Proc. Nat. Acad. Sci. USA
Vol. 69, No. 12, pp. 3624–3628, December 1972

Altered Sequences Changing the Operator-Binding Properties of the *Lac* Repressor: Colinearity of the Repressor Protein with the *i*-Gene Map

(amino-acid substitutions/amino-terminal/proteolysis/mutant/*E. coli*/suppressor)

KLAUS WEBER, TERRY PLATT*, DON GANEM, AND JEFFREY H. MILLER†

Biological Laboratories, Harvard University, Cambridge, Massachusetts, 02138

Communicated by Charles Yanofsky, October 5, 1972

ABSTRACT **A technique is described for mapping point mutations in the first 59 amino-acid residues of the *lac* repressor from *Escherichia coli*, using less than 0.1 μmol (4 mg) of the purified protein. This technique was used to localize five mutations affecting the ability of the *i*-gene product to repress *in vivo*. These alterations are located at four different sites in the amino-terminal region of the repressor molecule. Three of these are missense mutations and result in changes from serine to proline (residue 16), threonine to alanine (residue 19), and alanine to valine (residue 53). Each amino-acid substitution alone is sufficient to eliminate repression *in vivo*, presumably by altering the operator binding activity. The remaining two independently-isolated mutations are identical, and result in a change from a glutamine codon at position 26 to an amber (UAG) codon. Since suppression of this nonsense mutation with amber suppressors that insert leucine, tyrosine, serine, or glutamine restores repressor activity to the molecule, glutamine$_{26}$ cannot be crucial for the operator-binding function. A comparison of the position of each altered residue with the genetic map enabled us to estimate the physical distance between several deletion-group endpoints.**

Mutants of *Escherichia coli* that are deficient in the ability to repress expression of the lactose operon are called *lac* constitutives (1). In general, such mutants have either (*a*) a defective *lac* repressor or (*b*) an altered *lac* operator [the site on the DNA to which the repressor binds (2)]. An alteration in the operator site that decreases the efficiency of repression is termed an operator-constitutive (o^c) mutation. Such mutations weaken the repressor–DNA interaction by lowering the affinity between the repressor and the changed operator site (2, 3). The ability of the repressor to recognize the operator site may also be affected by mutations that alter the primary structure of the repressor itself; these mutations also result in high rates of *lac* enzyme synthesis.

Although studies on the mechanism of repression must eventually include a biochemical analysis of both types of mutation, at present a detailed investigation of i^- mutations appears more feasible. However, the general class of i^- mutations will exhibit multiple effects on the properties of the repressor because they result in incomplete, defective, or unstable polypeptide chains; therefore, the ability to bind inducer and/or operator, and also the ability to form tetramers will be jointly affected. Fortunately, there is a class of mutations in the *i*-gene that eliminates the ability of the *i*-gene product to repress *in vivo*, presumably by only altering the operator-binding activity. These mutations result in negatively-complementing repressor molecules that appear normal in all other respects. Such a mutation has been called an "i^{-d}" because it is *trans*-dominant in a heterozygous (i^+/i^{-d}) diploid; the diploid has a partial i^- phenotype. Since the wild-type *lac* repressor is a tetramer of identical subunits, the phenomenon of negative complementation is presumably due to the formation of mixed oligomers *in vivo* in which the mutant (nonrepressing) subunits are dominant (4).

A collection of missense point mutations of the i^{-d} class has been isolated and mapped (5, 6). Those mutants isolated thus far occur near the left-hand (5′) end of the *i*-gene (5, 6), which corresponds to the amino-terminal end of the *lac* repressor (7, 8). We have shown that some nonsense mutations also exhibit an i^{-d} character, due to internal reinitiation of translation after the amber block to produce a fragment of the *lac* repressor. In one case (i^{100}), the fragment lacks the first 42 amino-acid residues. The reinitiated i^{100} repressor fragment appears normal in all respects except in the ability to repress (9). The genetic mapping results and the biochemical characterization of the i^{100} repressor have supported the hypothesis that the amino-terminal region of the *lac* repressor is either directly involved in binding to the *lac* operator, or is required for the correct operator-binding conformation of the molecule. Since the amino-terminal sequence of the *lac* repressor is known for at least the first 59 residues (5, 9, 10), precise localization of i^{-d} mutants should be possible if a chemical protein mapping procedure is available to study specifically the amino-terminal part of the molecule. Such a procedure would also be of great value in the analysis of internal reinitiation signals, since two of these signals have already been found in the amino-terminal part of the *i*-gene (9).

