Proc. Nat. Acad. Sci. USA
Vol. 70, No. 12, Part I, pp. 3576-3580, December 1973

The Amino-Acid Sequence of *lac* Repressor

(*E. coli*/regulation/protein-DNA interaction)

KONRAD BEYREUTHER, KLAUS ADLER, NORBERT GEISLER, AND ALEX KLEMM*

Institut für Genetik der Universität zu Köln, 5 Köln 41, Germany (FRG)

Communicated by Jacques Monod, July 20, 1973

ABSTRACT **The amino-acid sequence of *lac* repressor from *Escherichia coli* has been determined. The sequence contains 347 residues in the subunit single peptide chain. It shows no similarities with the sequences of histones or the known part of β-galactosidase.**

If we want to understand repression (1), we have to elucidate the three-dimensional structure of a repressor–operator complex. The *lac* system of *E. coli* offers the possibility of such an analysis. *Lac* repressor has been isolated (2). It binds *in vitro* to operator DNA (3, 4) and represses *in vitro* transcription (5) and translation (6) of the *lac* operon. It can be isolated in large amounts (7, 8). Genetic analysis of mutants of the *lac* repressor-producing *i* gene has shown that the amino-terminus of repressor is involved in operator binding (9). The *lac* operator has also been analyzed genetically (10) and chemically: Gilbert succeeded recently in determining its structure (11). We report here the sequence of *lac* repressor. The sequence of its amino-terminus has been presented earlier (9, 12–14).

RESULTS

The amino-acid composition of *lac* repressor is given in Table 1. The amino-acid sequence is shown in Fig. 1. It was determined by analysis of fragments derived by treating the protein with cyanogen bromide; for the ordering of the fragments, tryptic peptides of repressor were used.

Lac repressor was purified from i^q and i^{pq} strains (7, 13) as described earlier (8). Gel electrophoresis in the presence of sodium dodecyl sulfate and endgroup determinations of repressor had indicated the presence of a single polypeptide chain with a molecular weight of 38,000–40,000. On the basis of the amino-acid composition, a molecular weight of 38,000 was calculated (8). The amino-acid sequence gives a molecular weight of 37,200 for the subunit and 148,800 for the functional tetramer.

Cyanogen Bromide Fragments. 10 Fragments were obtained after *lac* repressor was cleaved with cyanogen bromide (Table 2). The cyanogen bromide digest was chromatographed on a Sephadex G-75 column equilibrated with 1 M acetic acid (Fig. 2). The first peak, in the break-through volume, contained partially cleaved repressor and was not further analyzed. Peak II consists of an "overlap peptide" which proved to be fragment V joined to fragment III. Peak III contained the largest fragment with the two tryptophan residues of the chain. Several recyclings were necessary to separate it from peak IV. Fragments V and VI could be easily separated from the fraction of the small peptides. They correspond to sequences 2–42 (FVI) and 43–98 (FV) (9). Cyanogen bromide fragments VII–XII (including free homoserine) could be separated on DEAE-cellulose with pyridine–acetate buffers at pH 6.5. Gel electrophoresis of the isolated fragments II–VI in Na dodecyl sulfate and endgroup analysis established their molecular weight and their homogeneity. The molecular weights of fragments VII–XII could be determined with the method of Offord (16). The sum of the molecular weights of the ten fragments and the sum of the compositions of the ten fragments agreed with the molecular weight and the amino-acid composition of intact *lac* repressor (Tables 1 and 2).

Ordering the Cyanogen Bromide Fragments. Tryptic peptides were used to order the cyanogen bromide peptides. Thirty tryptic peptides were isolated after chromatography on high-resolution ion-exchange resins, gel chromatography, thin-layer electrophoresis, and thin-layer chromatography of a tryptic digest of carboxymethylated *lac* repressor. Five of these peptides contained methionine. The sequence of these peptides (Table 3) provided the overlaps that allowed the cyanogen bromide fragments to be ordered. The order of the cyanogen bromide fragments is: FXII-FVI-FV-FIII-FX-FVIII-FIX-FXI-FIV-FVII. The sequence of the other 25 tryptic peptides has also been determined.

