Proc. Natl. Acad. Sci. USA
Vol. 75, No. 10, pp. 4824–4827, October 1978
Biochemistry

β-Galactosidase chimeras: Primary structure of a *lac* repressor–β-galactosidase protein*

(protein sequencing/hybrid proteins/deletion mutant)

ANTHONY J. BRAKE†‡, AUDREE V. FOWLER†, IRVING ZABIN†§, JURGEN KANIA¶, AND BENNO MÜLLER-HILL¶

† Department of Biological Chemistry, UCLA School of Medicine and Molecular Biology Institute, University of California, Los Angeles, California 90024; and ¶ Institut für Genetik der Universität zu Köln, 5 Weyertal 121, West Germany

Communicated by Emil L. Smith, July 26, 1978

ABSTRACT **A protein possessing both *lac* repressor and β-galactosidase activities in a single polypeptide of about 155,000 daltons was purified from a deletion mutant of *Escherichia coli* in which the *lacI* and *Z* genes are fused. A 77-residue cyanogen bromide peptide containing the fusion joint was isolated. A radioimmunoassay with an antibody prepared against CNBr2 (residues 3–92) of β-galactosidase was used to monitor its purification. The sequence of the joining peptide was determined by analysis of tryptic peptides and by automatic sequencer analysis. The site of joining is from residue 355 of *lac* repressor to residue 24 of β-galactosidase (or 356 to 25), indicating that the last 4 residues at the carboxyl terminus of *lac* repressor and the first 23 residues at the amino terminus of β-galactosidase are not essential for the activities of these two proteins. The exact site of the fusion is not known because *lac* repressor residue 356 and β-galactosidase residue 24 are both leucine residues. Examination of the nucleotide sequences around the two end points of the deletion revealed a homology of 9 identities in a stretch of 11 base pairs.**

Strains of *Escherichia coli* with fusions of the *lacI* and Z genes, coding for *lac* repressor and β-galactosidase, respectively, have been isolated and shown to produce proteins containing a portion of *lac* repressor covalently attached near the amino terminus of β-galactosidase (1). These deletion mutants were found among *lac*$^+$ revertants of strains carrying the *lacZ* U118 nonsense mutation. Because the site of this mutation in the gene corresponds to amino acid residue 17 (2), at least 17 amino acid residues at the amino terminus of β-galactosidase are absent in the repressor–galactosidase hybrids. Several of these strains produce chimeric proteins possessing both *lac* repressor and β-galactosidase activities (1). One of these was purified from strain 71-56-14 and was found to be composed of monomers of about 155,000 daltons containing an amino-terminal sequence identical to that of *lac* repressor (1).

These chimeric proteins have been useful in assessing the role of quaternary structure in the binding of *lac* operator DNA by *lac* repressor and have provided evidence that *lac* repressor recognizes, as a dimer, *lac* operator DNA but uses all four subunits for binding to DNA (3, 4).

We undertook determination of the exact site of the joining in the protein from strain 71-56-14. Because the amino acid sequences of *lac* repressor (5) and of β-galactosidase (6) are known, determination of the sequence of a single peptide overlapping both sequences would be sufficient to answer the question. The problem of locating the correct peptide in an unusually complex mixture (at least 33 cyanogen bromide peptides) was solved by using a peptide radioimmunoassay procedure that had been developed and used previously for the isolation of a tryptic peptide overlapping the sequences of two cyanogen bromide peptides of β-galactosidase (7).