RESULTS

The five mutations that we have localized (as described below) all exhibit i^{-d} character. Two of these mutations (100 and 738) were obtained spontaneously, and three (AP309, AP432, and AP46) were isolated after mutagenesis with 2-aminopurine (6). The phenotype corresponding to i^- was selected on minimal medium containing phenyl-β-D-galactoside as a carbon source (11), and the resulting candidates were screened for the ability to negatively complement a wild-type

Abbreviation: IPTG, isopropyl-β-D-thiogalactoside.

* Present address: Department of Biological Sciences, Stanford University, Stanford, Calif., 94305.
† Present address: Department de Biologie Moleculaire, Universite de Geneve, Geneva, Switzerland.

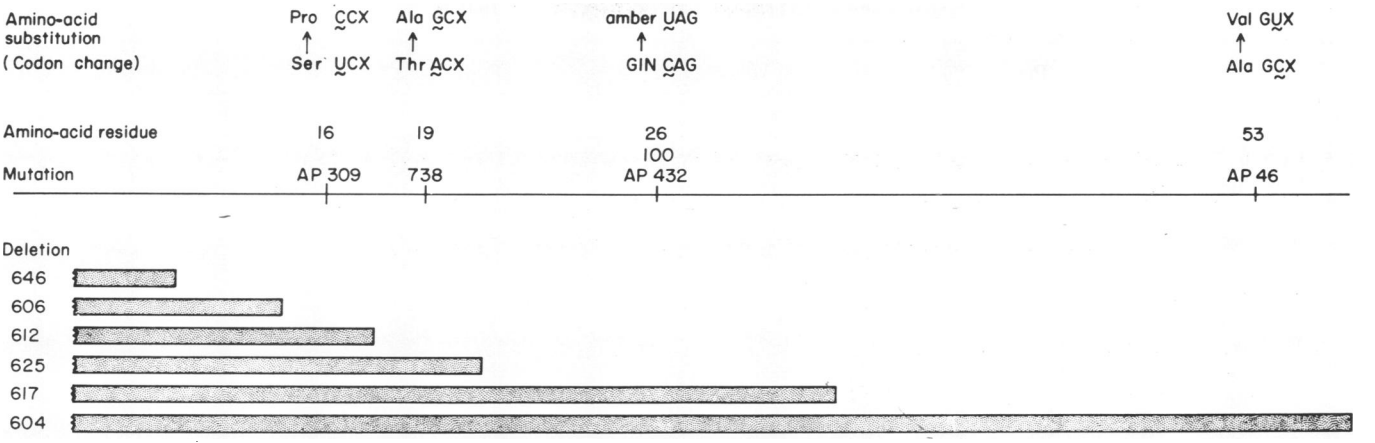

FIG. 1. Correlation between genetic and biochemical mapping results. The five mutations whose positions have been determined are shown on the central line, at intervals proportional to their location in the *i*-gene. The mutations AP309, 738, and AP46 are missense mutations; 100 and AP432 are amber mutations. The deletions that separate the mutations are indicated by the *bars* below the line (646, 606, 612, 625, 617, and 604). Details of the mapping procedure are described by Ganem (6). The techniques used to locate the amino-acid changes are given in the text. All of the observed changes can be accounted for by single base transitions, as indicated in the figure. The full sequence in the amino-terminal region of the *lac* repressor polypeptide chain is as follows (10):

```
1          5          10         15         20         25         30
Met-Lys-Pro-Val-Thr-Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-Ser-Tyr-Gln-Thr-Val-Ser-Arg-Val-Val-Asn-Gln-Ala-Ser-His-Val-
              35         40         45         50         55
Ser-Ala-Lys-Thr-Arg-Glu-Lys-Val-Glu-Ala-Ala-Met-Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Asn-Arg-Val-Ala-Gln-Gln-Leu-Ala-Gly-Lys-.
```