Sequence Analysis of the Cyanogen Bromide Fragments. The structure of cyanogen bromide fragments VII–XII was determined without further enzymatic digestions. Cyanogen bromide fragments II–VI were digested with various proteolytic enzymes (Fig. 3*A–D*). The peptide mixtures thus obtained were fractionated by the methods described for separation of

TABLE 1. *Amino-acid composition of* lac *repressor*

Amino acid	No. of residues	Amino acid	No. of residues
Lysine	11	Glycine	22
Histidine	7	Alanine	44
Arginine	19	Cysteine	3
Aspartic Acid	15	Valine	33
Asparagine	11	Methionine	9
Threonine	18	Isoleucine	17
Serine	30	Leucine	40
Glutamic acid	13	Tyrosine	8
Glutamine	28	Phenylalanine	4
Proline	13	Tryptophan	2
Total no. of residues		347	

Abbreviation: Hsr, homoserine.

* Deceased March 13, 1972.

TABLE 2. *Characterization of cyanogen bromide fragments of* lac *repressor*

Cyanogen bromide fragment	Molecular weight: SDS gel (± 10%)	Molecular weight: Electrophoretic mobility (± 5%)	No. of amino acids	NH_2-Terminus	COOH-Terminus	Sequences
FI	30,000–35,000	—	—	Lys,Ala,Val	Hsr,Gln	—
FII	20,000	—	170	Ala	Hsr	43–212
FIII	14,000	—	114	Val	Hsr	99–212
FIV	11,000	—	93	Arg	Hsr	242–334
FV	5,800	—	56	Ala	Hsr	43–98
FVI	4,500	—	41	Lys	Hsr	2–42
FVII	—	1,500	13	Gln	Gln	335–347
FVIII	—	1,100	10	Leu	Hsr	220–229
FIX	—	750	7	Leu	Hsr	230–236
FX	—	750	7	Ser	Hsr	213–219
FXI	—	450	5	Ala	Hsr	237–241
FXII	—	—	1	Hsr		1
Lac repressor	38,000–40,000	—	347	Met	Gln	1–347

SDS, sodium dodecyl sulfate; *Hsr*, homoserine.

