EDITH BRICKMAN, THOMAS J. SILHAVY, PHILIP J. BASSFORD, JR., HOWARD A. SHUMAN, AND JONATHAN R. BECKWITH*

Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115

Received for publication 6 March 1979

We describe the genetic analysis of 21 *Escherichia coli* strains in which the amino-terminal sequence of β -galactosidase has been removed and replaced by an amino-terminal sequence from one or another of the proteins involved in maltose transport. Genetic mapping of the *lacZ* end of these fused genes indicates that only those fusions in which fewer than 41 amino acids are removed from the amino-terminal sequence of β -galactosidase result in enzymatically active molecules. Within the region between amino acid 17 and amino acid 41 there are at least four or five sites where enzymatically active hybrid proteins can be formed.

One of the postulated mechanisms of genetic evolution is the fusion of two genes to create new "hybrid" genes. In the case of fusions between two genes which code for proteins, new hybrid proteins are produced. These gene fusion events can thus potentially generate new enzymatic activities, can give rise to bifunctional proteins which confer some selective advantage, or can alter the localization of one of the proteins involved. One example of a possible protein fusion with evolutionary significance is the variation among microbes in the structure of the enzyme tryptophan synthetase. In a number of procaryotes, including Escherichia coli, tryptophan synthetase is composed of two distinct polypeptide chains, each one exhibiting its own specific enzymatic activity. In contrast, in other organisms, including Neurospora crassa, tryptophan synthetase activity is found in a single polypeptide chain, which still also shows exactly the same two enzymatic activities (21). It may be that in evolution a gene fusion event generated a single gene from the two independent genes found in procaryotes.

One of the best studied cases of gene fusions involves the generation in the laboratory of hybrid proteins between the *E. coli* enzyme β galactosidase and other *E. coli* proteins. The enzyme β -galactosidase is a tetramer made up of monomers of molecular weight of approximately 116,000 (7, 8). Two lines of evidence demonstrate that a significant portion of the amino-terminal sequence of β -galactosidase is not essential for enzyme activity. First, it was shown by Accolla and Celada (1) that the enzymatically inactive protein produced by a lacZdeletion mutant, M15, could, in the presence of antibody to β -galactosidase, become active again. M15 removes DNA sequences corresponding to amino acids 11 to 41 of the protein (11). It is hypothesized that this portion of the protein may be essential for activity only in providing a stable structure for the tetramer. Secondly, Müller-Hill and Kania (14) showed that it is possible to remove a portion of the amino-terminal sequence of β -galactosidase and replace it with the amino-terminal sequence of another protein and still retain enzymatic activity. In this case, the substitute peptide was derived from the lactose repressor, the product of the lacI gene. Differing lengths of the repressor protein could be attached to β -galactosidase with enzymatic activity still being retained. In fact, in many cases, a hybrid bifunctional protein was produced which contained enough of the lacI gene product that it exhibited both repressor and β -galactosidase activity (13, 14).

The technique used by Müller-Hill and Kania to select for hybrid β -galactosidase molecules involved the use of a strain carrying the *lacZ* ochre mutation U118. This mutation alters the codon corresponding to amino acid 17 in the normal β -galactosidase monomer (22). Among revertants to *lac*⁺ of strains carrying the U118 mutation are ones due to deletions which remove the early portion of the *lacZ* gene, including the U118 site, the *lacP*, and O sites, and fuse the remainder of the *lacZ* gene to a point within the *lacI* gene. Casadaban (5) extended the range of possibilities with this technique, by developing a means of transposing the *lac* region to designated sites on the *E. coli* chromosome. Using a *lac* region carrying the U118 mutation in such experiments, he was able to detect protein fusions of β -galactosidase to the product of the *araB* gene, L-ribulokinase. This approach has now been used to isolate a number of such protein fusions (see, for example, 3, 18, 19).