i-gene in a *recA*$^-$ diploid. Depletion mapping of these five mutations (see refs. 9 and 12 for a deletion map of the entire *i*-gene) places them all in the early portion of the *i*-gene. Fine-structure mapping near the 5′ end of the gene allows the five mutations to be separated into deletion groups, as illustrated in Fig. 1. The amber mutations 100 and AP432 are indistinguishable from one another (although they were derived independently), and we shall show below that they occur at the same site, which corresponds to glutamine$_{26}$ in the primary structure of the repressor polypeptide chain. We have shown (9) that an amber codon at this position activates a reinitiation site at the codon corresponding to methionine$_{42}$, giving rise to a negatively-complementing fragment of repressor. This result explains why these two mutations have i^{-d} character, although they are nonsense rather than missense mutations. Suppression of these two amber mutations by amber suppressors that insert leucine, tyrosine, serine, or glutamine restores wild-type repressor activity to the molecule (Table 1). To locate the mutations, we chose to suppress 100 and AP432 with the leucine-inserting amber suppressor suVI (obtained from A. Garen), since there is only one naturally occurring leucine residue (at position 6) in the first 44 residues of the wild-type molecule (5, 9, 10).

The mutations were originally isolated on an F′ *lac-pro* episome in a host containing a *lac-pro* deletion; the episomal *i*-gene also carried the i^Q marker, which results in a 10-fold overproduction of the *lac* repressor (4). In order to purify reasonable amounts of the altered proteins, the mutations were crossed onto a temperature-inducible lysis-defective *lac* prophage (also carrying the i^Q marker); induction of the prophage yields a 20-fold increase in the amount of repressor, such that repressor comprises about 0.2% of the soluble protein in the cell (4). Each of the mutant repressors was purified from about 200 g of induced cells by ammonium sulfate fractionation and phosphocellulose chromatography, as described elsewhere (10). Since all five mutant repressors still possessed the capacity to bind inducer, it was possible to follow the activity throughout the purification. Each protein was checked for homogeneity by electrophoresis on sodium dodecyl sulfate–polyacrylamide gels (14) before being subjected to subsequent steps. Sucrose gradient centrifugation of each of the missense repressors demonstrated that they are in the tetrameric form, although they are unable to repress *in vivo*.

A 4-mg aliquot of each mutant repressor was dialyzed at 4° against 0.1 M ammonium bicarbonate, and digested with trypsin for 20–30 min at 37° (10). Under these conditions, the peptides accounting for residues 1–59 in the wild-type sequence are released from the amino-terminus of the native molecule. If this digest is passed through a Sephadex G-50

TABLE 1. *In vivo repression by amber 100 and suppressed derivatives*

Strain	Relative specific activity of β-galactosidase	
	−IPTG	+IPTG
i^Q/Su$^-$	2.2	98
i^-100/Su$^-$	160	153
i^-100/Su6$^+$	3.3	168

All *i*-gene mutations were present on a *lac proA,B* episome carrying the i^Q mutation and the *lac* promoter mutation L8. All were assayed in an F$^-$ (*lac pro*) deletion strain or its isogenic Su6$^+$ derivative (prepared by transduction with P1$_{vir}$). Each value given is the average of two independent, simultaneous determinations [β-galactosidase assays and units of specific activity as described by Miller (13)].

Similar results have been obtained with Su1, Su2, and Su3 suppressed derivatives (unpublished experiments).

TABLE 2. *Peptide composition used in mutant analysis*

	Peptide "*b*"		Peptide "*d*"		Peptide "*E*"		Peptide "*C*"	
	AP309	wild-type	738	wild-type	AP 432 or 100	wild-type	AP46	wild-type
Lysine	—	—	—	—	1	1	1	1
Histidine	—	—	—	—	1	1	—	—
Arginine	—	—	1	1	—	—	—	—
Asparagine	—	—	—	—	1	1	—	—
Threonine	—	—	—	1	—	—	—	—
Serine	—	1	1	1	2	2	—	—
Glutamine	—	—	1	1	—	1	2	2
Proline	1	—	—	—	—	—	—	—
Glycine	1	1	—	—	—	—	1	1
Alanine	1	1	1	—	2	2	1	2
Valine	1	1	1	1	3	3	2	1
Leucine	—	—	—	—	1	—	1	1
Tyrosine	1	1	—	—	—	—	—	—

Amino-acid compositions of the peptides that contain the amino-acid substitution in mutants AP309, 738, and AP46 and in the leucine-suppressed amber mutants 100 and AP432. The peptides were eluted from the fingerprints shown in Fig. 3*a* and *b*. Amides were assigned by the electrophoretic mobility and by digestion with leucine aminopeptidase (10). For peptide *C* of the mutant AP46, the point of alteration was found by Edman degradation (see text).

