*Proc. Natl. Acad. Sci. USA*
Vol. 74, No. 3, pp. 966–970, March 1977
Biochemistry

# Minimal length of the lactose operator sequence for the specific recognition by the lactose repressor*

(synthetic lactose operator/lactose operator-repressor interaction/protein-DNA interaction/gene regulation/oligonucleotide synthesis)

CHANDER P. BAHL†, RAY WU†§, JACEK STAWINSKY‡, AND SARAN A. NARANG‡

† Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853; and ‡ Division of Biological Sciences, National Research Council of Canada, Ottawa, Canada, KIA OR6

*Communicated by Jerrold Meinwald, December 27, 1976*

ABSTRACT **A number of specific duplex DNA sequences which correspond to varying lengths of the lactose operator region have been synthesized by a combination of chemical and enzymatic methods. Repressor binding studies on these synthetic duplex operator molecules show that all the nucleotides essential for full lactose operator-repressor interactions are included in a 17-nucleotide-long duplex DNA that constitutes the minimal recognition sequence for this DNA–protein interaction.**

The regulation of the *lac* operon in *Escherichia coli* is controlled by the *lac* repressor protein, which binds specifically to the *lac* operator DNA sequence (4, 5). The nucleotide sequence in the region of the *lac* operator (Fig. 1, structure *A*) has been determined by Gilbert and Maxam (8), and an extended sequence (Fig. 1, structure *B*) has been reported by Dickson *et al.* (6). We (7) have synthesized a 21-nucleotide-long *lac* operator duplex DNA (Fig. 1, structure *C*) and have shown that this sequence binds specifically to the *lac* repressor. Our choice of this particular sequence was based on the fact that it contained the region of 2-fold symmetry within the 24-nucleotide operator sequence of Gilbert and Maxam (8) and on our belief that a 21-nucleotide sequence should be more than sufficient for specificity of recognition and stability of the operator-repressor complex (9). Recently, it has been shown that the 21-nucleotide synthetic operator is functional not only *in vitro* (7) but also *in vivo* (2, 3, 10). We have now determined the minimal recognition sequence of the *lac* operator. In this communication we show that a 17-nucleotide stretch of duplex DNA (Fig. 1 structure *D*) has all the essential features for the specific recognition by the *lac* repressor protein in an *in vitro* assay system.

## MATERIALS AND METHODS

The source of all chemicals and reagents for oligonucleotide synthesis has been described in an earlier publication (11).

Snake venom phosphodiesterase (oligonucleate 5′-nucleotidohydrolase, EC 3.1.4.1) was purchased from Worthington Biochemical Co. and purified before use (12). Lactose repressor was purified according to Platt *et al.* (13) or was a gift of Drs. Arthur Riggs and John Rosenberg. T4 DNA ligase (14, 15) was purified by published procedures. The RNA-dependent DNA polymerase (AMV reverse transcriptase) was kindly provided by Dr. Joseph Beard of Life Sciences Inc. The procedures for characterization and labeling of DNA have been described in detail by Wu *et al.* (16).

**Chemical Synthesis of Oligonucleotides.** All the oligonucleotides used in this study were synthesized by the modified phosphotriester method developed in our laboratory (7, 11, 17–19). The synthesis of the oligonucleotides d(A-A-T-T-G-T-G-A-G-C-G-G), d(A-A-T-T-G-T-T-A-T-C-C-G-C-T-C), d(A-A-T-T-G-T-G-A-G-C-G-G-A-T-A-A-C-A-A-T-T), d(A-A-T-T-G-T-T-A-T-C-C-G-C-T-C-A-C-A-A-T-T), and d(A-C-A-A-G-C-T-T-G-T) has been previously reported (3, 7, 11). The two nonanucleotides, d(G-A-G-C-G-G-A-T-A) and d(T-A-T-C-C-G-C-T-C), and the two tridecanucleotides, d(T-T-G-T-G-A-G-C-G-G-A-T-A) and d(T-T-G-T-T-A-T-C-C-G-C-T-C), were synthesized by published procedures (7, 11, 17–19) except for the new deblocking conditions used for the dimethoxytrityl and *p*-chlorophenyl protecting groups. The dimethoxytrityl group was removed by treatment with 2% (wt/vol) benzene sulfonic acid in chloroform; the *p*-chlorophenyl group was removed by treatment with concentrated ammonium hydroxide for 4–6 hr. The reaction conditions and the yields of the four new synthetic oligonucleotides (two 13-mers which include two 9-mers as intermediates) are described in Table 1. The oligonucleotides were characterized by two-dimensional electrophoresis-homochromatography (20, 21) of their partial digestion products (Fig. 2).

