*J. Mol. Biol.* (1975) **96**, 185–199

# Paucity of Sites Mutable to Constitutivity in the *araC* Activator Gene of the L-Arabinose Operon of *Escherichia coli*

NEIL M. NATHANSON AND ROBERT SCHLEIF

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

*(Received 23 December 1974)*

A large number of high-level and low-level constitutive mutations in the *araC* gene of *Escherichia coli* were shown by deletion mapping to lie almost exclusively in two regions of the *araC* gene. Recombination data show that the high-level constitutive mutations are located within two very small regions, each probably less than ten base-pairs, while the low-level constitutive mutations are spread over two broader areas, each centered at the same two regions. All constitutive mutations isolated in either the presence or absence of D-fucose, an analog of L-arabinose which antagonizes induction by arabinose, are altered from the wild type in their response to this analog. A nonsense mutation that maps in one of the constitutive regions can be suppressed to wild type, "low-level" constitutive, or "high-level" constitutive phenotypes, depending on the amino acid inserted at the site of the mutation. This demonstrates that changing a single amino acid can cause dramatic alterations in the regulatory properties of the *araC* activator protein.

## 1. Introduction

Studies of mutational alteration of activity contribute toward understanding the structure and function of proteins. While mutations abolishing the activity of a protein can lie anywhere in the gene, mutations altering specificity may be restricted in their locations. Mutations in the *lac* repressor altering either binding to *lac* operator DNA ($i^{-d}$) or to inducer ($i^{s}$) are each confined to a relatively limited number of sites in the *i* gene, suggesting that these mutations occur in the regions of the protein which determine the binding sites for operator and inducer, respectively (Adler *et al.*, 1972; Pfahl, 1972; Pfahl *et al.*, 1974).

In this paper we present an analysis of mutations to constitutivity in the activator protein of the L-arabinose operon of *Escherichia coli*, the product of the *araC* gene. Induction of the L-arabinose operon, which in a wild-type cell is dependent on the presence of both arabinose and the product of the *araC* gene, is prevented by the L-arabinose analog D-fucose (Englesberg *et al.*, 1965; Irr & Englesberg, 1970; Sheppard & Englesberg, 1967). Mutants resistant to the fucose antagonism can be isolated; such mutants are frequently found to be constitutive, having high levels of enzymes even in the absence of arabinose. Fucose-resistant mutants having less than a normally induced level of enzymes in the absence of inducer have also been reported (Englesberg *et al.*, 1965; Beverin *et al.*, 1971; Greenblatt & Schleif, 1971; Doyle *et al.*, 1972).

Spontaneous mutations to constitutivity occur at a frequency of approximately $10^{-8}$, at least 100-fold lower than the frequency of point mutations abolishing the activity of the *araC* activator. This low mutation rate implies that only a small fraction of all possible mutations in the gene lead to constitutivity, and hence that there are only a limited number of sites where mutations to constitutivity occur. Englesberg *et al.* (1965) examined the recombination of a number of *araC* constitutive ($C^c$) mutants with five point mutations in the *C* gene, and concluded that the majority of $C^c$ mutations were grouped around two of the $C^-$ point mutations. Beverin *et al.* (1971) mapped seven $C^c$ mutations with respect to three partial deletions of the *C* gene and concluded that there were at least three regions where the mutations mapped. Since a study of mutations that affect the regulatory properties of the *araC* activator protein could aid in the elucidation of its mechanism of action, we have examined in detail mutations to constitutivity in *araC*. Using a number of techniques to determine the location and nature of mutations to constitutivity, we found that: (1) mutations to constitutivity occur in only a very limited number of regions of the *araC* gene, (2) the two regions where the vast majority of high-level constitutive mutations occur are both very small, encompassing less than ten base-pairs, (3) low-level constitutive mutations are spread over a broader area in these same two regions, and (4) a change of only one amino acid at a specific site in the *araC* protein can confer either a wild-type or a variety of constitutive phenotypes on the resultant *araC* product.

## 2. Materials and Methods

### (a) *Chemicals and media*

YT broth, minimal salts medium for plates, and M9 minimal medium have been described by Schleif (1969). M10 medium is M9 medium supplemented with $5 \times 10^{-5}$ M-$MnCl_2$, to which were added 0·2% sugar, 0·001% thiamine, and 0·01% amino acids, as noted. D-fucose, L-arabinose, adonitol (ribitol), B1 and amino acids were obtained from Sigma. Penicillin G ($K^+$) was obtained from Squibb. Diethyl sulfate was obtained from Aldrich.

### (b) *Bacterial strains*

All strains used were derived from *E. coli* K12. See Table 1.

#### (i) *Construction of suppressor strains*

Strain NMN127, possessing 3 nonsense mutations, was obtained by mating NMN125 with NMN126. Strain NMN130 was constructed by selection of an $Ara^-$ derivative of NMN127 on minimal medium containing arabinose, glycerol and ribitol (Katz, 1970; Schleif, 1972), followed by P1 transduction to $Ara^+$ $Thr^-$ from strain NMN122. NMN132 was constructed by transduction of NMN130 to $Thr^+$ $Ara^-$ $Leu^-$ using P1 grown on RFS882. The suppressor mutations were then transduced into NMN132 by selection on minimal plates containing glucose, B1, and leucine.

#### (ii) *Isolation of constitutive mutations*

Table 2 lists the constitutive mutations and their methods of isolation.