the tryptic peptides. The sequences of the purified peptides were determined by the most appropriate current techniques:

```
Met-Lys-Pro-Val-Thr-Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-
                                                        10
Ser-Tyr-Gln-Thr-Val-Ser-Arg-Val-Val-Asn-Gln-Ala-Ser-His-Val-
        20                                              30
Ser-Ala-Lys-Thr-Arg-Glu-Lys-Val-Glu-Ala-Ala-Met-Ala-Glu-Leu-
                                                        40
Asn-Tyr-Ile-Pro-Asn-Arg-Val-Ala-Gln-Gln-Leu-Ala-Gly-Lys-Gln-
        50                                              60
Ser-Leu-Leu-Ile-Gly-Val-Ala-Thr-Ser-Ser-Leu-Ala-Leu-His-Ala-
                                                        70
Pro-Ser-Gln-Ile-Val-Ala-Ala-Ile-Lys-Ser-Arg-Ala-Asp-Gln-Leu-
        80                                              90
Gly-Ala-Ser-Val-Val-Val-Ser-Met-Val-Glu-Arg-Ser-Gly-Val-Glu-
                                                       100
Ala-Cys-Lys-Ala-Ala-Val-His-Asn-Leu-Leu-Ala-Gln-Arg-Val-Ser-
       110                                             120
Gly-Leu-Ile-Ile-Asn-Tyr-Pro-Leu-Asp-Asp-Gln-Asp-Ala-Ile-Ala-
                                                       130
Val-Glu-Ala-Ala-Cys-Thr-Asn-Val-Pro-Ala-Leu-Phe-Ile-Ile-Phe-
       140                                             150
Ser-His-Gln-Asp-Gly-Thr-Arg-Leu-Gly-Val-Glu-His-Leu-Val-Ala-
                                                       160
Leu-Gly-His-Gln-Gln-Ile-Ala-Leu-Leu-Ala-Gly-Pro-Leu-Ser-Ser-
       170                                             180
Val-Ser-Ala-Arg-Leu-Arg-Leu-Ala-Gly-Trp-His-Lys-Tyr-Leu-Thr-
                                                       190
Arg-Asn-Gln-Ile-Gln-Pro-Ile-Ala-Gln-Arg-Glu-Gly-Asp-Trp-Ser-
       200                                             210
Ala-Met-Ser-Gly-Phe-Gln-Gln-Thr-Met-Leu-Asn-Glu-Gly-Ile-Val-
                                                       220
Pro-Thr-Ala-Met-Leu-Val-Ala-Asn-Asp-Gln-Met-Ala-Leu-Gly-Ala-
       230                                             240
Met-Arg-Ala-Ile-Thr-Glu-Ser-Gly-Leu-Arg-Val-Gly-Ala-Asp-Ile-
                                                       250
Ser-Val-Val-Gly-Tyr-Asp-Asp-Thr-Glu-Asp-Ser-Ser-Cys-Tyr-Ile-
       260                                             270
Pro-Pro-Leu-Thr-Thr-Ile-Lys-Gln-Asp-Phe-Arg-Leu-Leu-Gly-Gln-
                                                       280
Thr-Ser-Val-Asp-Arg-Leu-Leu-Gln-Leu-Ser-Gln-Gly-Gln-Ala-Val-
       290                                             300
Lys-Gly-Asn-Gln-Leu-Leu-Pro-Val-Ser-Leu-Val-Lys-Arg-Lys-Thr-
                                                       310
Thr-Leu-Ala-Pro-Asn-Thr-Gln-Thr-Ala-Ser-Pro-Arg-Ala-Leu-Ala-
       320                                             330
Asp-Ser-Leu-Met-Gln-Leu-Ala-Arg-Gln-Val-Ser-Arg-Leu-Glu-Ser-
                                                       340
Gly-Gln
  347
```

FIG. 1. The amino-acid sequence of *lac* repressor from *E. coli*.

manual and automatic Edman degradation (17, 18), dansyl-Edman procedure (19), and digestions with carboxy- and aminopeptidases and dipeptidyl aminopeptidase.

First, the carboxymethylated cyanogen bromide fragments II–VI were digested with trypsin. Since the sequence of the tryptic peptides from carboxymethylated *lac* repressor had been previously determined by us, small amounts (1–2 μmol) of the cyanogen bromide fragments were sufficient for the isolation of tryptic peptides to provide the necessary information. Only for fragment V did the isolated tryptic peptides not provide enough information. They accounted for only 31 of 56 amino acids. The sequence of the missing tryptic peptide was derived by examining all chymotryptic and thermolytic peptides of fragment V (Fig. 3*C*). In order to find the overlaps for the tryptic peptides, chymotryptic and thermolytic pep-

