnstiel, M. L. (1980). *Nature (London)*, **285**, 147–151.
Hirsh, J. & Schlief, R. (1977). *Cell*, **11**, 545–566.
Kolodrubetz, D. & Schleif, R. (1981). *J. Mol. Biol.* **151**, 215–227.
Lee, N., Wilcox, G., Gielow, W., Arnold, L., Cleary, P. & Englesberg, E. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 634–638.
Maxam, A. & Gilbert, W. (1980). In *Methods in Enzymology* (Grossman, L. & Moldave, K., eds), vol. 65, pp. 499–560, Academic Press, New York.
Ogden, S., Haggerty, D., Stoner, C. M., Kolodrubitz, D. & Schleif, R. (1980). *Proc. Nat. Acad. Sci., U.S.A.* **77**, 3346–3350.
Salser, W., Gesteland, R. F. & Bolle, A. (1967). *Nature New Biol.* **215**, 588–591.
Sanger, F. & Coulson, A. R. (1978). *FEBS Letters*, **87**, 107–110.
Schleif, R. (1969). *J. Mol. Biol.* **46**, 185–196.
Schleif, R. & Favreau, M. A. (1982). *Biochemistry*, **21**, 778–782.
Smith, B. R. & Schleif, R. (1978). *J. Biol. Chem.* **253**, 6931–6933.
Stoner, C. M. & Schleif, R. (1983). *J. Mol. Biol.* In the press.
Wallace, R. G., Lee, N. & Fowler, A. V. (1981). *Gene*, **12**, 179–190.

Edited by S. Brenner