Proc. Nat. Acad. Sci. USA
Vol. 69, No. 4, pp. 897–901 April 1972

Translational Restarts: AUG Reinitiation of a *lac* Repressor Fragment

(protein biosynthesis/antibody/gel electrophoresis/amino-acid sequence/*E. coli*)

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

The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138

Communicated by J. D. Watson, January 12, 1972

ABSTRACT **An early, spontaneous amber mutation in the *lac i*-gene allows translational reinitiation, which results in a mutant *lac* repressor. Comparison of the amino-terminal sequence of this mutant repressor with the partial amino-acid sequence of the wild-type *lac* repressor shows that reinitiation occurs at the first internal AUG codon, and results in a mutant protein lacking 42 residues at the amino-terminal end. This protein binds the inducer isopropyl-β-D-thiogalactoside with normal affinity, and is capable of maintaining a tetrameric structure; however, it does not repress *in vivo*. These data suggest that the amino-terminal portion of the wild-type *lac* repressor is necessary either for direct binding to the *lac* operator or for the correct conformation for binding to DNA.**

The codon AUG functions as an initiation signal in protein biosynthesis (1). An AUG triplet occurring internally codes for the insertion of methionine; under normal conditions, no initiation occurs within a gene. However, internal initiation sites have been found in certain cistrons. These sites have little or no effect on the wild-type gene product, and function only in the presence of a nearby terminator codon, permitting the reinitiation of translation and the synthesis of a C-terminal fragment of the gene product. Once the site is made available by chain termination, initiation presumably occurs by the normal mechanism (2).

Sarabhai and Brenner have described a reinitiation site in the rII-B cistron of bacteriophage T4; chain termination is required for this site to function in initiation (3). Both the nonsense fragment and a C-terminal fragment of the rII-B gene product are synthesized, and the insertion of frameshift mutations shows that reinitiation is independent of the phase between the nonsense codon and the starter site.

Newton (4, 5), and Zipser and others (6–8) have found initiation sites within the *z*-gene of the *lac* operon in *Escherichia coli*. These have no effect during normal translation, and do not affect the gene product, β-galactosidase. As in the rII-B cistron, however, chain termination close to one of these sites results in the synthesis of a reinitiated polypeptide, which is the C-terminal fragment of β-galactosidase, that can be detected both *in vivo* and *in vitro* (8, 9).

These results demonstrate that internal initiation signals exist, and are able to function in protein biosynthesis under certain conditions. It is not known what information is necessary for initiation beyond the requirement for an initiating codon.

In a genetic study to be described in a subsequent paper, we have demonstrated the existence of two natural reinitiation sites early in the *i*-gene of the *lac* operon of *E. coli*. We report here the characterization of a *lac* repressor polypeptide that arises by reinitiation at the earlier of these sites as the result of chain termination at an amber codon. To locate the site of reinitiation, we determined the sequence of the amino-terminal regions of the mutant and wild-type repressors. Our results imply that reinitiation at an AUG codon occurs in the mutant strain. Since the reinitiated polypeptide retains some wild-type activity, conclusions about structure and function relationships in the *lac* repressor may be drawn from a comparison of the functional properties of the mutant and wild-type molecules.

RESULTS

We have found five amber mutations that result in reinitiation in a collection of 51 spontaneous and independently isolated nonsense mutations in the *i*-gene of the *lac* operon. The details of this genetic study will be presented in a subsequent paper. The evidence for reinitiation is that these five amber mutants (i^{100}, i^{970}, i^{136}, i^{919}, and i^{971}) map early in the gene (Fig. 1), yet display partial negative complementation in diploid tests with the wild-type *i*-gene in a nonsuppressing (Su^-) background. An early amber mutation in a gene should produce a small amber fragment (from the amino-terminus to the amber block) possessing few properties of the wild-type gene product. However, negative complementation is thought to reflect an interaction in mixed oligomers between mutant and wild-type subunits, in which the mutant subunits are dominant, to produce a partial relief of repression in the absence of inducer (10). Thus, the negatively complementing repressor activity in strains carrying an early amber mutation in the *i*-gene is more likely to be due to a reinitiated polypeptide than to the amber fragment itself.