(1) *Isolation in the presence of fucose.* *F'araC6902-C6926* and *C6928-C6961* were derived from RFS740. Purified colonies of RFS740 were picked and grown overnight to stationary phase in YT broth, and approximately $2 \times 10^9$ cells were spread on minimal/arabinose/glucose plates (Englesberg *et al.*, 1965) containing 0·1% L-arabinose and 0·2% D-fucose. After 2 days' incubation at 35°C, 0 to 50 colonies per culture appeared. "High-level" constitutive strains (*C6902-C6920* and *C6928-C6957*) were screened on minimal plates containing 0·2% glycerol and 0·45% ribitol. The high level of ribulokinase present in a high-level constitutive strain phosphorylates the ribitol to yield ribitolphosphate, which is

TABLE 1

*Bacterial strains used*

| Strain no. | Genotype | Comments |
|---|---|---|
| CGSC2597 | Hfr *phoA5 rel-1 tonA22 supF34* $T_2^R$ | su3$^+$. Strain H12R8a of A. Garen |
| CGSC2598 | Hfr *phoA4 rel-1 tonA22*ul *supD32* $T_2^R$ | sul$^+$. Strain 26R1e of A. Garen (Garen *et al.*, 1965) |
| CGSC2599 | Hfr *phoA4 rel-1 tonA22 supE33* $T_2^R$ | su2$^+$. Strain S26R1d of A. Garen (Garen *et al.*, 1965) |
| CGSC4487 | Hfr *phoA4 rel-1 tonA22 sup-51* $T_2^R$ | su6$^+$. Strain S26, Su*6*$^+$ of A. Garen (Chan & Garen, 1969) |
| CGSC4619 | Hfr *phoA6 rel-1 tonA22 supG46* $T_2^R$ | su5$^+$. Strain U11r1d of A. Garen (Gallucci & Garen, 1966) |
| CSH4 | F$^-$ *trp lacZ strA thi* | Amber mutations in *trp* and *lacZ* (Miller, 1972) |
| JTL28 | F$^-$ *leu thr lac74 tsx28* | Lis & Schleif (1973) |
| NMN118 | F′ *thr*$^+$ *ara*$^+$ *leu*$^+$/*ara-498 leu498 lac74 thi tonA* | From RFS740 |
| NMN119 | F′ *thr*$^+$ *ara*$^+$ *leu*$^+$/*ara-498 leu498 lac74 thi tonA str* | From NMN118 |
| NMN122 | F$^-$ *leu thr lac74 tonA str tsx28* | From JTL28 |
| NMN125 | F$^-$ *trp lacZ strA thi tonA* | From CSH4 |
| NMN126 | Hfr *metB argE thi rif* | From PX4 |
| NMN127 | F$^-$ *trp lacZ argE rif thi strA tonA* | Progeny of cross between NMN125 × NMN126 |
| NMN130 | F$^-$ *argE trp lacZ thr tonA rif thi strA* | From 127 |
| NMN132 | F$^-$ *argE trp lacZ leu araC882 rif thi strA tonA* | From NMN130, contains partial deletion of *araC* from RFS882 |
| NMN134 | Like NMN132, *supG46* | Contains su5$^+$ of CGSC4619 |
| NMN135 | Like NMN132, *supF34* | Contains su3$^+$ of CGSC2597 |
| NMN136 | Like NMN132, *supD32* | Contains sul$^+$ of CGSC2598 |
| NMN137 | Like NMN132, *supE33* | Contains su2$^+$ of CGSC2599 |
| NMN138 | Like NMN132, *sup-51* | Contains su6$^+$ of CGSC4487 |
| PX4 | Hfr *metB argE thi* | *argE* is amber (strain CSH70 of Miller, 1972) |
| RFS726 | F$^-$ *ara-498 leu-498 lac74 thi* | Deletion of *leu* and at least *araC*, *B* and *A*; *RV lac* deletion |
| RFS740 | F′ *thr*$^+$ *ara*$^+$ *leu*$^+$/*ara-498 leu-498 lac74 thi* | KLF1 episome in RFS726 (Schleif, 1972) |
| RFS882 | F$^-$ *leu-73*△*(gal att bio uvr B) lac 74 str lam830 araC882* | Deletion of most of *araC* (Schleif, 1972) |
| RFSF′240 | F′ *thr*$^+$ *araC240 D591 leu*$^+$/*ara-498 leu-498 thi lac74* | Nonsense mutations in *araC* and *D* (Schleif, 1972) |
| RFSF′273 | F′ *thr*$^+$ *araC273 D591 leu*$^+$/*ara-498 leu-498 thi lac74* | Nonsense mutations in *araC* and *D* (Schleif, 1974) |
| RFSF′280 | F′ *thr*$^+$ *araC280 D591 leu*$^+$/*ara-498 leu-498 thi lac74* | Nonsense mutations in *araC* and *D* (Schleif, 1972) |

All strains used were *E. coli* K12. Strains whose numbers beginning with CGSC were obtained from the *E. coli* Genetic Stock Center, New Haven, Conn.

bacteriostatic (Katz, 1970). Preliminary candidates were verified by assay of L-arabinose isomerase. "Low-level" $C^c$ strains (*C6921-C6926*, *C6958-C6961*) were screened for the ability to grow on minimal medium containing 0·2% glycerol and 0·45% ribitol, but not on minimal medium containing 0·2% glycerol, 0·45% ribitol, 0·1% arabinose and 0·2% fucose. Arabinose isomerase assays were used to verify the conclusions of the screening on plates. Low-level $C^c$ strains are not growth-inhibited on minimal medium containing glycerol and ribitol, but are fucose-resistant and therefore are fully induced in the presence of arabinose and fucose. Thus these as well as high-level $C^c$ strains will not grow on minimal medium containing glycerol, ribitol, arabinose and fucose. High-level constitutive strains F′*ara-C6969-C6976* were isolated after mutagenesis of NMN118 with 2-ethyl sulfate: 0·1 ml of cells from an overnight culture was diluted into 5 ml of M10 medium and 3 drops of diethyl sulfate were added. After 15 min incubation at 35°C, 0·05 ml was diluted into 5 ml of YT broth, grown overnight, then plated on minimal/arabinose/fucose medium as before. Diethyl sulfate mutagenesis caused approximately 100-fold increase in the number of fucose-resistant colonies compared to the number obtained spontaneously. To prevent isolation of siblings, no more than one high-level and one low-level constitutive mutant was taken from the same culture.

(2) *Isolation of high-level $C^c$ mutations without fucose.* F′*araC6927* was isolated from RFS740 and F′*C6962* was isolated from NMN 118. Cells were grown in M10 medium containing B1 and glycerol at 37°C to a density of $1 \times 10^8$/ml; ribitol was added to a final concentration of 0·45%, and the cells were grown for an additional 30 min. Penicillin G (2000 to 6000 units/ml), was added and the culture was incubated for a further 1·5 to 3 h, until the turbidity of the culture stopped decreasing. Surviving cells were then concentrated by centrifugation, washed once with M10, resuspended in fresh medium, and grown. This cycle was repeated twice more; cells were then plated out on YT plates. The colonies that exhibited ribitol sensitivity following replication onto plates containing glycerol and ribitol were grown for enzymatic assay of arabinose isomerase. Two out of 20 cultures treated in this way yielded constitutive mutants.