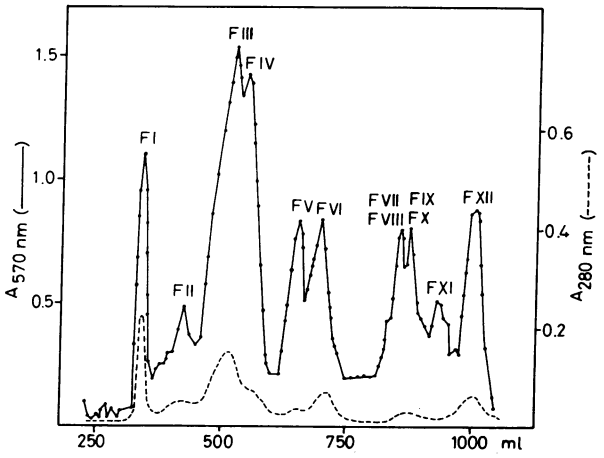

FIG. 2. Separation of the cyanogen bromide fragments of *lac* repressor on a 2 × 200-cm column of Sephadex G-75 equilibrated with 1.0 M acetic acid. The digest of 7.5 μmol (300 mg) was dissolved in 10 ml of 2.0 M acetic acid; 2-ml portions were applied to the column at once and eluted with 1.0 M acetic acid. The column was developed at 22° at a rate of 30 ml/hr. The effluent was monitored both at 280 nm (*broken line*) and by ninhydrin analysis after alkaline hydrolysis of 10-μl aliquots (*solid line*).