It has been possible to restore enzymatic activity to β -galactosidase by substituting its amino-terminal sequence with the amino terminus of a variety of proteins with very different properties: an outer membrane protein-namely the bacteriophage λ receptor (19); two periplasmic proteins-the maltose-binding protein and alkaline phosphatase (3; A. V. Sarthy and J. R. Beckwith, unpublished results); a cytoplasmic enzyme (araB); and a cytoplasmic DNA binding protein-the lactose repressor (13, 14). In several cases, the cellular location of β -galactosidase was altered. Thus, it appears that the amino-terminal sequence of nearly any protein is sufficient to restore enzymatic activity to a β -galactosidase molecule with its own amino terminus removed.

One question concerning these hybrid proteins is how much of the amino-terminal sequence of β -galactosidase can be removed and substituted for and still have activity retained. Another question is whether fusion points between the two proteins are limited to one or a few positions within the amino-terminal sequence of β -galactosidase, or can they be anywhere within this particular region. It is clear already from the properties of the deletion strain M15, which in the absence of antibody exhibits absolutely no enzymatic activity, that one cannot simply remove a portion of this region and retain activity. Furthermore, Kania and Müller-Hill (in H. Biswanger and E. M. Schimke-Ott, ed., Multifunctional Proteins, in press), have presented evidence that it may be impossible to fuse a lacZgene carrying the M15 mutation to another gene and restore β -galactosidase activity.

In this paper, we present genetic studies which confirm that the amino-terminal sequence of β galactosidase is removed in the strains which are producing hybrid β -galactosidase molecules. In addition, our results indicate that there are a number of different possible sites of fusion joints in the *lacZ* gene in various hybrid proteins.

MATERIALS AND METHODS

Bacterial strains. Bacterial strains used are listed in Table 1.

Media and chemicals. Media and chemicals used are described elsewhere (12).

LacZ mapping—approach 1. Diploid strains carrying fusions on the chromosome and F'lac pro episomes with various mutations were constructed as

TABLE 1. Bacterial strains"

Strain	Genotype				
E7151.1	F'lacZ2 lacI3 pro A^+B^+/Δ (lac-proAB) XIII recA supE nalA				
E7153.1	F'lacZ131 pro $\dot{A}^+B^+/\Delta(lac\text{-}proAB)XIII$ recA supE nalA				
E7154.1	F'lacZ84 proA ⁺ B ⁺ /Δ(lac-proAB)XIII recA supE nalA				
E7155.1	$F'lacZ\omega proA^+B^+/\Delta(lac-proAB)XIII recA supE nalA$				
E7089.2	F'lacZ118 proA ⁺ B ⁺ /Δ(lac-proAB)XIII recA supE nalA				
X7131	$\mathbf{F}^{-\Delta}(lac\text{-}proAB)XIII val^{r} rpsE$				
MC4100	F⁻araD139 ∆lac169 rpsL relA thiA				

^a Fusion strains are all in MC4100 background. Gene abbreviations are according to Bachmann et al. (2).

follows. The donor strains carrying different F'lacZpro episomes were mated with the fusion strains for 1 h, and dilutions were spread on LB agar containing streptomycin. Since there was no way to select for those clones which had received the F', a number of colonies from the selective agar were tested for their ability to transfer the pro⁺ marker to another recipient, X7131. The latter strain was spectinomycin resistant, which permits selection against the fusionderivative strains, which were spectinomycin sensitive. Those derivatives of the fusion strain which carried the F' factor were purified, grown up in broth culture, and crossed with X7131 to detect any Lac⁺ recombinant F' factors which arose from recombination between the *lacZ* point mutation and the fusion deletion. These were detected as Lac⁺ spectinomycin-resistant colonies on selective media.

In a number of cases, a very low frequency of apparent Lac⁺ recombinants was observed in cases where we expected none. One explanation for these anomalous recombinants was that the fusion itself was being transferred by F' mobilization of the chromosome into the recipient. If this were the case, one would expect that the β -galactosidase in such recombinants would not be inducible by isopropyl- β -D-thiogalactoside and that the recombinants would be Mal⁻. Therefore, in all such cases, Lac⁺ recombinant colonies were picked from the selective media and streaked on appropriate indicator media to determine their inducibility properties. In many cases, recombinants were detected which were Mal⁻ and not isopropyl- β -D-thiogalactoside inducible.