MATERIALS AND METHODS

Isolation of *lac* Repressor–Galactosidase. Strain i^{q1} 71-56-14 [(lac pro) $\Delta F' pro^+ i^{q1} i^+ Z^+$] (1) was grown in minimal medium with 1% glycerol at 35°. The initial steps in the isolation of repressor–galactosidase from 650 g of cells were identical through the ammonium sulfate precipitation to those used for β-galactosidase (8). The 0–35% ammonium sulfate precipitate was then dialyzed against 0.08 M Tris/1 mM $MgCl_2$/1 mM dithiothreitol/0.1 mM EDTA, pH 7.5 and applied to a 2.5 × 40 cm column of DEAE-Bio-Gel A (Bio-Rad) equilibrated with the same buffer. After washing, a linear gradient of 0.01–0.15 M NaCl in 1200 ml of buffer was applied. The main peak of β-galactosidase activity was pooled and the protein was chromatographed again on a 1.5 × 20 cm column of DEAE-Bio-Gel under the same conditions but with a gradient of 500 ml. Enzyme (9) and α-complementation assays (10) were carried out as before.

Radioimmunoassay. The procedure used was similar to that described (7) with several modifications. Magnesium was omitted from all buffers. The tryptic peptide T8 (residues 60–140) of β-galactosidase (11) was labeled with ^{125}I by using lactoperoxidase and then separated from free ^{125}I on a Bio-Gel P2 column equilibrated with buffer containing 0.25% Triton X-100. The peptide-containing fractions were pooled, and bovine serum albumin was added to a concentration of 1 mg/ml before storage at −40°. The radioimmunoassay was then carried out with antiserum prepared against CNBr2 (residues 3–92) of β-galactosidase (12). Unlabeled CNBr2 was used to calibrate the assay. Because T8 contains the 33 carboxyl-terminal residues of CNBr2 and the 48 amino-terminal residues of CNBr3, this assay detects peptides that crossreact with the carboxyl-terminal portion of CNBr2.

Peptide Purification and Sequence Determination. Cyanogen bromide peptides were purified by CM-cellulose chromatography, gel filtration on Sephadex G-50, and sulfopropyl (SP)-Sephadex chromatography as described for β-galactosidase (12–14). Trypsin digestion was performed at a trypsin/peptide ratio of 1:100 (wt/wt) in 0.1 M NH_4HCO_3/2 M urea at 40° for 16 hr, and peptides were separated by gel filtration on Sephadex G-25 in 30% acetic acid followed by paper electrophoresis at pH 1.9. Two tryptic peptides were resolved by preparative thin-layer chromatography on cellulose

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* This work was part of a dissertation submitted by A.J.B. as partial fulfillment for the degree of Doctor of Philosophy.

‡ Present address, Department of Bacteriology and Immunology, University of California, Berkeley, CA 94720.
§ To whom reprint requests should be addressed.

plates (Eastman) developed with butanol/pyridine/acetic acid/water, 15:10:3:9 (vol/vol). Peptides separated by thin-layer chromatography or paper electrophoresis were stained with ninhydrin or by the arginine-specific phenanthrene quinone procedure (15). Peptides were treated with bovine liver pyroglutamate aminopeptidase (Boehringer) in 0.1 M NH_4HCO_3/0.01 M 2-mercaptoethanol at 37° for 24 hr. Automatic sequencer analysis was carried out with a Beckman 890C sequencer using the 0.1 M Quadrol program.

RESULTS

Isolation of *lac*-Repressor-β-Galactosidase. The protein (290 mg) was purified approximately 40-fold from 650 g of cells by ammonium sulfate precipitation and ion-exchange chromatography on DEAE-Bio-Gel A. The yield was 30% based on β-galactosidase activity. The protein was more than 90% pure on sodium dodecyl sulfate gels (16) and migrated with a mobility equal to that of the β-subunit of RNA polymerase, suggesting a subunit molecular weight of about 155,000. The β-galactosidase activity was about 280,000 units/mg compared to 450,000 units/mg for pure β-galactosidase. After correction for the difference in subunit molecular weights (155,000 vs. 116,400), this indicates that β-galactosidase in the hybrid had a specific activity greater than 80% that of wild-type β-galactosidase. A cyanogen bromide digest of *lac* repressor-β-galactosidase was added to an extract of the *lacZ* deletion mutant M15 (12, 17) and was found to have less than 0.1% of the α-donor complementing activity of that in a cyanogen bromide digest of wild-type β-galactosidase.