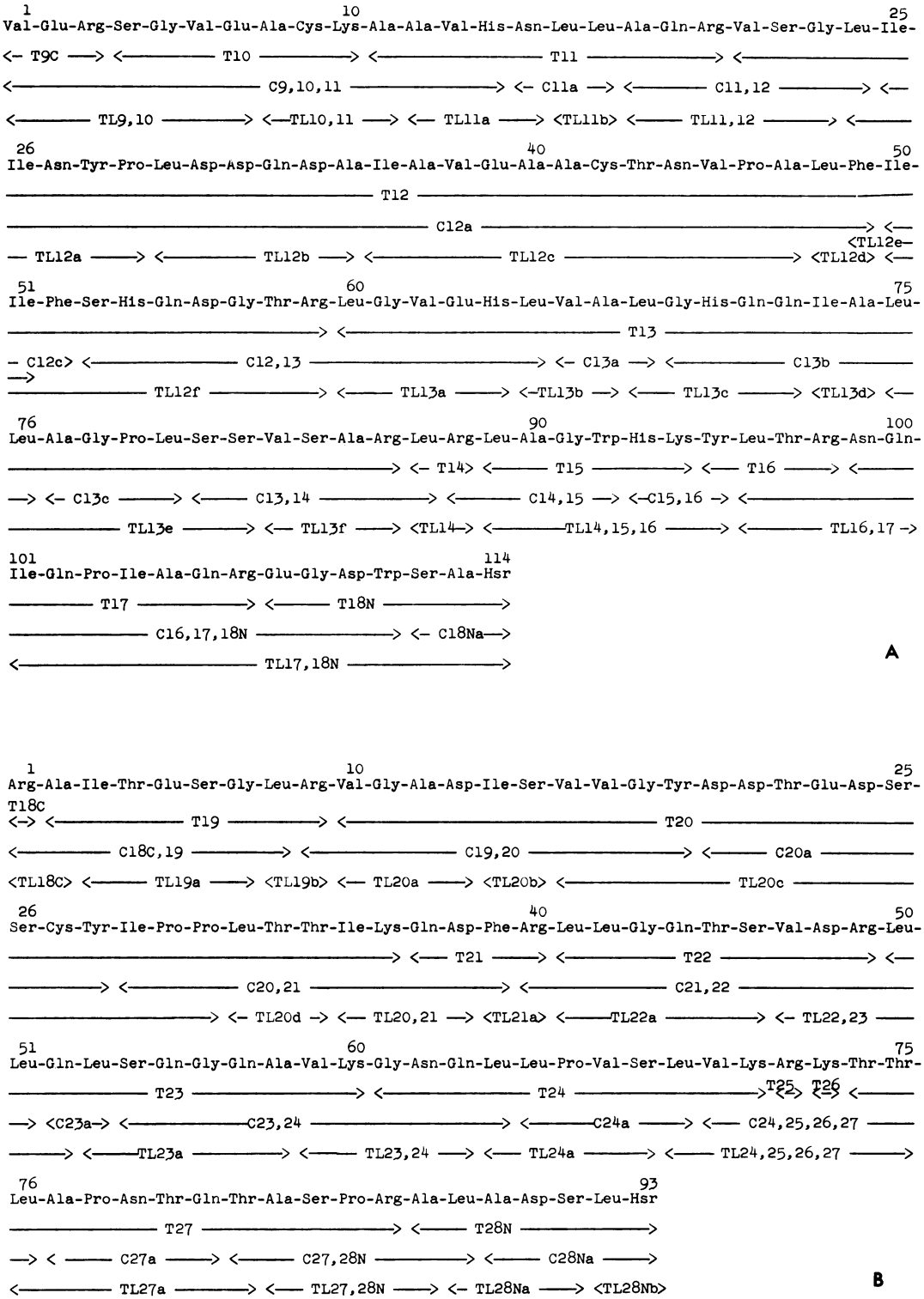

FIG. 3. The amino-acid sequence of cyanogen bromide fragments III (*A*), IV (*B*), V (*C*), and VI (*D*) and alignment of peptides after hydrolysis with trypsin (*T*), chymotrypsin (*C*), and thermolysin (*TL*).

tides were isolated from the cyanogen bromide fragments. Their alignment is shown in Fig. 3*A–D*. The amide groups of asparagine and glutamine residues were located by measuring the electrophoretic mobilities of peptides containing them (16).

DISCUSSION

There is no obvious feature in the sequence of *lac* repressor that could explain its tight and specific binding to *lac* operator. However, genetic analysis of the *i*-gene mutants has shown that the amino-terminus is involved in operator-DNA bind-