F′*araC6962-C6968* was obtained in a similar fashion except that cultures of NMN119, a streptomycin-resistant derivative of NMN118, were first mutagenized with diethyl sulfate. Streptomycin (0·01%) was present in the medium during growth.

(iii) *Mapping of constitutive mutations*

The constitutive mutations, all isolated on episomes, were mapped with a large number of deletions extending into the *araC* gene (Schleif, 1972). If an episomal constitutive mutation is mated into a strain carrying at least that part of the wild-type *araC* gene allelic to the $C^c$ mutation, recombination can occur, removing the $C^c$ mutation from the active *araC* gene on the episome and substituting the $C^+$ allele. The chromosomal *araC* gene is not functional in these experiments because part of it is deleted, so that the only active *araC* gene in the cell changes from $C^c$ to $C^+$. The small fraction of cells in a population accomplishing such a recombination were detected on plates allowing growth of $AraC^+$ cells and not allowing growth of $AraC^c$ cells. If the deletion on the chromosome extends past the location of the $C^c$ mutation on the episome, then a functional $C^+$ gene cannot be obtained by recombination. The sensitivity to ribitol of cells possessing high levels of ribulokinase allows this selection. High-level constitutive strains will not grow on minimal glycerol medium containing ribitol. Low-level constitutive strains are not growth-inhibited on minimal glycerol ribitol medium, but are fucose-resistant and thus are fully induced in the presence of arabinose plus fucose; therefore, these, as well as high-level constitutive strains, will not grow on minimal medium containing glycerol, ribitol, arabinose and fucose. Thus, by scoring for growth on the appropriate medium following matings with strains deleted of portions of the *araC* gene, the genetic locations of $C^c$ mutations were determined.

Cells were grown and mated in polypropylene blocks as previously described (Schleif, 1972). After 3 to 5 h mating at 35°C, samples of the mating mixture were spotted onto minimal plates containing B1 and glycerol and either 0·45% ribitol, for the mapping of high-level $C^c$ mutations, or 0·1% arabinose, 0·2% fucose and 0·45% ribitol for the mapping of low-level $C^c$ mutations. The mating mixtures were also spotted onto minimal plates

containing glucose and streptomycin as a control to ensure that lack of growth on medium containing ribitol was not from failure to transfer the episome into the female deletion strain. Crosses were scored after 36 to 48 h.

(iv) *Recombination studies*

Constitutive mutations were introduced onto the chromosome of strain NMN122 by P1 transduction to Leu$^+$, followed by screening for cotransduction of constitutivity. Diploids were then constructed by mating the episomal C$^c$ mutations into the female constitutive strains. After purification of progeny, diploids were grown to stationary phase in YT broth, concentrated by centrifugation, resuspended in M10 medium, and plated at various dilutions on minimal plates containing B1, glycerol and either 0·45% ribitol, for high-level/high-level diploids, or 1·3% ribitol, 0·1% arabinose and 0·2% fucose for low-level/low-level or low-level/high-level diploids. Constitutive strains are unable to grow, due to accumulation of ribitol phosphate synthesized by phosphorylation of ribitol by ribulokinase. If 2 C$^c$ mutations are not at the same site, they should be able to recombine to yield a C$^+$ gene. Since C$^+$ is dominant to C$^c$ (Englesberg *et al.*, 1965), such a recombinant will be phenotypically C$^+$. This recombinant can be detected amongst a large population of C$^c$ mutants utilizing the ribitol sensitivity of constitutive strains. Homozygous diploids were found to produce ribitol-resistant colonies at a frequency of approximately $10^{-8}$. These could result from mutation to either B$^-$ or C$^-$, followed by recombination and homozygote formation, reversion of either C$^c$ mutation to C$^+$, or mutations to ribitol resistance occurring outside the arabinose operon.

(c) *Enzyme assays*

Assays of L-arabinose isomerase were performed as previously described (Schleif, 1972).

## 3. Results

(a) *Isolation of C$^c$ mutations*

Induction of the *ara* operon in a cell possessing a wild-type *araC* gene is prevented by the L-arabinose analog D-fucose. These cells will not grow on a minimal plate containing L-arabinose as the sole carbon source if D-fucose is present. Very rare mutations in the *araC* gene can occur, allowing cells to grow on these plates; such mutants, resistant to the fucose antagonism, have widely variable levels of the *ara* enzymes in the absence of arabinose plus fucose. High-level constitutives are defined

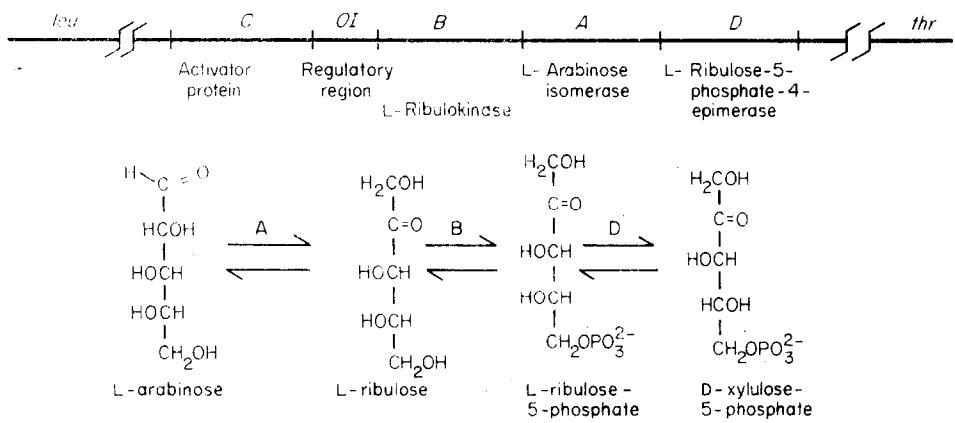


FIG. 1. The L-arabinose operon of *E. coli*, its genetic map and associated biochemical reactions (Englesberg, 1971).

as those mutants possessing, in the absence of arabinose, the levels of *ara* enzymes normally present following induction by arabinose. Mutants which are fully resistant to fucose but possess less than half the normally induced levels of enzymes in the absence of any inducer are defined as low-level constitutives.