These five negatively complementing amber mutations map into two groups. It seemed convenient to begin our study with one of the earlier group, and we chose the i^{100} strain. We shall show that the negatively complementing species is indeed a reinitiated polypeptide, and refer to it as the i^{100} repressor protein. We describe here the characterization of the *lac* repressor polypeptide isolated from strains carrying the i^{100} mutation. To obtain enough material for chemical studies, the amber mutation i^{100} was crossed from the i^{Q} episome, on which it was originally isolated, onto a temperature-inducible

Abbreviations: IPTG, isopropyl-β-D-thiogalactoside; SDS, sodium dodecyl sulfate.

* Present address: Institute de Biologie Moleculaire, Universite de Genéve, Geneva, Switzerland.

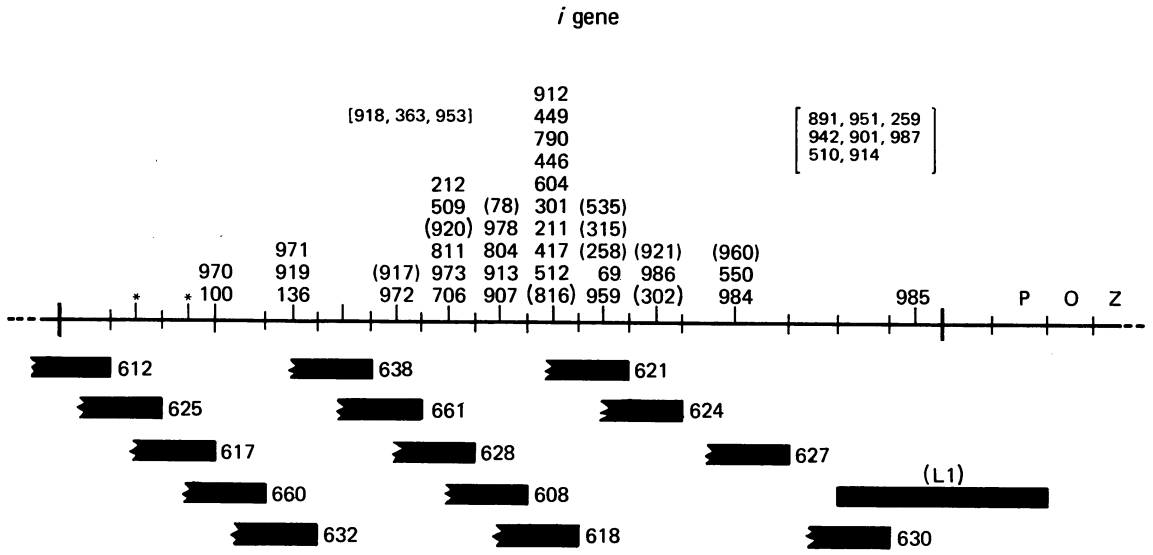

FIG. 1. Nonsense mutations in the *i*-gene. Fourteen deletions cutting in from the left end of the *i*-gene have been used to map 51 nonsense mutations. The precise location of markers in brackets and parentheses has not been determined. More detailed mapping and a description of the methods used are given in a separate genetic study, which will be reported in a subsequent paper. The five amber mutations, which result in negatively complementing material, are i^{100}, i^{970}, i^{136}, i^{919}, and i^{971}. Since the direction of reading of the *i*-gene is from left to right (17, 18), these are the five earliest nonsense mutations in the gene. The *asterisks* mark the position of mutations isolated by D. Gho and J. H. Miller, used to order the early deletions. The relative order of the deletions used above has also been determined independently by M. Pfahl.

λ*h80 dlac* prophage, which also carried the i^Q marker (10). The i^Q promoter overproduces repressor 10-fold, and induction of the lysis-defective prophage gives a further 20-fold increase in the number of repressor molecules per cell (10).

Inducer binding and *s* value of i^{100} repressor

Extracts from these i^{100} cells contain material that binds IPTG (isopropyl-β-D- thiogalactoside) and crossreacts with antibody made against wild-type repressor, although at a level 10-times lower than in the isogenic wild-type strain. Precipitation with 33% ammonium sulfate gives an 8-fold purification of the i^{100} protein, and chromatography on a DEAE-cellulose column provides an additional 3-fold increase in specific activity. At this stage, the binding constant of the i^{100} protein for IPTG can be determined by equilibrium dialysis against radioactive IPTG (11, 12). The value obtained, 1.0 μM, agrees within experimental error with that of wild-type repressor (Table 1). This finding confirms that the actual amount of i^{100} protein in cell extracts is only 10% of the normal level. Sucrose gradients run in parallel with wild-type repressor and four marker proteins yield an *s* value for the i^{100} protein about 8% less than that of the wild-type tetramer, corresponding to a molecular weight for the mutant species of about 135,000.