column, the profile shown in Fig. 2 is obtained. The amino-terminal tryptic peptide of the *lac* repressor (residues 1–22) elutes at peak *B* in the figure. The fractions comprising this peak were pooled, lyophilized, and further digested with chymotrypsin. These chymotryptic peptides were separated on Whatman 3 MM paper by electrophoresis and chromatography (Fig. 3*a*). The four peptides eluted (given in the legend to Fig. 3) account for residues 1–7 ("*c*"), 8–13 ("*a*"), 13–17 ("*b*"), and 18–22 ("*d*") of the repressor polypeptide chain. This procedure allowed us to locate the amino-acid changes resulting from the two earliest mutations, AP309 and 738. The chymotryptic fingerprints obtained were similar to those shown by the wild type. No significant mobility changes were apparent.

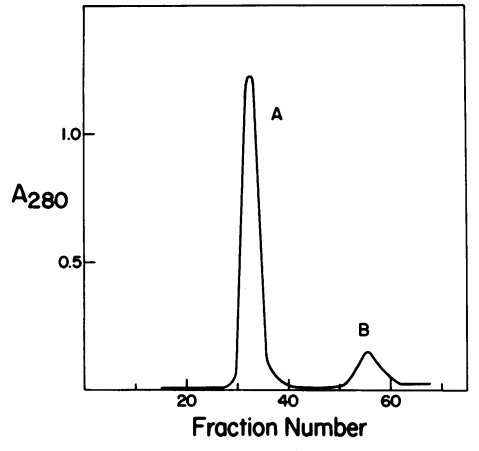

FIG. 2. Gel filtration of a digest of repressor (4 mg) with trypsin under native conditions. The sample was passed through a Sephadex G-50 column (1.2 × 20 cm) equilibrated with 0.1 M ammonium bicarbonate at a flow rate of 15 ml/hr; 0.4-ml fractions were collected. The column profile shows the absorption at 280 nm. Peak *A* represents the trypsin-resistant repressor core molecule, peak *B* corresponds to the amino-terminal tryptic peptide (residues 1–22), which contains three tyrosine residues (10).

Therefore, all peptides were eluted and characterized by amino-acid composition. In the case of the mutant AP309, peptide "*b*" showed a replacement of the serine in position 16 by a proline residue (Table 2 and Fig. 1). With mutant 738, only peptide "*d*" differed from the wild type. The threonine residue in position 19 was replaced by an alanine residue (Table 2 and Fig. 1).

A slightly different procedure was used to analyze amino-acid changes occurring between residues 23 and 59 (10). After tryptic digestion under native conditions, the sample was lyophilized and extracted directly with pH 3.5 electrophoresis buffer. The soluble peptides were separated in two dimensions as described above; the fingerprint obtained after staining with ninhydrin is shown in Fig. 3*b*. The four tryptic peptides of wild-type repressor that account for residues 23–59 are given in the legend to Fig. 3*b*. The missense mutation AP46 and the two amber mutations 100 and AP432 (suppressed with a leucine suppressor) were localized by this procedure.

Tryptic fingerprints of leucine-suppressed repressor from mutants 100 and AP432 showed that one peptide ("*E*") has an increased chromatographic mobility. Amino-acid composition analysis of peptide "*E*" showed the replacement of a glutamine by a leucine residue in position 26 (Table 2, Fig. 1). Thus, the amber mutations 100 and AP432 must occur at the position corresponding to residue 26 in the wild-type amino-acid sequence.

We used the same procedure to locate the mutation AP46, but no mobility changes were observed on the fingerprint. The amino-acid composition (Table 2) of peptide "*C*", however, revealed the substitution of an alanine residue by a valine residue. Since this peptide normally contains two residues of alanine, Edman degradation was used to localize the amino-acid change, which was found to occur in position 53 (Fig. 1).

DISCUSSION

We have developed a simple fingerprint analysis for *lac* repressor mutants in the amino-terminal part of the polypeptide chain. The technique is based on our previous finding (10),

that digestion of *lac* repressor under certain conditions yields a limited number of tryptic peptides, which can all be assigned to the amino-terminal part of the molecule. The material requirement for the mapping is small, and we have found 0.1 μmol of repressor (4 mg) to be quite sufficient in the analysis of five different mutants. Since most DNA-binding mutants are found in the amino-terminal region (5), the fingerprint technique used in this study should allow the characterization of this important group of *lac* repressor mutants.