Two selective methods were used to isolate constitutive mutations. The first was based on the fucose sensitivity of wild-type cells. The second was based on the ribitol sensitivity of high-level constitutive strains. The latter was developed to test whether the mutations isolated by their fucose resistance are a subset of all possible *ara* constitutive mutants. Cells possessing high levels of ribulokinase (Fig. 1) do not grow in the presence of ribitol (Katz, 1970), presumably due to the resulting accumulation of bacteriostatic ribitolphosphate. Rare high-level $C^c$ mutants in an $AraC^+$ culture can therefore be isolated by first inhibiting the growth of $C^c$ cells by addition of ribitol, then killing the growing $C^+$ cells by penicillin. Since only cells with high levels of ribulokinase are growth-inhibited by ribitol, only high-level $C^c$ mutations will be isolated with this method.

TABLE 2

*$C^c$ mutations used*

| Mutation numbers | Level of constitutivity | Method of selection |
|---|---|---|
| *C6902–C6920, C6928–C6957* | High | Spontaneously on minimal/arabinose/fucose |
| *C6921–C6926, C6958–C6961* | Low | Spontaneously on minimal/arabinose/fucose |
| *C6927, C6962* | High | Spontaneously by ribitol/penicillin |
| *C6969–C6976* | High | Diethyl sulfate-induced on minimal/arabinose/fucose |
| *C6963–C6968* | High | Diethyl sulfate-induced by ribitol/penicillin |

For details of isolation see Materials and Methods. Map positions of constitutive mutations are shown in Fig. 2.

Table 2 lists the constitutive mutations and their methods of isolation. All ten high-level $C^c$ strains isolated on the basis of ribitol sensitivity without the use of D-fucose as a selection against wild-type cells were found to be fucose-resistant. As has been previously noted, low-level constitutive mutants, while fucose-resistant and fully induced in the presence of either arabinose alone or arabinose plus fucose, are induced to various extents by fucose alone (Beverin *et al.*, 1971; Doyle *et al.*, 1972; Greenblatt & Schleif, 1971).

### (b) *Deletion mapping of high and low-level constitutive mutations*

As the distance separating point mutations decreases, the ordering of mutations using recombination frequencies becomes inaccurate due to multiple crossovers in localized regions (Amati & Meselson, 1965; Sadler & Smith, 1971). Deletion mapping

becomes a better technique for fine structure mapping since experimentally one merely tests whether or not a functional gene can be reconstructed (Benzer, 1961). The availability of a large number of independent deletions ending in the *araC* gene (Schleif, 1972) allowed this technique to be used with high resolution.

The constitutive mutations, all located on episomes, were mapped after transfer of the episomes into a large number of strains containing independent deletions ending in the *araC* gene. The rationale for this is described in Materials and Methods. The constitutive mutations were found to map almost exclusively in two clusters in the *C* gene (Fig. 2), one at the end of the gene adjacent to the regulatory region, the other approximately one-third the length of the *C* gene from the operator. High and low-level $C^c$ mutations, both spontaneous and diethyl sulfate-induced, are found in each region, although the operator proximal region contains largely high-level $C^c$ mutations and the other region contains predominantly low-level $C^c$ mutations.

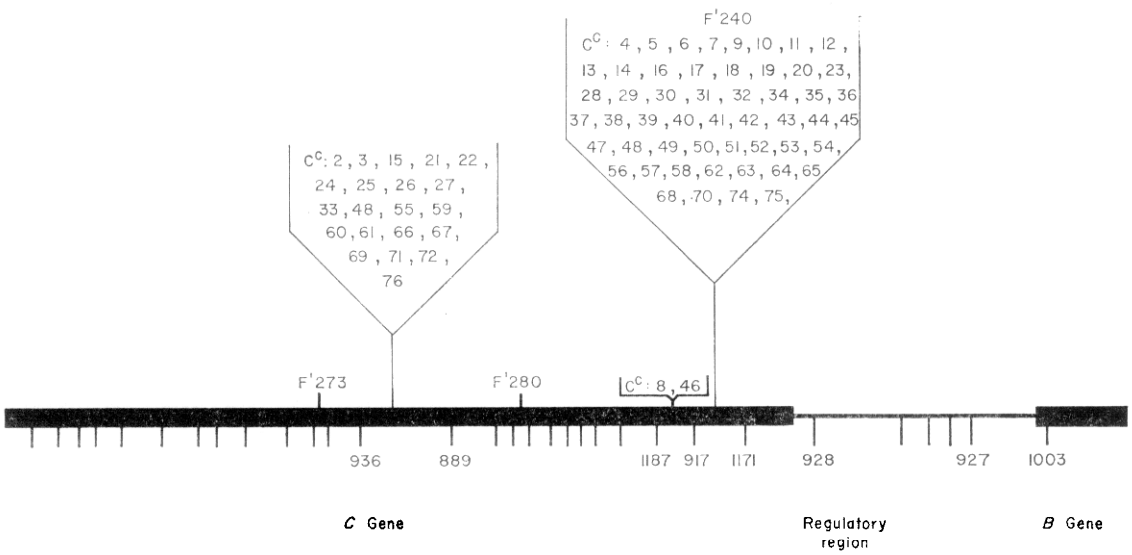


FIG. 2. Deletion map of *araC* constitutive mutations. Each hashmark represents a set of deletions which are separated from the adjacent sets of deletions by either point mutations (Schleif, 1972) or endpoint of portions of the arabinose operon carried on λ*para* transducing phages (Schleif & Lis, 1975). F'240, F'273 and F'280 are *araC* nonsense mutations (Schleif, 1972). Constitutive mutation number *xy* on the map corresponds to mutation *araC69xy* in the text. The various methods used to isolate the $C^c$ mutations are described in Materials and Methods and listed in Table 2.