Lac mapping—approach 2. Isolation of malT mutants in mal-lac fusion strains. A direct selection for malT mutants exists. Among mutants selected for resistance to bacteriophage λ , a high proportion are malT. This result is attributable to the fact that the λ receptor protein, product of the lamB gene, is under maltose control (16). In the case of fusions of lacZ to malE and malF, direct selection of λ -resistant derivatives using λvir often yielded mutants with greatly reduced rates of β -galactosidase synthesis exhibiting a Lac⁻ phenotype. These were presumed to be malT mutations. Only malT mutations with low reversion frequencies were used. In the case of fusions of *lacZ* to *lamB*, the strains are already λ resistant, so that this direct selection for *malT* cannot be employed. However, it is possible in these cases to transfer the fusion into a *malT* background. Since fusions are isolated by a procedure which results in the fused operon being located adjacent to a λ phage, UV irradiation of such strains yields λ transducing phages carrying the fusion. Such transducing phages can then be used to lysogenize a *malT* strain. Even though these latter strains are λ resistant, the phage can be introduced by phenotypic mixing with phage ϕ 80 (19).

Diploids of the character $F'lacZ pro^+/lacZ$ -fusion, malT, were constructed by the same techniques as in approach 1. These diploids were grown up overnight in L broth, and 0.03 ml was dropped onto a lactoseminimal agar. Lac⁺ recombinants were counted after 3 days. In both approaches, all crosses were done at least in duplicate.

RESULTS

We have described a series of strains which were presumed to produce protein hybrids between β -galactosidase and one or another of several proteins involved in maltose transport

(3, 18, 19). The genes coding for the proteins of the maltose transport system are clustered in the malB region of the bacterial chromosome. Five such genes have been found in this region. They are clustered in two operons transcribed in opposite directions from the same controlling element region (9). The lamB gene codes for an outer membrane protein which acts both as a receptor for bacteriophage λ and is a component of the maltose transport system (16, 20). The malE gene codes for a periplasmic maltose-binding protein (10), and the *malF* gene codes for a cytoplasmic membrane protein essential for maltose transport. The malK (9) and malG (Silhavy et al., unpublished results) genes, whose products have not been identified, are also essential for maltose transport.

The strains carrying fusions between lacZ and mal genes are listed in Table 2. In all cases, an amino-terminal portion of β -galactosidase is thought to be removed and replaced by an amino-terminal sequence from one of the mal-tose-specific proteins. In work described below,

TABLE 2. Recombination of lacZ mutations with lacZ-mal fusions"

					•	
Fusion –		Gene to which				
	U131	2	X84	ω	U118	lacZ is fused"
41-4	+++	_	-	NT	_	malE
11-1	+++	-	-	NT	_	malF
3-21	+++	-	-	NT	-	malE
11-3	+++	-	-	NT	-	malF
69-1	+++	-	-	NT	-	malG
62-37	+++	+	-	NT	_	malE
53-1	+++	+	-	NT	-	malF
57-1	+++	+	-	NT	-	malF
66-2	+++	+	-	NT	-	malF
57-3	+++	+	-	NT	-	malF
72-47	+++	++	-	NT	_	malE
4-1	+++	++	-	NT	-	malE
14-4	+++	++	_	NT	_	malF
179-3	+++	++	-	NT	-	malE
4-81	+++	++	-	NT	_	malE
41-2						malF
42-1	+++	++	+	-	_	lamB
72-27	+++	++	+	-	-	malE
6-3	+++	++	+	-	-	malF
62-20	+++	++	+	+	_	malE
11-2	+++	++	+	+	-	malF
179-7	+++	+++	+++	+++	+++	malE

^a Recombination frequencies are indicated by the symbols as follows: +++, greater than 10^{-6} Lac⁺ recombinants per diploid; ++, approximately 10^{-6} ; +, between 6×10^{-8} and 2×10^{-7} ; -, less than 5×10^{-8} . Frequencies below 5×10^{-8} are comparable to reversion frequencies. Those fusions which have been described elsewhere are 11-1 (MC4416, ref. 18); 42-1 (42-1 or pop3186, ref. 19); 41-4, 4-81, 72-47, 62-37 and 179-3 (ref. 3). NT, Not tested. All fusions are in the MC4100 background (5).