Isolation of the Joining Peptide. It seemed likely from the size of the chimeric protein that a part of the sequence of CNBr2 (residues 3–92) of β-galactosidase was present in the protein. Therefore, a cyanogen bromide digest of the latter was tested for crossreaction against anti-CNBr2 in a radioimmunoassay (7) and was indeed found to react to an extent approximately equal to that of an equimolar amount of CNBr2. The cyanogen bromide peptides were then separated on CM-cellulose (Fig. 1A) and assayed for crossreactivity in the same way. The two main peaks of immunological activity on CM-cellulose were purified separately on Sephadex G-50. The immunologically active fractions from each column had the same size and nearly identical amino acid compositions, suggesting that they resulted from separation of the homoserine and homoserine lactone forms of the same peptide. When these fractions were combined and purified on SP-Sephadex at pH 2.5, they yielded one peak of activity (Fig. 1B). The elution volume on Sephadex G-50 suggested a size of approximately 80 amino acid residues.

Sequence Determination of the Cyanogen Bromide Joining Peptide. The purified peptide had an amino acid composition similar to that of CNBr2 of β-galactosidase (Table 1). No free amino-terminal residue was detected by the dansyl procedure. This was expected because the carboxyl-terminal cyanogen bromide peptide of *lac* repressor begins with a glutaminyl residue which readily cyclizes to a pyrrolidone carboxyl residue. The cyanogen bromide peptide was treated with trypsin, and eight peptides were purified by Sephadex gel filtration followed by paper electrophoresis. Two peptides were purified further by preparative thin-layer chromatography. The compositions and sequence analyses are presented in Table 1.

Two peptides, T1 and T2, are also present in the carboxyl-terminal cyanogen bromide peptide of *lac* repressor, from residues 348–355. Peptides T4–T8 are identical to tryptic peptides from CNBr2 of β-galactosidase and represent residues

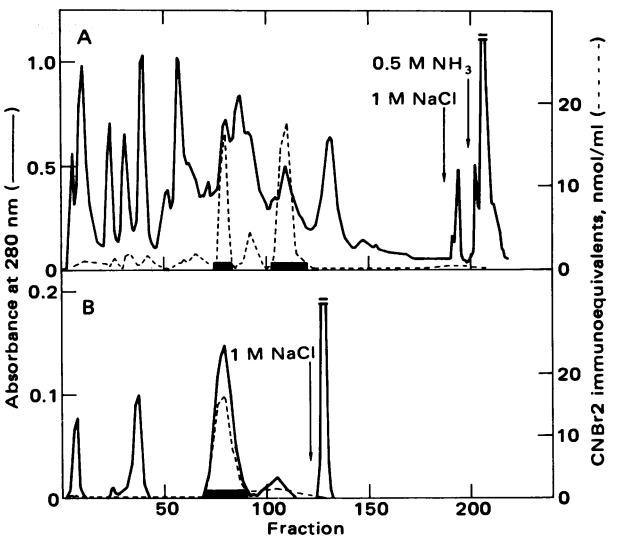

FIG. 1. Purification of *lac* repressor-β-galactosidase cyanogen bromide joining peptide. (*A*) Cyanogen bromide digest (1.3 μmol) of [^{3}H]carboxymethyl *lac* repressor-β-galactosidase was dialyzed against 0.02 M ammonium acetate/8 M urea, pH 5.0, applied to a 1.5 × 20 cm column of CM-cellulose, and eluted with a linear gradient of 0.01–0.08 M NaCl in 1000 ml in the same buffer. Fractions (5 ml) were collected at a flow rate of 25 ml/hr and tested for immunological crossreaction with anti-CNBr2 (dashed line). Fractions containing crossreactivity were pooled (solid bars) and dialyzed before gel filtration on a 2.5 × 200 cm column of Sephadex G-50. (*B*) Crossreacting fractions from the gel filtration step were combined and evaporated, dissolved in 0.01 M ammonium formate/8 M urea, pH 2.5, applied to a 2.5 × 15 cm column of SP-Sephadex, and eluted with a linear gradient of 0.02–0.15 M NaCl in 600 ml of the same buffer. Fractions (5 ml) were collected at a flow rate of 45 ml/hr. Fractions containing crossreactivity were pooled (solid bar), dialyzed, and added to a 2.5 × 30 cm column of Sephadex G-25 in 30% acetic acid.