### (c) *Recombinations between $C^c$ mutations*

#### (i) *Recombination of high-level constitutive mutations with high-level constitutive mutations*

Recombination between $C^c$ mutations mapping in the same region was investigated in an attempt to place limits on the size of the regions: constitutive mutations not at the same site should be able to recombine to form a $C^+$ allele. Since a $C^+$ is normally dominant to $C^c$ (Englesberg *et al.*, 1965), the phenotype of such a recombinant will be $C^+$. It should be noted that the deletion mapping technique described in the previous

section does not require the dominance of $C^+$ to $C^c$. However, the recombination studies described here do require that $C^c$ be recessive to $C^+$. All the $C^c$ mutations we have tested, namely those used in the recombination studies, were recessive to $C^+$. Recombinant $C^+$ cells can be selected for in the $C^c$ population by the ribitol sensitivity of constitutive strains (Katz, 1970). Recombination was examined in diploid strains containing a $C^c$ mutation on the chromosome and a second $C^c$ mutation on the episome. The following $C^c$ mutations were transduced onto the chromosome of a female strain: *C6915*, *C6920*, *C6927* and *C6928*. Altogether, 28 different diploids were constructed, each carrying one of each of the above mutations and one of the following F′$C^c$ mutations: *C6902*, *C6904*, *C6915*, *C6920*, *C6927*, *C6928* and *C6907*.

This recombination mapping method was tested in two ways. First, diploids containing two $C^c$ mutations, one from each of the two main regions of constitutivity, were tested for the recombination which should occur by plating on minimal medium containing glycerol and ribitol. All such diploids gave rise to $C^+$ recombinant colonies at a frequency of approximately $10^{-5}$. A homozygous diploid containing two identical $C^c$ mutations should not be able to recombine to form a $C^+$ allele. Nonetheless homozygous diploids containing two identical $C^c$ mutations from either region were found to yield ribitol-resistant colonies, but at a frequency of $10^{-8}$. These result from phenomena other than recombination, as described in Materials and Methods. Their frequency constitutes the lower limit of resolution in the recombination tests. Thus, two $C^c$ mutations, one from each of the two main constitutive regions in the *C* gene, yield recombinants at a frequency 1000-fold higher than the background. As a further test, recombination with a constitutive mutation located between the two main regions of constitutivity was studied. F′*araC6908*, which maps between the two main regions of constitutivity, was found to recombine with *C6915* and *C6927* at a frequency of approximately $10^{-5}$, and to recombine with the closer *C6920* and *C6928* at a frequency of approximately $10^{-6}$.

Recombination between two non-identical $C^c$ mutations from the same region was examined to determine whether the constitutive mutations which could not be separated by deletion mapping could be shown to be separable by recombination. In all cases tested, no recombination greater than the $10^{-8}$ background could be detected. Thus, while recombination between $C^c$ mutations mapping in different regions of the C gene is readily detected, no recombination is seen between high-level $C^c$ mutations which map in the same region of the gene, implying that the size of each region is very small.

(ii) *Recombination of low-level constitutive mutations with high-level and low-level constitutive mutations*

While no recombination between two high-level constitutive mutations mapping in the same region can be detected, recombination between low-level constitutive mutations with either high-level or other low-level constitutive mutations mapping in the same region can be observed. Two female strains were constructed, one with low-level $C^c$ mutation *C6923*, the other with low-level $C^c$ mutation *C6960*; diploids were constructed by addition of an episome containing one of the following high-level or low-level constitutive mutations: *C6915* and *C6920* (high-level), and *C6923*, *C6925* and *C6960* (low-level). $C^+$ recombinants were detected on plates containing glycerol, ribitol, arabinose, and fucose, as described in Materials and Methods. The background

frequency of ribitol-resistant colonies seen in an isogenic diploid containing the same constitutive mutation on both the chromosome and episome was approximately $10^{-8}$. *C6923* recombines with *C6915*, *C6920*, *C6927* and *C6960* at a frequency of $10^{-5}$. *C6960* recombines with *C6915* and *C6927* at a frequency of $10^{-6}$, and recombines with *C6920* and *C6925* at a frequency of $10^{-5}$. In contrast to recombination between two high-level constitutive mutations, a low-level $C^c$ mutation will recombine with both high and low-level $C^c$ mutations which map in the same region, implying that the sites where low-level constitutive mutations can occur are distinct from those for high-level constitutive mutations and encompass a larger region.

(iii) *Luria–Delbrück fluctuations in recombination*

Recombination between two $C^c$ mutations could occur at any time during the 30 generations of growth prior to being plated on selective medium. Differences in the time of this recombination would produce large fluctuations in the number of $C^+$ recombinants ultimately found, making the determination of recombination frequencies irreproducible. This situation is analogous to that investigated by Luria & Delbrück (1943) for fluctuations in the numbers of colonies resistant to phage T1. A measure of the magnitude of the fluctuation problem is given by the ratio of the standard deviation in the observed number of recombinant colonies in many cultures to the average number of recombinants. If the ratio is large, as it was for the case studied by Luria & Delbrück, then fluctuations will cause wide variations from culture to culture in the observed number of colonies; if this ratio is small, fluctuations will not be a problem. The ratio can either be calculated from the data obtained on many cultures, or predicted from data obtained on a few cultures by the formula $SD/r = \sqrt{C}/\ln[NCa]$ (equation 12 of Luria & Delbrück, 1943) where: $SD$ is the standard deviation, $r$ is the observed average number of colonies per culture, $C$ is the number of cultures, $a$ is the recombination frequency, $N$ is the number of cells in each culture.

In our case, $10^8$ cells yield 1000 recombinants for a recombination frequency of $10^{-5}$, and the ratio of the standard deviation to the measured number of recombinants in the culture is $SD/r = 1/\ln[10^3] = 0{\cdot}14$. Thus for recombination frequencies as high as our case, the standard deviation is much smaller than the number of observed recombinants, and Luria–Delbrück-type fluctuations are not large.

(d) *Suppression of an* araC *nonsense mutation to constitutivity*

$AraC^-$ nonsense mutations can be used to create aberrant *araC* proteins; specific amino acids can be inserted at the sites of the nonsense mutations by utilizing a variety of characterized suppressor mutations. Two complexities arise in this type of experiment: first, the varied efficiency with which the different suppressors insert an amino acid and second, suppressors can add amino acids to the carboxyl terminus of proteins by suppression of the normal chain termination signal (Lu & Rich, 1971). For our study of constitutive mutations it was first necessary to show that a wild type and a constitutive *araC* gene in each of the suppressor strains had the same regulatory properties as they do in non-suppressing strains. Table 3 shows this to be the case.

