

## Translational Reinitiation: Reinitiation of *lac* Repressor Fragments at Three Internal Sites Early in the *lac i* gene of *Escherichia coli*\*

(initiation codons/antibody/amber mutants/mRNA secondary structure)

JAMES G. FILES, KLAUS WEBER, AND JEFFREY H. MILLER†

Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138; and † Department of Molecular Biology, University of Geneva, Geneva, Switzerland

Communicated by Bernard D. Davis, November 1, 1973.

**ABSTRACT** Three early amber mutations in the *lac i* gene have been shown to arise from the codons corresponding to residues 7, 12, and 17 of the *lac* repressor polypeptide chain. All three mutations allow translational reinitiation at the same two sites, resulting in the synthesis of two *lac* repressor fragments. The amino-terminal sequences of these fragments show that the first site is the triplet coding for valine residue 23, while the second is the first internal in-phase AUG codon corresponding to residue 42.

Translational reinitiation appears to be a common event in *E. coli*, since there are at least three such sites in the first 70 in-phase codons of the *i*-gene messenger RNA, and all amber mutants found in this region show translational reinitiation. Only one of these sites involves an AUG codon; the other two involve an *in vivo* ambiguity of the genetic code, in that the same codon can be translated into two different amino acids depending on whether it is recognized during initiation or elongation of protein biosynthesis. The two non-AUG codons are the codons corresponding to leucine residue 62 and valine residue 23 of the *lac* repressor.

The most successful approach so far to understanding the specificity of the initiation process of protein biosynthesis has been the isolation and sequence analysis of ribosomal binding sites of natural mRNA molecules. Such studies have clearly demonstrated the exclusive use of AUG as the initiator codon in all eight currently known initiation sequences (1-6).† In contrast, earlier experiments on protein biosynthesis *in vitro* suggested that not only AUG but also GUG, and to some extent UUG and GUA, could function as initiator codons (7-12).

Another approach to evaluating initiation sequences makes use of the phenomenon of reinitiation of protein synthesis after chain termination at nonsense codons. Such internal initiation sites have been described for the rIIB cistron of bacteriophage T4 (13) and the *z* gene (14, 15) and the *i* gene (16, 17) of the lactose operon of *Escherichia coli*. If the amino-acid sequence of the wild-type protein and the amino-terminal sequence of the "restart" fragment are known, one can deduce the nature of the initiation codon for the translational restart. The genetics of the *i* gene and the protein

Abbreviations: SDS, sodium dodecyl sulfate; dansyl, 1-dimethylaminonaphthalene-5-sulfonyl; IPTG, isopropyl- $\beta$ -D-thiogalactoside.

\* This is no. III in the series "Translational Reinitiation." Ref. 17 is no. II.

† After this manuscript was prepared, we learned that GUG has been proposed as the starting codon for the maturation protein of bacteriophage MS2 (28).

chemistry of its product, the *lac* repressor, have advanced to a level that permits such an analysis.

We have previously shown (16) that the amber mutant 100, in the codon for residue 26 of the *lac* repressor polypeptide chain (a glutamine), exhibits translational reinitiation at the first internal in-phase AUG codon, corresponding to methionine, residue 42 (site  $\pi_{42}$ ). Moving the amber block beyond  $\pi_{42}$ , as in mutant 136, activates a new reinitiation site,  $\pi_{62}$ , which we have located at the leucine codon responsible for residue 62 (17). The finding that a triplet can code in elongation for a different amino acid (leucine) than it does in initiation (*N*-formylmethionine) raises the question whether those amber mutants mapping before 100 use  $\pi_{42}$  exclusively or if there is another potential initiation site earlier in the gene involving a codon other than AUG.

The results presented here demonstrate that there is a reinitiation site utilizing a codon other than AUG before  $\pi_{42}$ , at the codon specifying the valine residue 23 ( $\pi_{23}$ ). Three early amber mutations, mapping before this site, activate both the  $\pi_{23}$  and the  $\pi_{42}$  reinitiation sites. During elongation this codon specifies valine (residue 23), but in reinitiation it seems to code for *N*-formylmethionine. This is the first time that a valine codon, most likely GUG, has been shown to be an *in vivo* initiator of protein synthesis in a bacterial message.