27–92 of β-galactosidase. One peptide, T3, Leu-Asn-Arg, is not found in tryptic digests of either the *lac* repressor carboxyl-terminal cyanogen bromide peptide or of CNBr2 from β-galactosidase nor in complete tryptic digests of either protein. The presence of Asn rather than Asp was demonstrated by the mobility of T3 in paper electrophoresis at pH 6.5 (20).

These peptides were ordered from the known sequences of the *lac* repressor and β-galactosidase. The order was confirmed by treatment of the cyanogen bromide peptide with pyroglutaminyl aminopeptidase which gave pyrrolidone carboxylic acid-free peptide in low yield but sufficient for automatic sequencer analysis.

The sequence is shown in Fig. 2. The peptide contains 77 residues and joins *lac* repressor residue 355 to β-galactosidase residue 24 (or 356 to 25).

DISCUSSION

The chimeric repressor-galactosidase used for the sequence analysis described here had nearly the full activities of the wild-type repressor and β-galactosidase. It was purified by a procedure that has been generally useful for the isolation of β-galactosidase-related proteins from a number of *lacZ* gene mutants of *E. coli*. Use of a peptide radioimmunoassay allowed the identification of fractions containing the desired joining peptide in a mixture of at least 33 peptides expected in a cyanogen bromide digest of repressor-β-galactosidase.

Studies of this peptide showed that it was distinct from any present in *lac* repressor or β-galactosidase. The actual site of joining was determined to be from Arg-355 of *lac* repressor to Leu-24 of β-galactosidase or from Leu-356 of *lac* repressor to

Table 1. Analyses of *lac* repressor–β-galactosidase cyanogen bromide joining peptide and derived tryptic peptides

Amino acid	CNBr peptide[a]	Tryptic peptides[a] T1	T2	T3	T4	T5	T6	T7	T8
Tryptophan[b]	2.1 (5)				(1)			0.3 (1)	2.5 (3)
Lysine					0.6				0.2
Histidine	1.1 (1)				0.7 (1)		0.2		
Arginine	7.4 (8)	0.9 (1)	1.0 (1)	1.0 (1)	1.1 (1)	0.7 (1)	1.9 (2)	0.8 (1)	
CM-cysteine	1.0 (1)								1.0 (1)
Aspartic acid	6.4 (7)		0.2	1.1 (1)	0.3	0.9 (1)	1.0 (1)	1.0 (1)	3.1 (3)
Threonine	2.4 (2)				0.4		0.9 (1)		1.2 (1)
Serine	6.2 (7)	0.2	0.9 (1)		1.1 (1)	0.9 (1)	1.0 (1)	0.8 (1)	2.0 (2)
Glutamic acid	10.9 (12)	1.1 (1)	1.1 (1)		0.7	2.0 (2)	2.0 (2)	1.1 (1)	4.7 (5)
Proline	6.7 (8)				2.3 (2)		1.2 (1)		4.4 (5)
Glycine	2.2 (1)	0.2	0.3	0.4	0.6	0.2		1.1 (1)	0.5
Alanine	8.2 (9)	1.0 (1)			2.6 (3)	1.0 (1)	0.4		3.8 (4)
Valine	4.4 (5)		0.9 (1)						3.6 (4)
Isoleucine	0.7				0.2				0.4
Leucine	7.5 (7)	0.9 (1)	0.2	0.9 (1)	1.0 (1)		1.0 (1)	0.8 (1)	2.1 (2)
Tyrosine	0.4								0.2
Phenylalanine	2.7 (3)				0.5 (1)				1.6 (2)
Homoserine[c]	0.5 (1)								0.4 (1)
Amino terminus[d]	—	—	Gln	Leu	Leu	Asn	Thr	Ser	Phe
Total residues	77	4	4	3	11	6	9	7	33
Yield, %[e]	22	19	14	35	10	21	29	30	53
Purification steps[f]		G-25, PE, TLC	G-25, PE, TLC	G-25, PE	G-25, PE	G-25,PE	G-25,PE	G-25,PE	G-25, G-50
Residues sequenced[g]	2–12[h]		1–4	1–3	1–6	1–6	1–8	1–7	1–11