^b The appropriate mal gene was determined by genetic mapping and complementation studies.

we have taken each of these strains and by genetic crosses demonstrated that, in fact, a portion of the lacZ gene corresponding to the amino terminus of the protein has been removed.

Approach 1. Both approaches involve crossing into fusion strains F' plasmids carrying lacZ mutations located early in the gene, and then determining whether Lac⁺ recombinants can arise in these diploids for the lac region. Since the original fusion strains were selected for their Lac⁺ phenotype, it is, in most cases, difficult to seek Lac⁺ recombinants in the diploid strains themselves. Therefore, in our first approach we detect those F' plasmids which have become Lac^+ , due to a recombination with the fusion, by transferring these plasmids into another strain which is Lac⁻ (see Materials and Methods). F'lac pro plasmids carrying the early lacZ mutants, U118, ω , X84, 2, and U131 (Fig. 1), were crossed into recipient strains carrying a mallacZ hybrid. Such lac diploid clones were then crossed with a recipient strain which carried a total deletion of the lac region to detect Lac⁺ recombinants.

This approach can be used for the mapping of the lacZ end of any protein fusion deletion. However, we developed a more convenient method for characterizing these strains subsequently. The data obtained with approach 1 (unpublished results) are all consistent with data presented below. The above details are included because the approach is of more general applicability than approach 2.

Approach 2. To avoid the requirement for two bacterial crosses for each lacZ mapping experiment involved in approach 1, we have developed an alternative approach which is made possible by knowledge of the mechanism of regulation of the maltose genes. The proteins of maltose metabolism are regulated by the product of the *malT* gene which is a positive

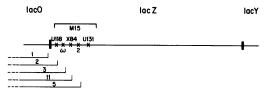


FIG. 1. lacZ endpoints of deletions which fuse β galactosidase to maltose transport proteins. The map order of the lacZ point mutations was described previously (15, 17). The U131 mutation has been shown not to recombine with the M15 deletion (Brickman and Beckwith, unpublished results). The numbers in the different deletion intervals are the total numbers of fusions with endpoints in that interval. This map is not drawn to scale. control activator protein (6). Mutations which inactivate this gene result in non-inducibility for the *mal* genes and a Mal⁻ phenotype. When *malT* mutations are introduced into *mal-lac* fusions strains, the strains become Lac⁻.

malT mutations were isolated in all of the *mal-lac* fusion strains as described in Materials and Methods. These phenotypically Lac⁻ derivatives of the fusions were then crossed with the same series of early *lac* point mutations carried by F' plasmids. Diploids which were heterozygous for the *lac* region (*lac⁻* point mutant/*mallac* fusion) were detected and purified. The frequency of Lac⁺ recombinants in these diploids was then measured.

The results of the crosses when done by both approaches were similar. These results using approach 2 are presented in Table 2 and Fig. 1. In all but one case, fusion 179-7, a presumed *malE-lacZ* protein fusion, we were unable to obtain Lac⁺ recombinants with the *lacZ* mutation U118. In contrast, fusion 179-7 yielded recombinants with all *lacZ* mutations tested, including U118. We believe that, in this case, the Lac⁺ character is due, in part, to the reversion of the U118 mutation, and that no hybrid protein is produced in this strain.

DISCUSSION

The technique developed by Müller-Hill and Kania (14) permitted the isolation of hybrid proteins between β -galactosidase and the lactose repressor. In one case, the hybrid nature of the protein product was directly demonstrated by the ability of the hybrid protein to interact with antibodies to both repressor and β -galactosidase and more recently by amino acid sequence analysis (4). In the case of mal-lac fusion strains, we have presented several lines of evidence suggesting that hybrid proteins are being produced. In many cases, the subcellular location of β -galactosidase is altered by the fusion event, indicating its covalent attachment to a portion of another protein (3, 18, 19). In addition, many presumed protein fusion strains produce proteins of higher molecular weight than β -galactosidase which appear to be the product of the hybrid gene (3, 18, 19). Finally, in one case (fusion 11-1, Table 2), we have demonstrated that a particular purified hybrid protein has antigenic determinants of both the malF gene product and β -galactosidase (H. A. Shuman and J. R. Beckwith, unpublished results).