### RESULTS

Three early *i*-gene mutations, M1, M2, and 26, generate restart fragments. These fragments have been characterized by described protein chemical procedures (16). In each case the reinitiated *lac* repressor fragments were concentrated from cell extracts by ammonium sulfate precipitation and purified by precipitation with antibody against *lac* repressor followed by electrophoresis of the antigen-antibody complex on Na dodecyl sulfate (SDS)-polyacrylamide gels (for details see Fig. 1). The gels showed two closely spaced protein bands of restart material with molecular weights of 34,000 and 36,000, obtained in approximately equal amounts (arrows). The two restart species are not very well resolved on preparative gels and were therefore cut out and eluted from the gel together and subjected to the modified dansyl (1-dimethylaminonaphthalene-5-sulfonyl)-Edman procedure. This procedure has been described in detail (18); for M1 the first four residues at the amino terminus gave (Ala,Val)-(Glx,Asx)-(Leu,Glx)-(Asx,Ala).

Comparison of this sequence with the previously identified restarts (refs. 16 and 17; see also Fig. 2) shows that it could



FIG. 1. *Lac* repressor restart polypeptides found in amber mutant M1. The gel shown is a guidestrip cut from a preparative gel on which an antibody-antigen complex was separated. The guidestrip was stained with Coomassie brilliant blue and destained as described (16, 18). The isopropyl- $\beta$ -D-thiogalactoside (IPTG)-binding material of cell extracts was concentrated by an initial 30% ammonium sulfate precipitation and then further purified by precipitation with antibody against wild-type *lac* repressor. After overnight incubation with the antibody at 4°, the insoluble material was collected by centrifugation and washed carefully with buffer solution. After three washing and centrifugation cycles, the antigen-antibody complex was dissociated with SDS and 2-mercaptoethanol at 95° for 3 min, and the polypeptides were separated on SDS-polyacrylamide gels [for details see Platt *et al.* (16)]. The bands are, in order from the left: immunoglobulin G heavy chain (molecular weight 50,000), M1 reinitiation protein  $\pi_{23}$  (molecular weight 36,000), M1 reinitiation protein  $\pi_{42}$  (molecular weight 34,000), and immunoglobulin G light chain (molecular weight 23,000).

contain the  $\pi_{42}$  restart sequence Ala-Glx-Leu-Asx-, which is activated by amber mutant 100 (16). Such an assignment is supported by the molecular weight of 34,000, which is identical to that reported for the  $\pi_{42}$  fragment. By difference, the larger species (36,000 molecular weight) would then have the amino-terminal sequence Val-Asx-Glx-Ala. There is only one such sequence in the first seventy residues of the wild-type repressor, occurring at residues 24-27 (Fig. 2).<sup>§</sup> A fragment starting at position 24 should have a molecular weight of approximately 36,000, since the wild-type protein has a molecular weight of 38,000 (16).

To verify this assignment, we determined the amino-terminal residue of each of the two restart fragments. Since direct amino-terminal analysis requires only small amounts of protein (18), the resolution on 12-cm polyacrylamide gels is sufficient to detect separation of the two proteins on guidestrips. The protein of 36,000 molecular weight gave amino-terminal valine and the protein of 34,000 molecular weight gave alanine. Further proof for the identification of the two restart fragments was obtained by direct dansylation of the repressor-antibody complex in SDS. The dansylated protein mixture, purified by acetone precipitation, was separated on 10% polyacrylamide gels in SDS. The two well-separated bands of *lac* repressor restart protein, visible directly under ultraviolet light, were each eluted separately from the gel. Amino-terminal identifications again gave valine for the larger restart protein and alanine for the smaller one.

Restart material isolated from mutants M2 and 26 was identical to that of M1, as judged by the amounts of the two protein species, their molecular weights, and their amino-terminal residues. In particular it should be noted that mutant 26, the earliest of the three amber mutants (Fig. 3), exhibited no additional earlier reinitiation site.