[a] Number in parentheses is the integral number of residues based on analysis of tryptic peptides and previously determined sequence of CNBr2 of β-galactosidase (12).
[b] Determined from A_{280} and analysis of tryptic peptides.
[c] Homoserine plus homoserine lactone.
[d] Determined by dansyl chloride procedure (18).
[e] Yield based on 1.3 μmol of cyanogen bromide digest and 210 nmol of tryptic digest.
[f] Purification steps were: G-25, Sephadex G-25 chromatography in 30% acetic acid; G-50, Sephadex G-50 chromatography in 30% acetic acid; PE, paper electrophoresis at pH 1.9; TLC, thin-layer chromatography on cellulose plates.
[g] Sequence determination was by the dansyl chloride-Edman procedure of Gray (19).
[h] Partial data from automatic sequencer analysis of 50 nmol of the pyroglutamate aminopeptidase-treated peptide.

Asn-25 of β-galactosidase, resulting in a polypeptide chain 1353 residues long. The uncertainty results from the fact that the leucine residue could correspond to a leucine present in either *lac* repressor or β-galactosidase.

The essentially full activities of the *lac* repressor–β-galactosidase chimera shown in this and previous studies (1) demonstrates that at least 4 residues at the carboxyl terminus of *lac* repressor and 23 residues at the amino terminus of β-galactosidase are not essential for the activities of either protein. These residues are thus unlikely to be buried within the interior of the