Amino-terminal sequence of the i^{100} repressor

Pure i^{100} protein was obtained by the technique of antibody purification described by Rosenbusch and Weber (manuscript in preparation). This procedure yields the molecular weight of the polypeptide chain, and provides pure material that can be sequenced by the modified Dansyl–Edman technique developed by Weiner, A. M., Platt T. & Weber, K., manuscript in preparation.

After the initial ammonium sulfate fractionation, the i^{100} species represents about 0.1–0.2% of the total protein in the sample. Antibody is added, and the mixture is incubated overnight at 4°C. The insoluble antibody–antigen complex is isolated by centrifugation, carefully washed, dissociated with 2-mercaptoethanol and sodium dodecyl sulfate, and run on 12-cm 10% polyacrylamide–sodium dodecylsulfate gels (13, 14). On analytical gels of this type, the *lac* repressor runs as a single band, with a molecular weight of 38,000. Gels that compare wild-type repressor with antibody-purified i^{100} protein are shown in Fig. 2. The molecular weight of the mu-

TABLE 1. *Physical properties of the i^{100} protein and the wild-type lac repressor*

	Repression *in vivo*	IPTG binding constant, μM	*s* value	Subunit size	State of aggregation	Amino-terminal sequence
i^{100} Protein	No	1.0	6.6	34,000	Tetramer	Ala-Glx-Leu-Asx-
lac Repressor	Yes	1.3	7.2	38,000	Tetramer	Met-Lys-Pro-Val-

The table shows a comparison of the properties of the two molecules, as determined by methods described in the text. The significant functional difference between the two molecules is that the i^{100} protein is unable to repress *in vivo*. The wild-type amino-terminal sequence was found to be identical for the i^+, i^Q, and i^{SQ} gene products. The other wild-type properties have been reported (10, 11). The *s* values correspond to molecular weights of 135,000 and 150,000 for the native i^{100} protein and *lac* repressor, respectively.

tant monomer is about 34,000. This result, together with the *s* value of the i^{100} protein, demonstrates that the native state of i^{100} repressor is tetrameric.

On a preparative scale, we used antibody to purify 0.5 mg of the crossreacting i^{100} protein, after precipitation with 33% ammonium sulfate from an extract prepared from 80 g of i^{100} cells. The antibody–antigen complex was run in parallel on five 0.6 × 12 cm gels, and the band corresponding to the i^{100} protein was eluted. The modified Dansyl–Edman procedure yielded an unambiguous amino-terminal sequence for the i^{100} protein: Ala-Glx-Leu-Asx-.

Amino-terminal sequence of the wild-type repressor

To locate the site of reinitiation, we determined the amino-terminal sequence of the wild-type molecule, shown in Fig. 3 for residues 1–51. We obtained this sequence by standard techniques from a cyanogen bromide fragment (residues 2–42), and an overlapping tryptic peptide (residues 36–51). Amino-terminal sequencing of the whole protein by the modified Dansyl–Edman procedure yielded residues 1–16. This sequence was derived from a strain containing the mutation i^{SQ}, which leads to a 30- to 50-fold overproduction of *lac* repressor, and is thought to affect the *i*-promoter (12). To determine whether the promoter mutations i^{SQ} or i^{Q} also