It was also possible to determine the position of the amber block in each of the three mutants M1, M2, and 26. This was done by protein chemical mapping of amino-acid substitutions in the mutant repressor molecules isolated from strains carrying amber suppressors. Su 1, a serine-inserting

suppressor, was used for 26, since 26 maps earlier than the first serine residue (Figs. 2 and 3). Su 6, a leucine-inserting suppressor, was used for M1 and M2 since these mutations were derived using a mutator gene, *mutT*, suggesting that the original mutation was from tyrosine to amber (22). We have documented that tryptic digestion of native *lac* repressor releases only a small number of peptides, most of which are from the amino-terminal portion of the molecules (21, 23). Peptide mapping is used to separate the tryptic peptides. Amino-acid sequence analysis of the pure peptides allows unambiguous identification of amino-acid replacements in the first 59 residues of the *lac* repressor. We have previously used this procedure to identify the location of six different mutations in the polypeptide chain (21). When suppressed 26 *lac* repressor was subjected to this procedure, tyrosine residue 7 was found to be replaced by serine. Tyrosine-to-leucine changes were found in residue 12 of the suppressed M1 repressor and in residue 17 of the suppressed M2 repressor. Further amino-acid sequence analysis showed that in the three suppressed amber mutants the amino-acid sequence between residues 18 and 33 is unchanged, indicating the new initiation codon ( $\pi_{23}$ ) has not been achieved by mutagenesis.

Because of the specificity of induction of mutations by *mutT*, i.e.,  $\frac{A}{T} \rightarrow \frac{C}{G}$  transversions (22), only UAU codons can be converted to amber by this mutagen. Therefore the tyrosine residues at positions 12 and 17 are specified by the codon UAU. However, the codon specifying tyrosine residue 7 can mutate to amber spontaneously (mutant 26) but has not been converted to amber by *mutT*; therefore, it is highly probable that the codon at this position is UAC.

Our recent data allowing comparison of the genetic map of the *i* gene and the amino-acid sequence of the *lac* repressor are shown in Fig. 3. The characterization of mutants 26, M1, and M2 described above is included. In addition, the early missense mutant 40 (isolated after ethylmethane sulfonate mutagenesis), has been characterized as a threonine-to-methionine change in amino-acid residue five. The localization of M1 shows that deletion 606 has its endpoint between the codons responsible for amino-acid residues 12 and 16. This amino-terminal deletion fuses the remainder of the *i* gene to the *trp* operon and produces negatively complementing *lac* repressor, suggesting the possibility of reinitiation (24). Isolation of *lac* repressor fragments from 606 shows the same two restart fragments observed with the three early amber mutants described above. This finding proves that  $\pi_{23}$  and  $\pi_{42}$  can also act as translational initiation sites in the polycistronic *trp-i* gene messenger RNA.

#### DISCUSSION

The three early amber mutations in the *i* gene described in this paper map at sites corresponding to amino-acid positions 7, 12, and 17 of *lac* repressor. All three amber mutations give rise in *su*<sup>-</sup> strains to the same two translational reinitiation *lac* repressor fragments. One of these has been previously described (16); the other is novel, having an amino-terminal sequence (Val-Asx-Glx-Ala) that corresponds to residues 24-27 of the wild-type polypeptide chain. These results suggest the presence of an internal initiation codon at or before residue 24, but past position 17 where one of the amber mutations (M2) activating the restart site has been localized. Wild-type repressor has no methionine residue in this region (Fig. 2) so we do not expect an AUG codon there. The three amber mutants cannot contain AUG codons cre-

<sup>§</sup> The full amino-acid sequence has become available (19) and shows that the amino-acid sequence Val-Asx-Glx-Ala occurs only once in the *lac* repressor sequence.