The ability of suppressing strains to substitute a variety of amino acids at known locations allowed experiments to further test the origins and nature of constitutive mutations. First, by the study of several nonsense mutations in the *araC* gene, it was found as expected, that alteration of amino acids did not usually lead to constitutivity.

TABLE 3

*Effects of arabinose and fucose on F'AraC$^+$ and F'AraC$^c$ in various suppressor strains on levels of* L-*arabinose isomerase*

| Strain | Addition | | | |
|---|---|---|---|---|
| | None | Arabinose | Fucose | Arabinose + fucose |
| F'AraC$^+$/Su$^-$ | 91 | 24,563 | 121 | 638 |
| F'AraC$^+$/Su1 | 53 | 28,750 | 85 | 613 |
| F'AraC$^+$/Su2 | 99 | 29,080 | 111 | 549 |
| F'AraC$^+$/Su3 | 120 | 17,430 | 88 | 656 |
| F'AraC$^+$/Su5 | 91 | 27,720 | 80 | 560 |
| F'AraC$^+$/Su6 | 246 | 40,980 | 153 | 534 |
| F'AraC$^c$/Su$^-$ | 26,080 | 45,660 | 96,500 | 48,350 |
| F'AraC$^c$/Su1 | 23,030 | 26,700 | 80,340 | 24,990 |
| F'AraC$^c$/Su2 | 24,960 | 23,550 | 91,640 | 33,930 |
| F'AraC$^c$/Su3 | 114,440 | 106,760 | 198,740 | 146,100 |
| F'AraC$^c$/Su5 | 15,980 | 43,380 | 58,010 | 24,350 |
| F'AraC$^c$/Su6 | 35,330 | 85,710 | 185,330 | 90,280 |

F'AraC$^+$, wild-type KLF1 episome; F'AraC$^c$, F'*araC6920*; Su$^-$, NMN132; Su1, NMN136; Su2, NMN137; Su3, NMN135; Su5, NMN134; Su6, NMN138. Cells were grown in 10 ml M10/$B_1$/glycerol/arginine/tryptophan in 50-ml Erlenmeyer flasks at 37°C to a density of $1 \times 10^8$. Arabinose and fucose were present at 0·2% and 0·3%, respectively, as indicated. Toluene-treated cells were assayed for L-arabinose isomerase as described in Materials and Methods. Data are presented as monomers isomerase per cell. All assays were performed at least twice; the standard deviation in most cases is less than 10%, and in all cases is less than 20%.

TABLE 4

*Effects of arabinose and fucose on F'273 and F'280 in various suppressor strains on levels of* L-*arabinose isomerase*

| Strain | Addition | | | |
|---|---|---|---|---|
| | None | Arabinose | Fucose | Arabinose + fucose |
| F'273/Su$^-$ | 58 | 99 | 168 | 184 |
| F'273/Su1 | 156 | 16,380 | 164 | 559 |
| F'273/Su2 | 155 | 4825 | 126 | 199 |
| F'273/Su3 | 136 | 21,763 | 151 | 668 |
| F'273/Su5 | 121 | 1413 | 198 | 469 |
| F'273/Su6 | 206 | 40,350 | 153 | 650 |
| F'280/Su$^-$ | 54 | 76 | 199 | 208 |
| F'280/Su1 | 80 | 21,940 | 43 | 86 |
| F'280/Su2 | 59 | 19,710 | 81 | 270 |
| F'280/Su3 | 256 | 22,650 | 375 | 1913 |
| F'280/Su5 | 55 | 4275 | 123 | 415 |
| F'280/Su6 | 163 | 19,550 | 153 | 450 |

Episomal AraC$^-$ nonsense mutants in the various suppressor strains were grown and assayed as described in the legend to Table 3.

Table 4 shows such data from experiments with two $C^-$ nonsense mutations, F'273 and F'280, both lying outside the two main regions where constitutive mutations occur in the *C* gene. In none of the suppressor strains is either nonsense mutation suppressed to constitutivity; several of these strains were not normally inducible by arabinose, most likely due to an inability of the inserted amino acid to fully restore *araC* activity.

Fortunately, one *araC* nonsense mutation, F'240, was available in the operator proximal region of the *C* gene where a large number of constitutive mutations map. Table 5 shows that different *araC* protein activities can be produced by varying the amino acid inserted into this nonsense codon site. The insertion of serine leads to a "normal" $C^+$ wild-type phenotype–arabinose-inducible and fucose-sensitive. However, the insertion of either glutamine, tyrosine or leucine leads to a high-level constitutive phenotype; in the absence of arabinose or fucose the cells possess fully induced levels of L-arabinose isomerase. The effects of arabinose and fucose on the three strains are different, however: while the insertion of glutamine and tyrosine enables further induction by both arabinose and fucose, the insertion of leucine allows further induction only by arabinose. The insertion of lysine causes a novel phenotype: in the absence of inducer the strain appears as a "low-level" $C^c$, with a basal level approximately 20-fold higher than a normal $C^+$. Arabinose induces to the normal level; fucose, however, antagonizes arabinose induction, and reduces the elevated basal level. Thus "wild type" or a variety of "constitutive" phenotypes result, demonstrating that this site has a crucial role in determining the regulatory properties of the *araC* activator protein.

TABLE 5

*Effects of arabinose and fucose on F'240 in various suppressor strains on levels of L-arabinose isomerase*

| F'240 in strain | Amino acid inserted | Addition | | | |
|---|---|---|---|---|---|
| | | None | Arabinose | Fucose | Arabinose + fucose |
| $Su^-$ | — | 81 | 144 | 146 | 124 |
| Su1 | Serine | 105 | 23,500 | 80 | 535 |
| Su2 | Glutamine | 15,162 | 28,990 | 30,280 | 23,260 |
| Su3 | Tyrosine | 19,880 | 39,660 | 58,900 | 39,150 |
| Su5 | Lysine | 1850 | 22,140 | 759 | 2963 |
| Su6 | Leucine | 20,500 | 56,040 | 14,450 | 22,130 |

F'240 in the various suppressor strains was grown and assayed as described in the legend to Table 3.