Five different mutants have been characterized. The mutations result in a *lac* repressor that is nonfunctional with respect to operator binding. The mutations fall into two groups: three (AP309, 738, and AP46) are missense mutations that result in an amino-acid substitution. Two mutants (100 and AP432) are nonsense mutations that activate reinitiation of translation at the codon corresponding to methionine$_{42}$, and produce a repressor fragment lacking its amino-terminal 42 amino acids (9). In all cases the mutant repressors cannot repress *in vivo*, but they have not lost inducer binding ability and the capacity to form a tetramer. We have previously isolated a tryptic core with a polypeptide chain molecular weight of about 28,000, missing 80–100 amino acids at the amino terminus. This *in vitro* repressor derivative forms a tetrameric structure and shows the same binding of isopropyl-β-D-thiogalactoside (IPTG) as the wild-type repressor (10). Taken together, the results clearly indicate that the amino-terminal part of the molecule is not a necessary requirement for IPTG binding or for the ability to associate into a tetramer. The *in vitro* results when trypsin susceptibility is used as a probe of the structure of *lac* repressor argue that the amino-terminal part can be readily digested by the protease, and may be localized on the surface of the molecule. This interpretation supports the hypothesis of Adler *et al.*, who have assumed that the DNA-binding part of the *lac* repressor forms a protuberance (5).

There is a significant difference in the functional effect between the missense mutations and the suppressed amber mutations. In all three missense substitutions, a single amino-acid change destroys the operator-binding activity. This result seems especially surprising in the case of the conservative alanine to valine change at residue 53 (mutant AP46). On the other hand, when the amber codon resulting from mutation 100 or AP432 is suppressed by the insertion of leucine, tyrosine, serine, or glutamine (the original amino acid at that locus), the operator-binding function is restored (Table 1). Thus, amino-acid alterations at positions 16, 19, and 53 are sufficient to eliminate operator-binding activity, but at least four different amino acids are tolerated at position 26 with no apparent effect on any of the functional properties of the *lac* repressor.

These results are interesting in view of the specific model for repressor–DNA interaction proposed by Adler *et al.* (5). They have argued that the amino-terminal part of the molecule may be built into an α helix, which in turn can be arranged in a specific way to form hydrogen bonds and salt bridges to the DNA. Using model building, they predict a specific operator sequence based on the assumption that the five amino acids in positions 17, 18, 21, 25, and 26 of the polypeptide chain specify operator recognition by forming hydrogen bonds to the DNA. It is interesting to note that the amber mutants are in position 26, and that the repressor chain can tolerate four different amino-acid residues in the position. Since one of the residues is leucine, one has to conclude that the hypothetical contact of Gln$_{26}$ in the model of Adler *et al.* (5) is fortuitous, and is not necessary for repressor activity. Thus, if their particular model is true, the specificity must reside with the remaining four amino acids (residues17, 18, 21, and 25). Even in this case, it is surprising that an amino-acid

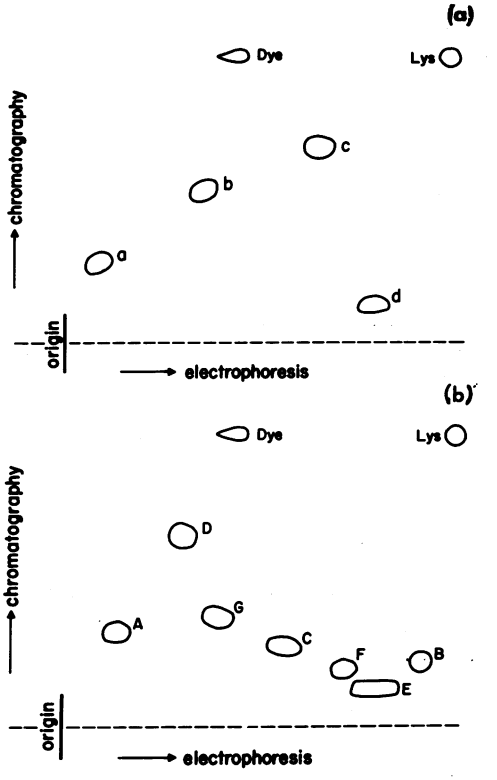

FIG. 3. Fingerprints of peptides released from intact *lac* repressor by digestion under native conditions (10). (*a*) Chymotryptic peptides spanning residues 1–22 of the repressor polypeptide chain. Peak *B* of the column profile shown in Fig. 2 was digested with chymotrypsin, and the peptides were separated on Whatman 3 MM paper by electrophoresis at pH 3.5 (5% acetic acid–0.5% pyridine) in the first dimension, followed by ascending chromatography (24.4% pyridine–37.8% *n*-butyl alcohol–7.6% acetic acid–30.2% water) in the second dimension. The peptides are:

c (residues 1–7): Met-Lys-Pro-Val-Thr-Leu-Tyr;
a (residues 8–12): Asp-Val-Ala-Glu-Tyr;
b (residues 13–17): Ala-Gly-Val-Ser-Tyr;
d (residues 18–22): Gln-Thr-Val-Ser-Arg.

(*b*) Fingerprints of the soluble tryptic peptides obtained from a digest of repressor under native conditions. The peptides were extracted with the electrophoresis buffer. The peptides are:

E (residues 23–33): Val-Val-Asn-Gln-Ala-Ser-His-Val-Ser-Ala-Lys:
B (residues 34–35): Thr-Arg:
G (residues 36–51): Glu-Lys-Val-Glu-Ala-Ala-Met-Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Asn-Arg:
C (residues 52–59): Val-Ala-Gln-Gln-Leu-Ala-Gly-Lys.

We reported the detailed sequence analysis previously (10). The peptides account for the first 59 residues of *lac* repressor. A similar sequence has been reported by Adler *et al.* (5). The peptides *D*, *F*, and *A* have been subjected to amino-acid sequence analysis (10): however, their exact location in the polypeptide chain is not known.

residue so close to the DNA in the hydrogen-bonding model of Adler *et al.* (5) can show such varied changes that do not interfere with repressor function, including changing of a hydrogen-bond acceptor to a hydrogen-bond donor. This result is even more surprising in view of the fact that the more subtle amino-acid replacements in mutants 738 and AP46 abolish repressor activity.

Three of the five mutants studied were obtained by mutagenesis with 2-aminopurine. In all cases, the amino-acid substitution (Fig. 1) was consistent with the previous observation that 2-aminopurine is specific for transition mutations (15). Comparison of the genetic and protein chemical map (Fig. 1) shows that the distances between the end points of the different deletions cutting into the *i*-gene vary. It is apparent, for instance, that the interval between deletions 612 and 625 is very short, perhaps comprising as few as 10 nucleotides, whereas the intervals between deletions 625 and 617 and 617 and 604 might be considerably larger. More precise location of the respective deletion end groups can be obtained by a study of more repressor mutants.

We thank Charlotte Hering for her expert technical assistance in purifying the mutant repressors, and Magnus Pfahl for initially mapping the mutant 738 and reporting this result to us. Terry Platt was supported by an NSF Predoctoral Fellowship. Additional support was provided by NIH Grants GM 17662-01 and GM 059541-10.

1. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356.
2. Gilbert, W. & Müller-Hill, B. (1967) *Proc. Nat. Acad. Sci. USA* **58**, 2415–2421.
3. Riggs, A. D., Bourgeois, S., Newby, R. F. & Cohn, M. (1968) *J. Mol. Biol.* **34**, 365–368.
4. Müller-Hill, B., Crapo, L. & Gilbert, W. (1968) *Proc. Nat. Acad. Sci. USA* **59**, 1259–1264.
5. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M. & Schmitz, A. (1972) *Nature New Biol.* **237**, 322–327.
6. Ganem, D. E. (1972) Honors Thesis, Biochemical Sciences, Harvard College.
7. Miller, J. H., Beckwith, J. R. & Müller-Hill, B. (1968) *Nature* **220**, 1287–1290.
8. Kumar, S. & Szybalski, W. (1969) *J. Mol. Biol.* **40**, 145–151.
9. Platt, T., Weber, K., Ganem, D. & Miller, J. H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 897–901.
10. Platt, T., Files, J. & Weber, K. (1972) *J. Biol. Chem.* **247**, in press.
11. Smith, T. F. & Sadler, J. R. (1971) *J. Mol. Biol.* **59**, 273–305.
12. Pfahl, M. (1972) *Genetics*, in press.
13. Miller, J. H. (1972) in *Experiments in Molecular Genetics*, (Cold Spring Harbor Press, Cold Spring Harbor, N.Y.), pp. 352–355.
14. Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244**, 4406–4412.
15. Yanofsky, C., Ito, J. & Horn, V. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 151–162.