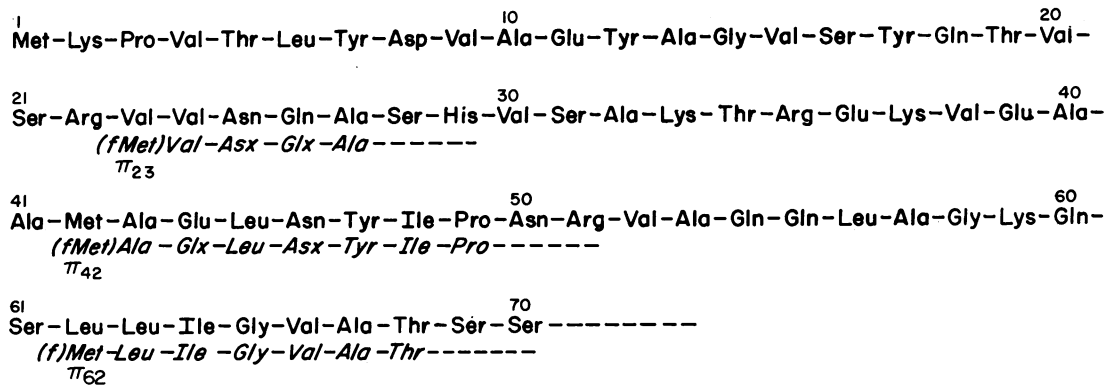


FIG. 2. Correlation between the amino-terminal sequences of the three restart fragments ( $\pi_{23}$ ,  $\pi_{42}$ , and  $\pi_{62}$ ) and the amino-acid sequence of the wild-type *lac* repressor. The restart protein sequences are given in *italics*. (*f*) and (*fMet*) indicate a formyl group and *N*-formylmethionine, respectively, not found on the purified restart protein but assumed to have been removed enzymatically *in vivo* after protein synthesis. The amino-acid sequence given for the first 70 residues of *lac* repressor has been reported previously by us (17), and the total amino-acid sequence reported recently by Beyreuther *et al.* (19) agrees with this sequence. The  $\pi_{42}$  and  $\pi_{62}$  restart fragments have been described for mutants 100 and 136, respectively (16, 17).

ated by mutagenesis in this region because the amino-acid sequence between residues 18 and 24 in suppressed ambers is unchanged from the wild type.

Since initiation of protein biosynthesis always occurs with *N*-formylmethionine (25), the simplest explanation for the amino-terminal sequence of this restart fragment is that the codon specifying valine residue 23 in the wild-type repressor acts as an internal reinitiation site ( $\pi_{23}$ ) and gives rise to a sequence *N*-formylMet-Val-Asx-Glx-Ala. The formyl group and the methionine residue would then be removed by secondary enzymatic activities (25), leaving an amino-terminal valine corresponding to residue 24 of the wild-type repressor sequence (Fig. 2). There is evidence that *N*-formylmethionine can be removed from the amino terminus of a newly synthesized *E. coli* protein when followed by a valine residue. Valine has been observed as a minor amino-terminal residue of *E. coli* proteins (26). Moreover, certain initiation mutants in the yeast isocytchrome *c* system produce a protein in which the initiating methionine residue is followed by valine, but 85% of the protein molecules have lost the methionine residue (27).

How can initiation of protein biosynthesis occur at a codon specifying a valine residue during elongation? There are two explanations, both of which involve an *in vivo* ambiguity in translation of this codon. The first model assumes that this codon is an AUG codon which, owing to contextual effects, can be misread by a valyl-tRNA during elongation. How-

ever, amino-acid sequence studies of the *lac* repressor indicate only valine for the amino-acid residue in position 23 (19, 23). A quantitative misreading of a hypothetical AUG codon by a valyl-tRNA is very unlikely. We therefore prefer a second model in which the initiation codon is indeed a valine codon that is recognized by *N*-formylmethionyl-tRNA (or another yet unidentified minor methionyl-tRNA) in initiation but by valyl-tRNA in elongation. Several investigators have shown that the valine codon GUG supports fMet-tRNA-dependent initiation of protein synthesis on synthetic polyribonucleotide templates *in vitro* (7, 10). Valine codon GUA has also been proposed to be a weak initiator *in vitro* (7, 10). Also, the use of GUG as the initiation codon for the A protein of the RNA bacteriophage MS2 has now been reported (28). Although the identification of the particular valine codon specifying residue 23 will have to await the characterization of an appropriate frameshift mutant, the codons GUG or GUA seem most likely.

The distribution of the early amber mutants in the *i* gene and of the translational reinitiation sites activated by these mutations is diagrammatically shown in Fig. 4. Amber mutants earlier than site  $\pi_{23}$ , as well as the deletion 606, activate the two restarts  $\pi_{23}$  and  $\pi_{42}$ . If the amber block is moved beyond  $\pi_{23}$ , as in 100, reinitiation at the first site,  $\pi_{23}$ , is abolished as expected and only the  $\pi_{42}$  restart is observed. Moving the amber beyond  $\pi_{42}$ , as in 136, leads to a loss of the  $\pi_{42}$  restart, and reinitiation occurs exclusively at a new site,  $\pi_{62}$  (17). The observation that some amber mutations activate