In this paper we present genetic evidence that, in all but one of the presumed protein fusion strains which we have tested, the portion of the lacZ gene corresponding to the amino-terminal sequence of β -galactosidase has been deleted. These results provide further evidence for our assumption that the hybrid proteins we are studying are indeed covalent fusions of β -galactosidase with amino-terminal sequences of different maltose transport proteins.

Protein fusions occur as very rare deletion events in the strains in which they are selected. Considering this low frequency, it seemed possible that, in some cases, Lac⁺ derivatives of the starting strains might arise as the result of double events: i.e., a reversion of the U118 mutation and a deletion which did not extend into the lacZ gene but which generated a fusion with an intact lacZ gene. In fact, of 22 presumed protein fusions examined, one (179-7) appears to be due to just such a double event. Thus, if one is to use presumed protein fusion strains for any sort of analysis, it is important to have some direct evidence, genetic or otherwise, that the strain is indeed what it is thought to be. The genetic mapping techniques described here could be emploved.

An analogous technique to approach 2 used in this paper to map *mal-lacZ* fusions can be used in those systems where it is possible to introduce regulatory mutations similar to *malT* into the fusion strains. Müller-Hill and Kania (14) have done preliminary mapping of a *lacI-Z* fusion in this way. In addition, we have mapped the *lacZ* end of protein fusions of *lacZ* to the *phoA* gene (structural gene for alkaline phosphatase) by introducing *phoB* mutations into the strain which result in loss of expression of *phoA* (Sarthy and Beckwith, unpublished results). However, in cases where no such regulatory mutations exist for a particular pathway, approach 1 can be used for mapping.

The genetic mapping data give some information on the amount of β -galactosidase which is removed by deletions which generate protein fusions. A number of these deletions remove the sites corresponding to the early *lacZ* mutations, U118, ω , X84, and 2. The lacZ U118 mutation occurs at a site corresponding to amino acid 17 in β -galactosidase (22). None of the fusion deletions remove the region corresponding to the U131 mutation. In addition, all the fusions give considerably higher recombination frequencies with the U131 mutation than with any of the other mutations. An internal lacZ deletion, M15. removes sites corresponding to all five of these mutations and is known to end at a position in the gene corresponding to amino acid 41 (11). These frequencies suggest that the U131 site is located at a considerable distance from the other four mutations (see Fig. 1). It appears that removal of this site by a deletion event which fuses *lacZ* to another gene cannot generate an active β -galactosidase. This result, together with the finding that active hybrid proteins cannot be obtained from strains which carry the M15 deletion, indicate that fusion endpoints rarely, if ever, extend as far as amino acid 41 in β -galactosidase. The *lacI-Z* hybrid endpoint determined by amino acid sequence analysis is between amino acid positions 23 and 24 of β -galactosidase (4).

The map of deletion endpoints shown in Fig. 1 must be somewhat qualified since we are dealing with very low recombination frequencies. The different mutations alter nucleotide sequences within a relatively small region. However, it still does appear that, within this region, endpoints have been found in all possible intervals. Thus, it seems likely that there are no serious limitations within this region on the fusion joint between other proteins and β -galactosidase. These results provide further indications that the amino acid sequence of this region of the protein can be varied pretty much at will without serious alteration of β -galactosidase activity.

ACKNOWLEDGMENTS

We thank Ann McIntosh and Ronnie MacGillivray for assistance.

This research was supported by a grant from the National Science Foundation to J. Beckwith.

P. Bassford is a fellow of the Helen Hay Whitney Foundation. H. Shuman was supported by a National Research Service Award.