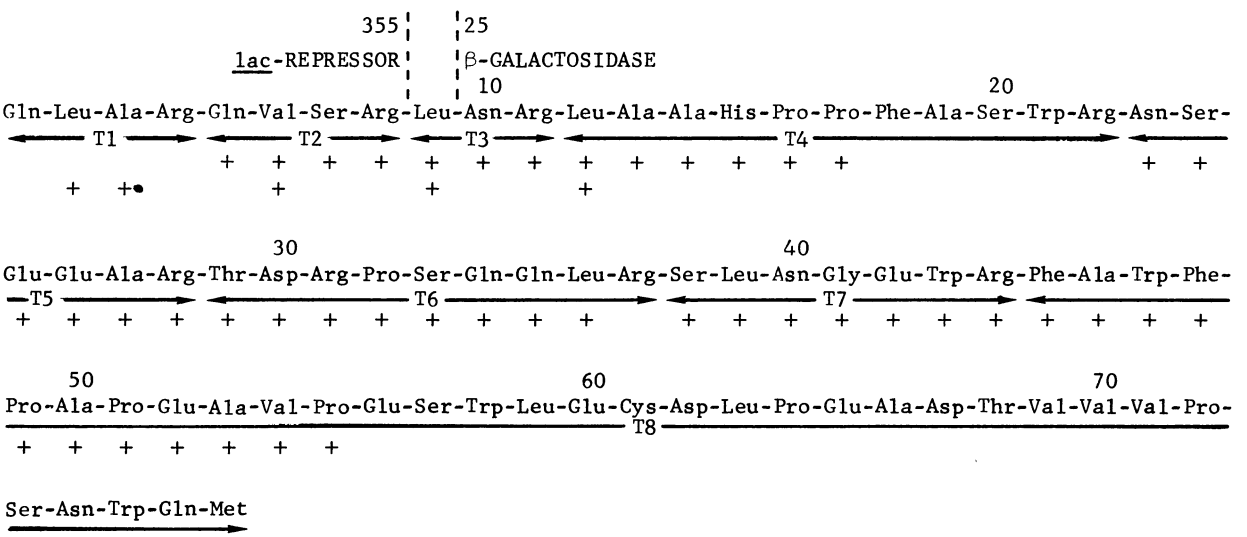

FIG. 2. Amino acid sequence of the cyanogen bromide joining peptide of *lac* repressor–β-galactosidase. Each of the two possible sites of joining are indicated by the dotted lines. Tryptic peptides isolated in this study are indicated by arrows beneath the sequence. +, Residues identified by dansyl-Edman analysis of tryptic peptides (upper row) or sequencer analysis of the pyroglutamate aminopeptidase-treated CNBr peptide (lower row).

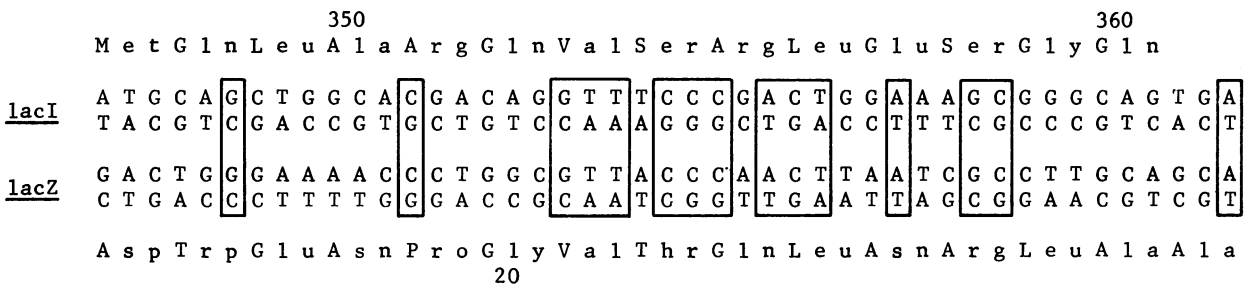

FIG. 3. Nucleotide sequences in the region of the deletion end points of mutant 71-56-14. Nucleotide sequences of the *lacI* (P. J. Farabaugh, personal communication) and *lacZ* (A. Maxam and W. Gilbert, personal communication) genes were aligned to conform to the site of fusion determined in this study. The corresponding amino acid sequences are indicated above the *lacI* sequence and below the *lacZ* sequence. Nucleotide identities are enclosed in boxes.

folded proteins because their replacement by unrelated sequences does not interfere with the proper folding of the *lac* repressor or β-galactosidase parts of this chimeric protein. Immunological studies of β-galactosidase and the *lacZ* deletion mutant M15 (F. Celada and I. Zabin, unpublished data) suggest, however, that at least part of CNBr2, residues 3–92, is buried in a subunit interface in β-galactosidase. The full activity of the chimera suggests that only part of the sequence beyond residue 24 can be contributing to the subunit interface.

Chimeric *lac* repressor–β-galactosidase has been used previously to demonstrate that the association of only two *lac* repressor subunits is sufficient to recognize *lac* operator DNA. The dispensibility of a portion of the carboxyl-terminal region of *lac* repressor is consistent with the lack of nonsense mutants mapping near the end of the *I* gene (21). Mapping of other deletions that allow retention of *lac* repressor or β-galactosidase activities should further delineate which amino acids are essential for their respective activities.