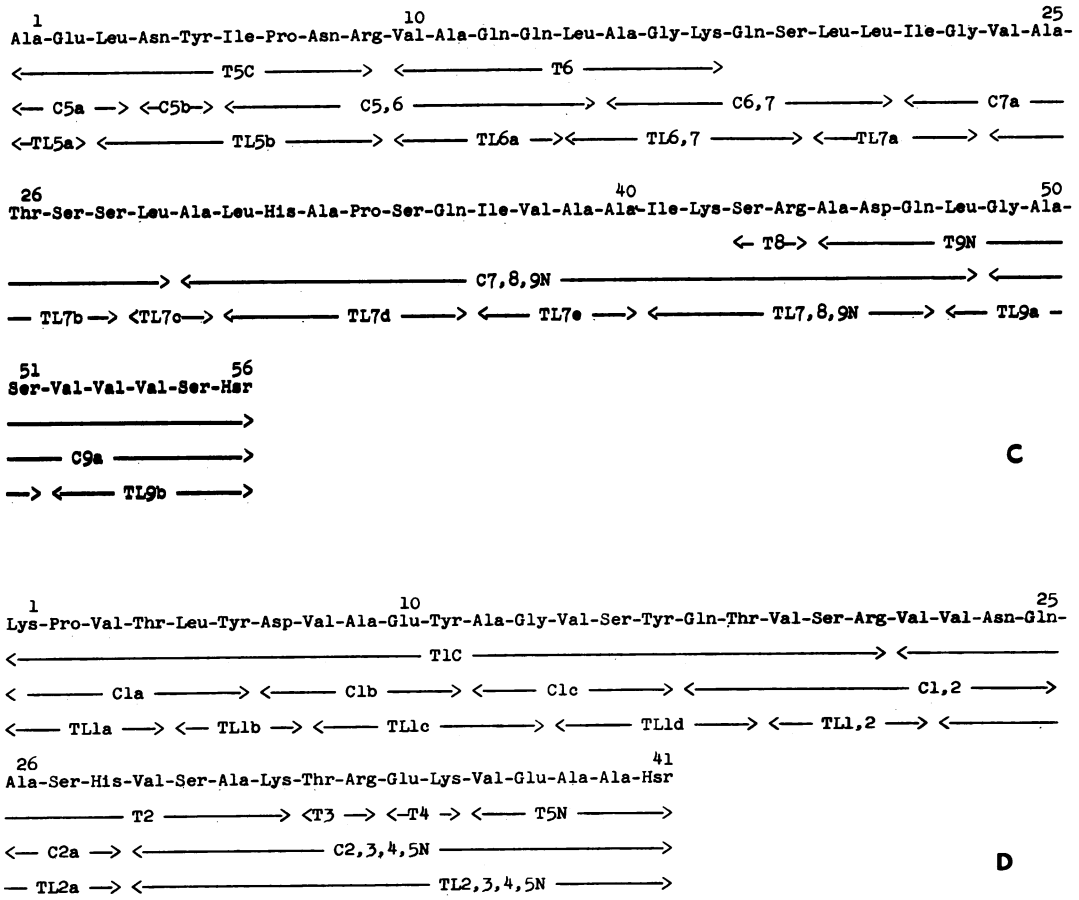

FIG. 3. (continued)

ing (9). The sequence does not resemble the sequence of histones (20). It may be pointed out that 50% of the tyrosine residues are found between sequences 1 and 50, the region which according to genetic analysis is involved in operator binding. Furthermore the two tryptophans occupy positions (190 and 209) in a region which is probably involved in inducer binding (21).

We found considerable deamidation of Asn25, Gln54, Gln55, Gln131, and Gln153, but of no other glutamines or asparagines. Asp154 could possibly have been an asparagine since extensive deamidation is known to occur at asparagine residues followed by a glycine residue (22, 23). Asn25 has been invoked in operator binding (9), and a similar involvement of Gln54 and Gln55 seems possible. Pure repressor has the annoying property of being rather inactive with regard to operator binding (8). Deamidation of asparagines and glutamines that are involved in operator-DNA recognition could explain the loss of operator-binding activity.

TABLE 3. *Sequence of tryptic peptides containing methionine*

Peptide	Overlap for cyanogen bromide fragments	Sequences
T 1	FXII–FVI	*Met*-Lys-Pro-Val-Thr-Leu-Tyr-Asp-Val-Ala-Glu-Tyr-Ala-Gly-Val-Ser-Tyr-Gln-Thr-Val-Ser-Arg (10 … 22)
T 5	FVI–FV	Val-Glu-Ala-Ala-*Met*-Ala-Glu-Leu-Asn-Tyr-Ile-Pro-Asn-Arg (14)
T 9	FV–FIII	Ala-Asp-Gln-Leu-Gly-Ala-Ser-Val-Val-Val-Ser-*Met*-Val-Glu-Arg (15)
T 18	FIII–FX–	Glu-Gly-Asp-Trp-Ser-Ala-*Met*-Ser-Gly-Phe-Gln-Gln-Thr-*Met*-Leu- (15)
	FVIII–FIX–	Asn-Glu-Gly-Ile-Val-Pro-Thr-Ala-*Met*-Leu-Val-Ala-Asn-Asp-Gln- (30)
	FXI–FIV	*Met*-Ala-Leu-Gly-Ala-*Met*-Arg (37)
T 28	FIV–FVII	Ala-Leu-Ala-Asp-Ser-Leu-*Met*-Gln-Leu-Ala-Arg (11)

Platt *et al.* (13) have shown that trypsin and chymotrypsin readily destroy operator binding of native *lac* repressor, leaving inducer binding intact. The trypsin-resistant core lacks 79 residues (14). 59 Residues are derived from the amino-terminus according to Platt *et al.* The present sequence shows that the remaining 20 residues originate from the carboxyl-terminus. In the same context, we would like to point out that our previous determination of the carboxyl-terminus with carboxypeptidase B (9) was wrong: the Arg-Lys sequence, which we erroneously believed to be carboxyl-terminal, was probably produced from the internal Lys-Arg-Lys (312–314) sequence by trypsin contaminating our carboxypeptidase B.