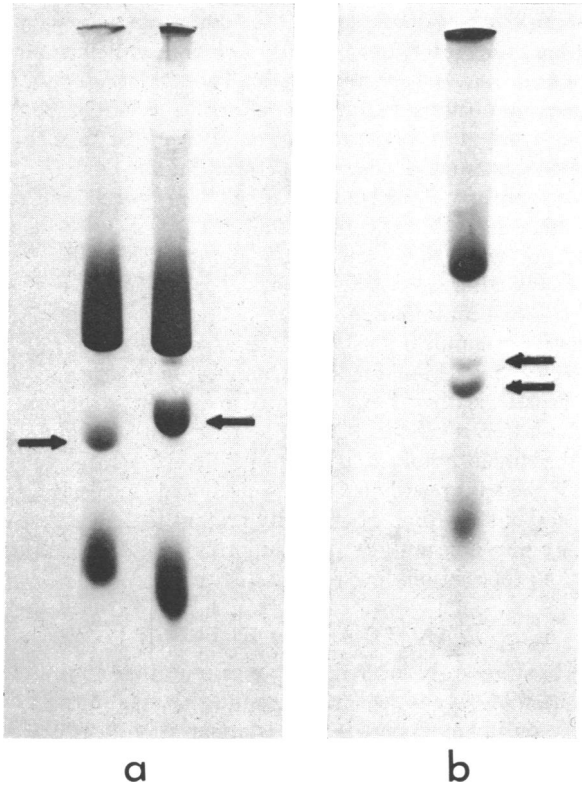

FIG. 2. Sodium dodecyl sulfate–polyacrylamide gels of antibody–antigen complexes. (*a*) Guidestrips from preparative gels, which are heavily overloaded to provide material for amino-terminal sequence determination. The middle band is the i^{100} protein (*left-hand gel*), or wild-type repressor (*right-hand gel*). No i^{100} material is visible on the wild-type gel. (*b*) An analytical gel used for the molecular weight determination. In order, from the top of the gel, the four bands are the heavy IgG (50,000), *lac* repressor (38,000), i^{100} protein (34,000), and light IgG (23,500) polypeptide chains. For this gel, 10 μg of pure *lac* repressor was added to i^{100} protein antibody–antigen complex.

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-Ser-Ala-Lys-Thr-
Arg-Glu-Lys-Val-Glu-Ala-Ala-Met-Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Asn-Arg-

FIG. 3. Amino-terminal sequence of wild-type *lac* repressor. Determination of this sequence will be reported elsewhere. The amino-terminal sequence of the i^{100} protein (underlined in the figure) corresponds to residues 43–46, and follows immediately the first internal methionine, which occurs at residue 42 of the wild-type polypeptide chain. The amino-terminal sequence of the wild-type molecule has been determined independently by Adler *et al.*‡

might affect the primary structure at the amino-terminus of the repressor, we sequenced 16 residues of the i^{Q} and 5 of the i^{+} gene-products by the modified Dansyl–Edman degradation, and found no differences in this region (Table 1). Since i^{SQ} and i^{Q} both map at the extreme left-hand end of the *i*-gene†, it is likely that the primary structure of the i^{+}, i^{Q}, and i^{SQ} gene-products are identical. Within this region, it can be seen that residues 43–46 (-Ala-Glu-Leu-Asn-) correspond to the amino-terminal sequence of the i^{100} protein, and immediately follow a methionine at residue 42. A polypeptide chain initiating at this position would lack 42 of the 350 residues of the wild-type repressor, or 12% by molecular weight. The smaller size of the i^{100} polypeptide chain determined by gel electrophoresis agrees closely with this prediction.

DISCUSSION

A comparison of the amino-terminal sequences of the i^{100} protein and the wild-type *lac* repressor suggests that reinitiation in the i^{100} messenger RNA occurs at the AUG codon specifying methionine$_{42}$. This reinitiation site is the first internal, inphase, AUG in the *i*-gene after the early amber block. During biosynthesis of the nascent i^{100} polypeptide, it may be assumed that the amino-terminal methionine corresponding to met$_{42}$ in the wild-type repressor is enzymatically removed, since it is followed by an alanine (1). The i^{100} amber codon is expected to precede met$_{42}$, and to provide the chain-terminating signal that allows this AUG to be used as a site for reinitiation. The amber codon cannot follow met$_{42}$, since this order would result in chain termination of the reinitiated polypeptide. The sequence of repressor from suppressed i^{100} cells should reveal the amber position. It is difficult to rule out the formal possibility of initiation before met$_{42}$, followed by limited proteolytic degradation; however, this hypothesis must assume initiation *in vivo* at a codon other than AUG, and cleavage that leads to a homogeneous amino-terminus.