## 4. Discussion

Our results indicate that mutations to constitutivity can occur in only a limited number of regions in the *araC* gene. The 75 independently isolated constitutive mutations map almost exclusively in two small regions of the gene; one region lies very near the operator proximal end of the *C* gene, while the other region is approximately one-third the length of the *C* gene from the operator end. Most of the constitutive mutations were isolated on the basis of resistance to D-fucose, an analog of

L-arabinose that in a wild-type cell prevents induction; spontaneous high-level $C^c$ mutations, spontaneous low-level $C^c$ mutations, and mutagen-induced high-level $C^c$ mutations were isolated. Constitutive mutations isolated on the basis of their fucose resistance might be only a subset of all possible constitutive mutations. Therefore, a second method for isolation of constitutive mutations was developed. This was based on the ribitol sensitivity of cells possessing a high level of ribulokinase. Each of the isolation methods yields constitutive mutations which map in either region; high-level $C^c$ mutations map predominantely at the operator proximal region and low-level $C^c$ mutations map predominantely in the middle region.

All ten $C^c$ mutations isolated on the basis of ribitol sensitivity without fucose were found to be fucose-resistant. Messer (1974) has also isolated an *araC* constitutive mutation using a selection procedure not based on fucose resistance. This $C^c$ mutation, like the ones we isolated without fucose, turned out to be fucose-resistant. Thus, while mutations producing fucose resistance do not always confer high-level constitutivity, mutations to high-level constitutivity have always been found to confer fucose resistance.

Isolation of mutations producing constitutive promoters for the structural genes might have been expected during the isolation of constitutive strains using ribitol sensitivity. However, a cell possessing such a mutation would be fucose-sensitive, since fucose antagonizes the *araC* protein-dependent induction of the arabinose permease (Englesberg *et al.*, 1965). None of our constitutive strains isolated on the basis of ribitol sensitivity exhibited such fucose sensitivity, so that these constitutive mutations are most likely not due to a promoter mutation. This is not surprising, since such promoter mutations ($I^c$) occur even more infrequently than $C^c$ mutations and their isolation requires more stringent selective techniques (Englesberg *et al.*, 1969; Gielow *et al.*, 1971).

Recombination between various constitutive mutations was examined to determine whether constitutive mutations which map at the same region could recombine to form a $C^+$ gene. Recombination between two high-level $C^c$ mutations mapping in the same region could not be detected above the background ($10^{-8}$) in the recombination test. However, a diploid consisting of one high-level $C^c$ mutation from each of the two main regions of constitutivity recombines at a frequency of approximately $10^{-5}$, 1000-fold higher than the background. This data provides sufficient information to permit an estimation of the sizes of each of the regions in which high-level constitutive mutations occur.

Analysis of a fine-structure deletion map of the arabinose operon showed that the frequency of deletions ending in an interval was roughly proportional to the size of the interval (Schleif, 1972). Later mapping in the arabinose operon by physical methods confirmed this (Schleif & Lis, 1975). The distance between the two regions containing high-level constitutive mutations can thus be estimated at 300 bases. Using the 300 base-pair separation between the regions as a recombination standard, the size of each of the regions can be estimated. Since the recombination frequency between these two regions was 1000-fold above the reversion background and since recombination within each of these regions was not detectable above the background, we could conclude that the size of each region is less than one-1000th of 300 bases. This reasoning is based on the assumption that recombination frequency is proportional to the distance separating two markers. However, this assumption is not fully correct in at least two ways. First, for distances less than several base-pairs, the

recombination frequency falls off more rapidly than expected by linear extrapolation from recombination between more widely separated markers (Yanofsky *et al.*, 1964, 1967; Yanofsky & Horn, 1972). Second, in some cases marker-specific effects introduce appreciable distortions in the linear relation between distance and recombination frequency (Drapeau *et al.*, 1968). In the light of these effects, we consider ten base-pairs to be a realistic estimation for the maximum size for each of the two regions in which high-level constitutive mutations occur.

In contrast to high-level constitutive mutations, the low-level constitutive mutations are clearly separable from each other and from high-level $C^c$ mutations mapping in the same region. The high values of recombination frequencies observed between a low-level constitutive mutation and high-level and low-level constitutive mutations in the same region, $10^{-6}$ to $10^{-5}$, approach those seen between high-level constitutive mapping in the two different regions. These results indicate that the sites where low-level constitutive mutations can occur encompass a larger region than those for high level constitutives, i.e. mutations to low-level constitutivity have a less stringent requirement for their exact location.

Due to their close structural similarities, arabinose and fucose have been postulated to bind to the same site on the *araC* protein (Beverin *et al.*, 1971; Doyle *et al.*, 1972). Since mutations to fucose resistance alter the inducer specificity to prevent antagonism by fucose of arabinose induction, they could change the structure of the arabinose binding site. If this were true, mutations to fucose resistance might be confined to one or a few localized regions of the *araC* gene, those forming the arabinose binding site on the *araC* activator protein. We found that constitutive mutations are localized almost exclusively in two distinct regions of the *C* gene. If these regions comprise the arabinose binding site, they should be close to each other in the teritary structure of the protein. The *araC* activator protein would then be similar to several other proteins whose ligand binding specificities are altered by changes in a few widely separated amino acid positions. A well-characterized example is immunoglobulin. The hypervariable regions of immunoglobulin occur at two or three widely spaced regions in the primary sequence of the L and H chains of the protein (Wu & Kabat, 1970). These regions have been shown by affinity labeling to lie at the antigen binding site (Singer *et al.*, 1971) and are in close spatial proximity in the three-dimensional structure of the protein (Poljak *et al.*, 1973,1974).