LITERATURE CITED

- Accolla, R. S., and F. Celada. 1976. Antibody-mediated activation of a deletion-mutant β-galactosidase defective in the α region. Fed. Eur. Biochem. Soc. 67:299-303.
- Bachmann, B. J., K. B. Low, and A. L. Taylor. 1976. Recalibrated linkage map of *Escherichia coli* K-12. Bacteriol. Rev. 40:116-167.
- Bassford, P. J., Jr., T. Silhavy, and J. Beckwith. 1979. Use of gene fusion to study secretion of maltose-binding protein into *Escherichia coli* periplasm. J. Bacteriol. 139:19-31.
- Brake, A. J., A. V. Fowler, I. Zabin, J. Kania, and B. Müller-Hill. 1978. β-Galactosidase chimeras: primary structure of a *lac* repressor-β-galactosidase protein. Proc. Natl. Acad. Sci. U.S.A. 75:4824-4827.
- 5. Casadaban, M. 1976. Transposition and fusion of the *lac* operon to selected promoters in *E. coli* using bacteriophage λ and Mu. J. Mol. Biol. 104:541-555.
- Débarbouillé, M., H. A. Shuman, T. J. Silhavy, and M. Schwartz. 1978. Dominant constitutive mutations in malT, the positive regulatory gene of the maltose regulon in E. coli. J. Mol. Biol. 124:359–371.
- Fowler, A. V., and I. Zabin. 1977. The amino acid sequence of β-galactosidase of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 74:1507-1519.
- Fowler, A. V., and I. Zabin. 1978. Amino acid sequence of β-galactosidase. XI. Peptide ordering procedures and the complete sequence. J. Biol. Chem. 253:5521-5525.
- Hofnung, M. 1974. Divergent operons and the genetic structure of the maltose B region in *Escherichia coli* K12. Genetics 76:169-184.

- Kellerman, O., and S. Szmelcman. 1974. Active transport of maltose in *Escherichia coli* K12: involvement of a "periplasmic" maltose binding protein. Eur. J. Biochem. 47:139-149.
- Langley, K. E., M. R. Villarejo, A. V. Fowler, P. J. Zamenhof, and I. Zabin. 1975. Molecular basis of βgalactosidase complementation. Proc. Natl. Acad. Sci. U.S.A. 72:1254-1257.
- Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Müller-Hill, B., G. Heidecker, and J. Kania. 1976. Repressor-galactosidase-chimeras, p. 167-179. In Proceedings of the Third John Innes Symposium.
- Müller-Hill, B., and J. Kania. 1974. Lac repressor can be fused to β-galactosidase. Nature (London) 249:561– 563.
- Newton, W. A., J. R. Beckwith, D. Zipser, and S. Brenner. 1965. Nonsense mutants and polarity in the *Lac* operon of *Escherichia coli*. J. Mol. Biol. 14:290-296.
- Randall-Hazelbauer, L., and M. Schwartz. 1973. Isolation of the phage lambda receptor from Escherichia

coli K-12. J. Bacteriol. 116:1436-1446.

- Schwartz, D. O., and J. R. Beckwith. 1969. Mutagens which cause deletions in *Escherichia coli*. Genetics 61: 371-376.
- Silhavy, T., M. J. Casadaban, H. A. Shuman, and J. R. Beckwith. 1976. Conversion of β-galactosidase to a membrane-bound state by gene fusion. Proc. Natl. Acad. Sci. U.S.A. 73:3423-3427.
- Silhavy, T., H. A. Shuman, J. Beckwith, and M. Schwartz. 1977. The use of gene fusions to study outer membrane protein localization in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 74:814-817.
- Szmelcman, S., M. Schwartz, T. J. Silhavy, and W. Boos. 1976. A comparison of transport kinetics in wildtype and λ-resistant mutants with the dissociation constants of the maltose-binding protein as measured by fluorescence quenching. Eur. J. Biochem. 65:13-19.
- Yanofsky, C., and P. St. Lawrence. 1960. Gene action. Annu. Rev. Microbiol. 14:317-320.
 Zabin, I., A. V. Fowler, and J. R. Beckwith. 1977.
- Zabin, I., A. V. Fowler, and J. R. Beckwith. 1977. Position of the mutation in beta-galactosidase ochre mutant U118. J. Bacteriol. 133:437-438.