The five earliest nonsense mutations in the *i*-gene all permit reinitiation, and map in two regions, as shown in Fig. 1. Preliminary experiments with i^{136} and i^{919}, which map in the latter group, show that they also yield a crossreacting frag-

† Miller, J. H. (1970) thesis, Harvard University.

‡ Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Muller-Hill, B., Pfahl, M. & Schmitz, A. (1972) submitted to *Nature New Biol.*

ment of *lac* repressor with IPTG-binding activity. These two amber mutations produce identical polypeptides, which differ from the i^{100} protein. The antibody purification yields a molecular weight for these fragments about 20% smaller than wild-type repressor, while the i^{100} protein is only 12% smaller. Their amino-terminus is methionine, and they have a different amino-terminal sequence than the i^{100} polypeptide. It remains to be seen whether an AUG codon is responsible for reinitiation at this later position, since we do not yet have the sequence of the wild-type repressor in this region.

The efficiency of reinitiation in the case of the i^{100} protein is at least 10%, based on the amount of IPTG-binding activity in cell extracts from mutant and wild-type strains. However, we have shown (13) that another mutant repressor is found at reduced levels in cell extracts because it is subject to degradation *in vivo*, although it is synthesized at the normal rate. Experiments are in progress to determine whether the rate of i^{100} repressor synthesis is higher than 10%.

Does the AUG codon at met_{42} also function as an initiation site in the wild-type *i*-gene? In the presence of the i^{100} amber mutation, reinitiation produces at least 10% of the normal level of *lac* repressor. In Fig. 2*a*, the heavily overloaded gel of wild-type repressor shows no material at the i^{100} protein position, although a 1% contaminant would easily be visible. Thus, initiation at the met_{42} AUG codon cannot occur at a significant level in the wild-type messenger. This finding is consistent with previous studies, showing that reinitiation sites function only if a nonsense codon is nearby (3–9), that suggested that the efficiency of reinitiation is related to the distance between the chain-terminating codon and the reinitiating locus. One may also ask whether any internal AUG codon may function in reinitiation, once polypeptide chain termination has occurred nearby. The reinitiation sites in the *i*-gene should permit testing of these hypotheses at the molecular level.

In their studies on yeast, Stewart *et al.* have recently shown that the precise location of an initiating AUG is not crucial to the synthesis of iso-1-cytochrome *c* (15). They selected revertants from a strain that produced no enzyme because a mutational event had altered the initiating AUG codon. In particular, a transition of this codon to GUG, which has been suggested as an alternative initiating codon by the *in vitro* experiments of Ghosh *et al.* (16), did not allow synthesis. The revertants were all of three types, and resulted from the creation of a new in-phase AUG codon either immediately before, at, or four codons after the original site. The synthesis of slightly altered cytochrome *c* molecules initiating at the first or last sites did not appear to be affected by the change in location of the initiating codon.

These results support our evidence that polypeptide initiation may occur at other AUG codons than those normally used for this function. Future experiments should reveal to what extent other elements in the sequence or structure of the surrounding mRNA may also be necessary for initiation.

It seems plausible that reinitiation at locations within a gene may be a relatively common phenomenon, although most restarts may not result in the formation of a stable and partially active fragment. The *lac* repressor may be unusual in retaining its oligomeric structure and the ability to bind inducer, even though a large amino-terminal segment of the molecule is missing.

The purification of the i^{100} polypeptide permits a determination of its size and amino-terminal sequence. This determination also allows us to verify with direct chemical evidence the direction of reading in the *i*-gene, which has previously been predicted by genetic studies (17) and by hybridization analysis (18). First, the physical position of the reinitiation site at residue 42 (Fig. 3) and the genetic location of the i^{100} amber mutation (Fig. 1) show that the left-hand end of the *i*-gene represents the amino-terminal end of the *lac* repressor. Second, the amber mutation i^{136}, which maps farther in from the left than i^{100}, gives rise to a restart polypeptide that is even shorter than the i^{100} repressor. Third, the mutant L1 repressor, which arises by a deletion (shown in Fig. 1) cutting into the right-hand end of the *i*-gene, has a homogeneous amino-terminal sequence identical to that of wild type (unpublished data).