The experiments on suppression of *araC* nonsense mutations show the effects on the regulatory properties of the *araC* activator protein caused by the insertion of a variety of amino acids at the sites of several nonsense mutations in the *araC* gene. Except for suppression of F'240, a nonsense mutation lying in one of the regions where constitutive mutations occur, variation in the amino acid had no effect on the regulatory properties of the *araC* protein. However, four distinguishable regulatory phenotypes are produced in nonsense mutation F'240 by the insertion of different amino acids by nonsense suppressing strains (Table 5). Insertion of serine at this site produces a wild-type C phenotype: low basal level, arabinose-inducible, fucose-sensitive. When either glutamine, tyrosine or leucine are inserted at this site in the *araC* protein, the cells have the phenotype of a high-level constitutive with near fully-induced levels of arabinose isomerase in the absence of arabinose. However, the responses of the three strains to fucose are not identical. When lysine is inserted the cell has the unusual phenotype of being "low-level" constitutive in the absence of arabinose with approximately a 20-fold elevated basal level. This is further induced

by arabinose to the normal fully induced level. This strain, however, is fucose-sensitive. A low-level fucose-sensitive constitutive mutant of this type would have eluded isolation by the methods used in this work to select constitutive mutants. Since F'240 is inducible by arabinose in all five suppressor strains, it is unlikely that the differences in the non-induced levels are due to differences in the efficiencies of the suppressor mutations used. Most likely the different constitutive phenotypes of F'240 result directly from an altered conformation of *araC* activator protein produced by the substituted amino acids. Suppression of nonsense mutations in the *araC* gene which lie outside the two main regions of constitutivity does not produce a constitutive phenotype (Table 4). An *araC*$^+$ gene in the various suppressor strains in all cases has the C$^+$ phenotype: a low basal level, arabinose-inducible and fucose-sensitive. This rules out the possibility that the constitutive phenotypes of F'240 could be due to alteration of either the *araC* or *araA* products due to the addition in the suppressor strains of an additional amino acid onto the normal carboxy terminal of the protein.

Such analysis of the effects of substituting known amino acids at specific sites in a protein can be valuable in determining the particular features in a protein which confer its structural and functional properties. For example, Weber *et al.* (1972), by analyzing the effects of various suppressors on a particular nonsense mutation in the *lacI* gene, were able to eliminate at least one of the interactions postulated by Adler *et al.* (1972) to occur between *lac* repressor and *lac* operator DNA. Detailed studies of the effects of various amino acid substitutions in the *araC* activator protein may similarly provide important clues into its structure and biological mechanism of action.

We thank Don Dean, Jeff Hall and Ross Feldberg for criticisms of the manuscript.

This work was supported by Public Health Service research grant GM-18277, Career Development Award K4-GM-38797, and Training Grant GM-212, from the National Institute of General Medical Sciences, National Institute of Health. This paper is publication no. 1021 from the Department of Biochemistry, Brandeis University.

## REFERENCES

Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M. & Schmitz, A. (1972). *Nature (London)*, **237**, 322–327.

Amati, P. & Meselson, M. (1965). *Genetics*, **51**, 369–379.

Benzer, S. (1961). *Proc. Nat. Acad. Sci., U.S.A.* **47**, 403–415.

Beverin, S., Sheppard, D. E. & Parks, S. S. (1971). *J. Bacteriol.* **107**, 79–86.

Chan, T. S. & Garen, A. (1969). *J. Mol. Biol.* **45**, 545–548.

Doyle, M. E., Brown, C., Hogg, R. W. & Helling, R. B. (1972). *J. Bacteriol.* **110**, 56–65.

Drapeau, G. R., Brammar, W. J. & Yanofsky, C. (1968). *J. Mol. Biol.* **35**, 357–367.

Englesberg, E. (1971). In *Metabolic Pathways* (Vogel, H., ed.) 3rd edtn, vol. 5, pp. 267–296, Academic Press, New York.

Englesberg, E., Irr, J., Power, J. & Lee, N. (1965). *J. Bacteriol.* **90**, 946–957.

Englesberg, E., Sheppard, D., Squires, C. & Meronk, F., Jr (1969). *J. Mol. Biol.* **43**, 281–298.

Gallucci, E. & Garen, A. (1966). *J. Mol. Biol.* **15**, 193–200.

Garen, A., Garen, S. & Wilhelm, R. C. (1965). *J. Mol. Biol.* **14**, 167–178.

Gielow, L., Largen, M. & Englesberg, E. (1971). *Genetics*, **69**, 289–302.

Greenblatt, J. & Schlief, R. (1971). *Nature New Biol.* **233**, 166–170.

Irr, J. & Englesberg, E. (1970). *Genetics*, **65**, 27–39.

Katz, L. (1970). *J. Bacteriol.* **102**, 593–595.

Lis, J. T. & Schleif, R. (1973). *J. Mol. Biol.* **79**, 149–162.

Lu, P. & Rich, A. (1971). *J. Mol. Biol.* **58**, 513–531.

Luria, S. & Delbrück, M. (1943). *Genetics*, **28**, 491–511.

Messer, A. (1974). *J. Bacteriol.* **120**, 266–272.
Miller, J. (1972). In *Experiments in Molecular Genetics*, p. 14, Cold Spring Harbor Press, New York.
Pfahl, M. (1972). *Genetics*, **72**, 393–410.
Pfahl, M., Stockter, C. & Gronenborn, B. (1974). *Genetics*, **76**, 669–679.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerly, R. P. & Saul, F. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 3305–3310.
Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerly, R. P. & Saul, F. (1974). *Proc. Nat. Acad. Sci., U.S.A.* **71**, 3440–3444.
Sadler, J. & Smith, T. (1971). *J. Mol. Biol.* **62**, 139–169.
Schleif, R. (1969). *J. Mol. Biol.* **46**, 185–196.
Schleif, R. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3479–3484.
Schleif, R. & Lis, J. (1975). *J. Mol. Biol.* in the press.
Sheppard, D. & Englesberg, E. (1967). *J. Mol. Biol.* **25**, 443–454.
Singer, S. J., Martin, N. & Thorpe, N. (1971). *Ann. N.Y. Acad. Sci.* **190**, 342–351.
Weber, K., Platt, T., Ganem, D. & Miller, J. H. (1972). *Proc. Nat. Acad. Sci., U.S.A.* **69**, 3624–3628.
Wu, T. T. & Kabat, E. A. (1970). *J. Exp. Med.* **132**, 211–250.
Yanofsky, C. & Horn, V. (1972). *J. Biol. Chem.* **247**, 4494–4498.
Yanofsky, C., Carlton, B. C., Guest, J. R., Helsinki, D. R. & Henning, U. (1964). *Proc. Nat. Acad. Sci., U.S.A.* **51**, 266–272.
Yanofsky, C., Drapeau, R., Guest, J. R. & Carlton, B. C. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 296–298.