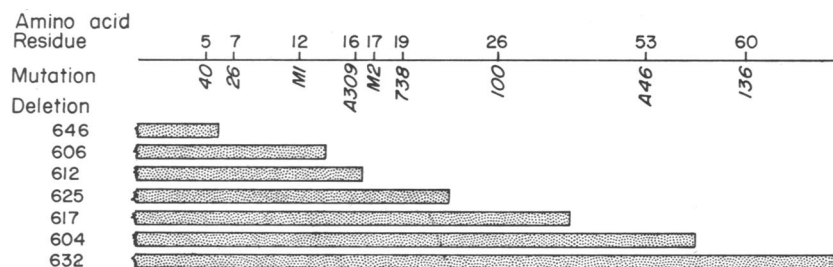


FIG. 3. Correlation between genetic and protein chemical mapping results. Nine mutations located early in the *i* gene have been mapped by a series of amino-terminal deletions. Details of the mapping procedure are described by Ganem (20). The mutations 40, A309, 738, and A46 are missense mutations. 26, M1, M2, 100, and 136 are amber mutations. Mutants A309, 738, 100, A46, and 136 have already been characterized (16, 17, 21). The positions of these mutations in the *lac* repressor protein are indicated by the amino-acid residue number above the *line*. Recombination tests show that deletion 612 ends between the two adjacent codons specifying amino-acid residues 16 and 17. Deletion 646 ends between the codons specifying amino-acid residues 5 and 7. Further details of the isolation of the three mutations (M1, M2, and 26) are in Miller, Ganem, and Lu (in preparation).

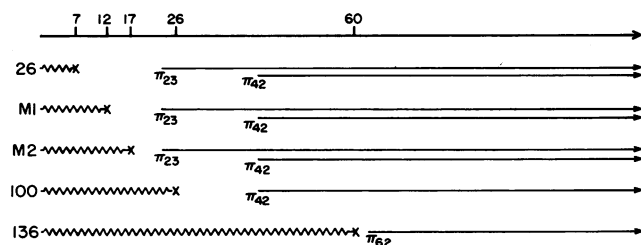


FIG. 4. Diagram of polypeptide chains synthesized in the early *i*-gene amber mutants 26, M1, M2, 100, and 136. ~ indicates amber fragments. These have no IPTG-binding activity and have not been detected so far. — indicates restart protein fragments. These fragments have IPTG-binding activity and are precipitable by antibody to wild-type *lac* repressor (16, 17, 23). X indicates the sites of chain termination at amber codons, and  $\pi$  indicates the sites of reinitiation of protein biosynthesis.

two restarts whereas others activate only one could be related to the secondary structure of the mRNA during protein synthesis. Certain internal initiation sites may be made available to a new ribosome only after translation has proceeded to an extent necessary to open secondary structure, thereby uncovering the initiation site. In such a model, secondary structure would interfere with internal reinitiation at  $\pi_{62}$  unless the amber block were moved sufficiently close to  $\pi_{62}$ , i.e., closer than the 100 block which activates  $\pi_{42}$  but not  $\pi_{62}$ . The simultaneous activation of  $\pi_{23}$  and  $\pi_{42}$  seen with the earliest amber mutants could suggest the absence of sufficient secondary structure between these two initiation sites.

Internal initiation does not occur at detectable levels in the wild-type *i* gene (16). This too might be explained by secondary structure of the mRNA. However, it is also possible that translation proceeds along the nascent mRNA involving the internal initiation sites in elongation of protein synthesis before they have any opportunity to act as initiation sites. The introduction of a nonsense codon will stop elongation before a restart site and make the site available for initiation of protein biosynthesis.

As is seen in Fig. 4, the distance between nonsense codon and initiation codon can vary greatly. The shortest distance is found in amber 136, where a single triplet separates the termination site from the  $\pi_{62}$  reinitiation site (17). The longest distance is found in amber mutant 26, in which the  $\pi_{42}$  reinitiation is separated by 105 nucleotides from the amber site. The data currently available are not sufficient to determine the importance of the distance between stop and start codons for the efficiency of reinitiation. All three early amber mutants have comparable amounts of inducer-binding activity, corresponding to approximately 10% of that of the wild type. However, a more quantitative study might reveal a subtle influence of the distance between stop and start codon on the efficiency of reinitiation.

The early part of the *i* gene shows three internal initiation signals. One of them,  $\pi_{42}$ , occurs at the only in-phase AUG codon in this part of the translated mRNA (16). The other two involve codons other than AUG:  $\pi_{23}$ , a valine codon, and as previously reported (17),  $\pi_{62}$ , a leucine codon. If these two codons should be GUG and UUG or CUG, the possibility arises that there may be considerable flexibility or ambiguity in recognition of the first base of the initiation codon by the initiator fMet-tRNA. However, such a hypothesis has to

await identification of the initiation codons with appropriate frameshift mutants.