Chemical characterization of the reinitiated polypeptide yields the location of the restart site; the partial activity retained by native i^{100} repressor allows an interesting comparison between the functional properties of the mutant and wild-type molecules (Table 1). The absence of 42 amino-terminal residues (12% of the polypeptide chain) does not interfere with the ability of the molecule to bind the inducer IPTG. Similarly, the i^{100} mutant protein is also a tetramer, suggesting that the missing residues are not involved in maintaining tetrameric structure in the wild-type *lac* repressor. The tetrameric form is thought to be necessary for binding to operator, but is clearly not sufficient since the i^{100} protein is unable to repress *in vivo*. This finding implies that some region before residue 43 is either (*a*) directly involved in the mechanism of repression, or (*b*) necessary for the repressor to assume the correct operator-binding conformation. Since there is nothing unusual about this portion of the sequence, except that it contains half of the eight tyrosine residues present in the protein, we cannot distinguish between these two alternatives. However, Adler *et al.*‡ have studied i^{-d} missense mutations, and have mapped them in comparison to our i^{100} mutant. Based on these genetic results and on molecular model building of the amino-terminal region of the *lac* repressor, they have proposed a specific structural hypothesis for repressor–DNA interaction that involves part of the amino-terminal sequence in direct operator binding. Their model is readily testable, since it makes specific predictions about the properties of any mutant repressor that contains a missence mutation or a suppressed nonsense mutation in the proposed binding region.

NOTE ADDED IN PROOF

We have recently located an amber mutation at residue 26 in the wild-type sequence. This mutation was derived by 2-aminopurine mutagenesis, and is a transition from the glutamine codon CAG to the nonsense codon UAG. Strains that carry this amber mutation also give rise to the i^{100} restart protein. In this case, the distance from the termination codon to the reinitiating AUG (corresponding to residue 42) is 48 nucleotides long.

We thank A. M. Weiner, J. P. Rosenbusch, and W. Gilbert for their helpful advice and criticism, and Dr. B. Müller-Hill for communication of unpublished results as reported in Adler *et al.*‡. T. P. was supported by a N. S. F. Predoctoral Fellowship, and J. H. M. by the Society of Fellows, Harvard University. Additional support was provided by N. I. H. Grants GM 17662-01 and GM 09541-10.

1. Capecchi, M. R. (1966) *Proc. Nat. Acad. Sci. USA* **55,** 1517–1524.
2. Lengyel, P. & Soll, D. (1969) *Bacteriol. Rev.* **33,** 264–301.
3. Sarabhai, A. & Brenner, S. (1968) *J. Mol. Biol.* **27,** 145–162.
4. Newton, A. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31,** 181–187.
5. Newton, A. (1969) *J. Mol. Biol.* **41,** 329–339.
6. Michaels, C. A. & Zipser, D. (1969) *J. Mol. Biol.* **41,** 341–347.
7. Zipser, D., Zabell, S., Rothman, J., Grodzicker, T. & Wenk, M. (1970) *J. Mol. Biol.* **49,** 251–254.
8. Grodzicker, T. & Zipser, D. (1968) *J. Mol. Biol.* **38,** 305–314.
9. Ullman, A. & Perrin, D. (1970) in *The Lactose Operon,* eds. Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Laboratory), pp. 143–172.
10. Muller-Hill, B., Crapo, L. & Gilbert, W. (1968) *Proc. Nat. Acad. Sci. USA* **59,** 1259–1264.
11. Gilbert, W. & Muller-Hill, B. (1966) *Proc. Nat. Acad. Sci. USA* **56,** 1891–1898.
12. Miller, J. H., Platt, T. & Weber, K. (1970) in *The Lactose Operon,* eds., Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Laboratory), pp. 343–351.
13. Platt, T., Miller, J. H. & Weber, K. (1970) *Nature* **223,** 1154–1158.
14. Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **244,** 4406–4412.
15. Stewart, J. W., Sherman, F., Shipman, N. A. & Jackson, M. (1971) *J. Biol. Chem.* **246,** 7429–7445.
16. Ghosh, H. P., Soll, D. & Khorana, H. G. (1967) *J. Mol. Biol.* **25,** 275–298.
17. Miller, J. H., Muller-Hill, B. & Beckwith, J. R. (1966) *Nature* **220,** 1287–1290.
18. Kumar, S. & Szybalski, W. (1969) *J. Mol. Biol.* **40,** 145–151.