The ambiguity concerning the exact point of fusion of the *lacI* and Z genes in this mutant strain as determined by the amino acid sequence prompted us to examine the nucleotide sequences around the possible end points of this deletion. The nucleotide sequences of each gene (P. J. Farabaugh, personal communication; A. Maxam and W. Gilbert, personal communication) were aligned, with the results shown in Fig. 3. There is a striking homology of 9 identities in an 11-base-pair sequence between the nucleotide sequences at positions 1057–1067 of *lacI* and 61–71 of *lacZ*. This raises the possibility that these small homologous regions may have been involved in the intramolecular recombinational event necessary to cause a deletion. Previous experiments that compared the frequency of *I-Z* fusions in *recA*$^+$ and *recA*$^-$ strains of *E. coli* have suggested that these deletions did not require a functional *recA*-recombination pathway (22). Perhaps small regions of homology may be responsible for the high frequency of these *I-Z* fusions, but a *recA*-independent pathway is involved. Examination of the nucleotide sequences involved in a number of similar gene fusions can test the possibility that base sequence homologies around their end points are a common feature. DNA sequence analysis of a number of spontaneous mutants with deletions within the *lacI* gene has demonstrated the frequent occurrence of base sequence repeats around these deletion end points (P. J. Farabaugh, U. Schmeissner, M. Hofer, and J. H. Miller, personal communication).

We thank Val Schaeffer, George Collins, and Paulette Osborne for excellent technical assistance. This work was supported in part by Grant AI 04181 and National Research Award 1 T32 GM 07104 from the U.S. Public Health Service.

1. Müller-Hill, B. & Kania, J. (1974) *Nature (London)* **249**, 561–563.
2. Zabin, I., Fowler, A. V. & Beckwith, J. R. (1978) *J. Bacteriol.* **133**, 437–438.
3. Kania, J. & Brown, D. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3529–3533.
4. Kania, J. & Müller-Hill, B. (1977) *Eur. J. Biochem.* **79**, 381–386.
5. Beyreuther, K., Adler, K., Fanning, E., Murray, C., Klemm, A. & Geisler, N. (1975) *Eur. J. Biochem.* **59**, 491–509.
6. Fowler, A. V. & Zabin, I. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1507–1510.
7. Brake, A. J., Celada, F., Fowler, A. V. & Zabin, I. (1977) *Anal. Biochem.* **80**, 108–115.
8. Fowler, A. V. (1972) *J. Bacteriol.* **112**, 856–860.
9. Horiuchi, T., Tomezawa, J. & Novick, A. (1962) *Biochim. Biophys. Acta* **55**, 152–163.
10. Lin, S. & Zabin, I. (1972) *J. Biol. Chem.* **247**, 2205–2211.
11. Fowler, A. V., Brake, A. J. & Zabin, I. (1978) *J. Biol. Chem.* **253**, 5490–5498.
12. Langley, K. E., Fowler, A. V. & Zabin, I. (1975) *J. Biol. Chem.* **250**, 2587–2592.
13. Fowler, A. V. (1975) in *Solid Phase Methods in Protein Sequence Analysis*, ed. Laursen, R. A. (Pierce Chemical Co., Rockford, IL), pp. 169–177.
14. Fowler, A. V. (1978) *J. Biol. Chem.* **253**, 5499–5504.
15. Yamada, S. & Itano, H. A. (1966) *Biochim. Biophys. Acta* **130**, 538–540.
16. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
17. Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J. & Zabin, I. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1254–1257.
18. Gray, W. R. (1972) *Methods Enzymol.* **25**, 121–138.
19. Gray, W. R. (1972) *Methods Enzymol.* **25**, 333–344.
20. Offord, R. E. (1966) *Nature (London)* **211**, 591–593.
21. Miller, J. H., Ganem, D., Lu, P. & Schmitz, A. (1977) *J. Mol. Biol.* **109**, 275–301.
22. Müller-Hill, B., Heidecker, G. & Kania, J. (1976) in *Proceedings of the Third John Innes Symposium*, ed. Markham, R. & Horne, R. W. (North Holland, Amsterdam, Netherlands), pp. 167–179.