Did β-galactosidase and *lac* repressor evolve from a common ancestor? We have shown that antibody against β-galactosidase does not react with *lac* repressor in an Ouchterlony test. Similarly there is no crossreaction between antibody against repressor and β-galactosidase. Since about 30% of the sequence of β-galactosidase is known (24), a comparison of the sequences cannot yet give the final answer. Visual inspection of the published 350 sequences of β-galactosidase did not show any homology with the sequence of *lac* repressor.

We thank Ellen Fanning for her help, Carolyn Murray, and Santosh Pinto for excellent assistance, B. Müller-Hill for giving us the opportunity to conduct this work and for his continual interest, and G. Braunitzer, B. Gutte, and B. S. Hartley for discussions. Part of the equipment was provided by Stiftung Volkswagenwerk. The work was supported by grants from Deutsche Forschungsmeinschaft through the SFB 74 to K.B. and B.M.H.

1. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3**, 318–356.
2. Gilbert, W. & Müller-Hill, B. (1966) *Proc. Nat. Acad. Sci. USA* **56**, 1891–1898.
3. Gilbert, W. & Müller-Hill, B. (1967) *Proc. Nat. Acad. Sci. USA* **58**, 2415–2421.
4. Riggs, A. D., Bourgeois, S., Newby, R. F. & Cohn, M. (1968) *J. Mol. Biol.* **34**, 365–368.
5. Chen, B., de Crombrugghe, B., Anderson, W. B., Gottesman, M. E. & Pastan, I. (1971) *Nature New Biol.* **233**, 67–70.
6. Zubay, G., Chambers, D. A. & Cheong, L. C. (1970) in *The Lactose Operon*, eds. Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor, N.Y.), pp. 375–391.
7. Müller-Hill, B., Crapo, L. & Gilbert, W. (1968) *Proc. Nat. Acad. Sci USA* **59**, 1259–1264.
8. Müller-Hill, B., Beyreuther, K. & Gilbert, W. (1971) in *Methods in Enzymology*, eds. Grossman, L. & Moldave, K. (Academic Press, New York), **21**, part D, pp. 483–487.
9. Adler, K., Beyreuther, K., Fanning, E., Geisler, N., Gronenborn, B., Klemm, A., Müller-Hill, B., Pfahl, M. & Schmitz, A. (1972) *Nature* **237**, 322–327.
10. Sadler, J. R. & Smith, T. F. (1971) *J. Mol. Biol.* **62**, 139–169.
11. Gilbert, W. & Maxam, A. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 3581–3585.
12. Platt, T., Weber, K., Ganem, D. & Miller, J. H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 897–901.
13. Weber, K., Platt, T., Ganem, D. & Miller, J. H. (1972) *Proc. Nat. Acad. Sci. USA* **69**, 3624–3628.
14. Platt, T., Files, J. G. & Weber, K. (1973) *J. Biol. Chem.* **248**, 110–121.
15. Miller, J. H., Platt, T. & Weber, K. (1970) in *The Lactose Operon*, eds. Beckwith, J. R. & Zipser, D. (Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor, N.Y.), pp. 343–351.
16. Offord, R. E. (1966) *Nature* **211**, 591–593.
17. Sjöquist, J. (1960) *Biochim. Biophys. Acta* **41**, 20–30.
18. Edman, P. & Begg, G. (1967) *Eur. J. Biochem.* **1**, 80–91.
19. Hartley, B. S. (1970) *Biochem. J.* **119**, 805–822.
20. DeLange, J. & Smith, E. L. (1971) in *Annual Review of Biochemistry*, ed. Snell, E. E. (Annual Reviews, Inc., Palo Alto, Calif.), Vol. **40**, pp. 279–314.
21. Pfahl, M. (1971) Thesis, University of Cologne.
22. Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378.
23. Shotton, D. M. & Hartley, B. S. (1970) *Nature* **225**, 802–806.
24. Fowler, A. V. (1972) *J. Biol. Chem.* **247**, 5425–5431.