Assuming three to four possible initiation codons and assuming that reinitiation at these codons is not strongly dependent on neighboring sequences, we might expect that one out of every 15–20 codons in a mRNA molecule can serve as an in-phase initiation site. We find in the *i*-gene message, reinitiation at three of the first 70 in-phase codons, at those specifying amino-acid residues 23, 42, and 62 of the *lac* repressor polypeptide chain. Thus, translational reinitiation is a common event in the *i* gene and may be expected to occur in other mRNAs with a comparable frequency.

We thank Donald Ganem and Marie-Francoise Aellen for helping us isolate the mutants used in this study, and Drs. A. Garen, M. van Montagu, and E. Signer for supplying us with suppressor strains. We also thank Dr. T. Platt and R. Chang for help during early work on mutant 606. This work was supported by Grants GM-17662 and GM-09541 from the National Institutes of Health. J.H.M. was supported by a grant from the Swiss National Foundation (F.N. 3.800.72).

1. Steitz, J. A. (1969) *Nature* **224**, 957–964.
2. Hindley, J. & Staples, D. H. (1969) *Nature* **224**, 964–967.
3. Staples, D. H. & Hindley, J. (1971) *Nature New Biol.* **234**, 211–212.
4. Staples, D. H., Hindley, J., Billeter, M. A. & Weissmann, C. (1971) *Nature New Biol.* **234**, 202–204.
5. Robertson, H. D., Barrell, B. G., Weith, H. L. & Donelson, J. E. (1973) *Nature New Biol.* **241**, 38–40.
6. Maizels, N. M. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3585–3589.
7. Ghosh, H. P., Söll, D. & Khorana, H. G. (1967) *J. Mol. Biol.* **25**, 275–298.
8. Sundararajan, T. A. & Thach, R. E. (1966) *J. Mol. Biol.* **19**, 74–90.
9. Dube, S. K., Rudland, P. S., Clark, B. F. C. & Marcker, K. A. (1969) *Cold Spring Harbor Symp. Quant. Biol.* **34**, 161–166.
10. Clark, B. F. C. & Marcker, K. A. (1966) *J. Mol. Biol.* **17**, 394–406.
11. Uhlenbeck, O. C., Baller, J. & Doty, P. (1970) *Nature* **225**, 508–510.
12. RajBhandary, U. L. & Ghosh, H. P. (1969) *J. Biol. Chem.* **244**, 1104–1113.
13. Sarabhai, A. & Brenner, S. (1967) *J. Mol. Biol.* **27**, 145–162.
14. Michels, C. A. & Zipser, D. (1969) *J. Mol. Biol.* **41**, 341–347.
15. Newton, A. (1969) *J. Mol. Biol.* **41**, 329–339.
16. Platt, T., Weber, K., Ganem, D. & Miller, J. H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 897–901.
17. Ganem, D., Miller, J. H., Files, J. G., Platt, T. & Weber, K. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3165–3169.
18. Weiner, A. M., Platt, T. & Weber, K. (1972) *J. Biol. Chem.* **247**, 3242–3251.
19. Beyreuther, K., Adler, K., Geisler, N. & Klemm, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3576–3580.
20. Ganem, D. E. (1972) Honors Thesis, Biochemical Sciences, Harvard College.
21. Weber, K., Platt, T., Ganem, D. & Miller, J. H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3624–3628.
22. Yanofsky, C., Cox, E. C. & Horn, V. (1966) *Proc. Nat. Acad. Sci. USA* **55**, 274–281.
23. Platt, T., Files, J. G. & Weber, K. (1973) *J. Biol. Chem.* **248**, 110–121.
24. Miller, J. H., Beckwith, J. & Müller-Hill, B. (1968) *Nature* **220**, 1287–1290.
25. Capecchi, M. R. (1966) *Proc. Nat. Acad. Sci. USA* **55**, 1517–1524.
26. Waller, J.-P. (1963) *J. Mol. Biol.* **7**, 483–496.
27. Stewart, J. W., Sherman, F., Shipman, N. A. & Jackson, M. (1971) *J. Biol. Chem.* **246**, 7429–7445.
28. Volckaert, G. & Fiers, W. (1973) *FEBS Lett.* **35**, 91–96.