**Synthesis of Operator Sequences by Repair Synthesis Using AMV Reverse Transcriptase.** A mixture of the template strand (100 pmol) and the primer strand (300 pmol) in 0.1 M Tris·HCl (pH 8.3) was heated to 90° for 1 min and quickly cooled to 0°. The mixture was incubated at 70° for 30 min and slowly cooled to room temperature and then to 4°. The mixture was adjusted to the following salt concentrations: 50 mM Tris·HCl, 50 mM KCl, 10 mM dithiothreitol, 10 mM $MgCl_2$, and 5–10 μM each of the desired dNTPs (one of them $^{32}P$-labeled). After the addition of 10 units of reverse transcriptase, the mixture was incubated at 23° for 4–6 hr for the extension of the primer strands (22–24). The reaction was terminated by the addition of EDTA, and the mixture was loaded onto a Sephadex G-50 (fine) column (0.6 × 40 cm) at 4° to remove the dNTPs. The column was eluted with a buffer containing 100 mM NaCl/10 mM Tris·HCl (pH 7.1). The fractions containing the extended duplex were concentrated in a vacuum desiccator to approximately 200 μl and three volumes of cold ethanol was added. After storage overnight at −20°, the samples were centrifuged at 12,000 × *g* for 1 hr. The DNA pellet was dissolved in 10 mM Tris·HCl (pH 7.1) and used for binding experiments. The extended DNA products were characterized by their mobility on homochromatography in Homo-mixture III (Fig. 3) and nearest neighbor analysis (16).

**Synthesis of 29-Nucleotide Duplex DNA Containing the 17-Nucleotide Duplex DNA Sequence of the Lactose Operator by Blunt-end Ligation (25) of a 9-Mer Duplex to *Hin*dIII Linker Molecule.** The procedure used for ligation was similar to one described earlier (3). The synthetic 9-mer duplex (12

* This is paper 7 in a series, "Studies on the lactose operon." Papers 4, 5, and 6 are refs. 1, 2, and 3, respectively.

§ To whom correspondence should be sent.

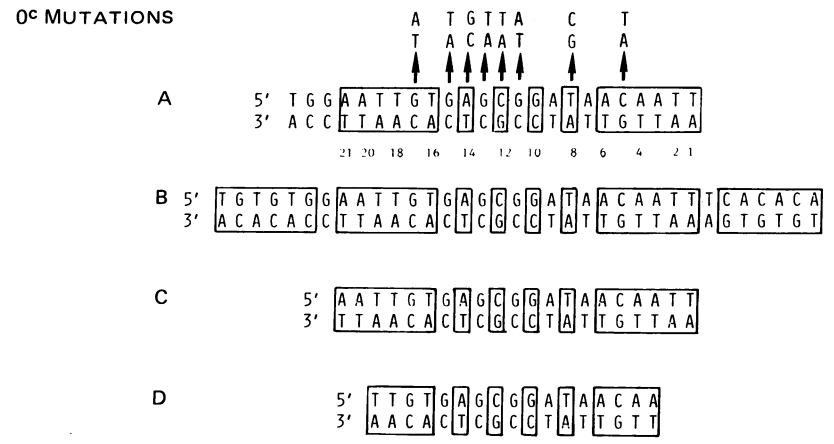


FIG. 1. DNA sequences in the lactose operator region. *A*. The 24 base-pair sequence reported by Gilbert and Maxam (8) as the piece of DNA protected by the *lac* repressor from pancreatic DNAse digestion. *B*. The 35 base-pair symmetrical sequence containing the operator sequence from the total sequence of *lac* operon control region, determined by Dickson *et al.* (6). *C*. The 21 base-pair operator sequence synthesized chemically and shown to bind to *lac* repressor by Bahl *et al.* (7). *D*. The 17 base-pair *lac* operator sequence (minimal recognition sequence) that binds to *lac* repressor reported in this paper.

pmol) and the synthetic 10-mer *Hin*dIII linker duplex (75 pmol) were joined end-to-end (Fig. 4) by incubation with 4 units of T4 DNA ligase in 100 $\mu$l of a solution containing 20 mM Tris·HCl (pH 7.5), 10 mM dithiothreitol, 10 mM $MgCl_2$, and 35 $\mu$M ATP at 20° for 6 hr. The reaction was terminated by addition of EDTA. The extent of the ligation reaction was determined by subjecting an aliquot of the reaction mixture to electrophoresis on 12% acrylamide gel (data not shown). The major part of the reaction mixture was fractionated on a Sephadex G-75 (0.6 × 75 cm) column and eluted with 100 mM NaCl containing 10 mM Tris·HCl (pH 7.5). Only the first two peaks (Fig. 5) contained ligation products and their size was determined either by homochromatography (Homo-mixture I) or by acrylamide gel electrophoresis.

**Binding of *lac* Repressor to the Synthetic *lac* Operator Fragments.** The repressor binding properties of various synthetic duplex DNA molecules were studied by the Millipore filter assay of Riggs *et al.* (26). The exact condition used has been described (7).

## RESULTS AND DISCUSSION

Gilbert *et al.* (27) sequenced a number of operator constitutive mutants and found that all these *lac* operator mutations lie within a stretch of 13 nucleotides around the axis of 2-fold rotational symmetry (see Fig. 1*A*, positions 5–17). This information together with our results showing the synthetic 21-mer duplex to be functional *lac* operator suggest that the minimal recognition sequence of the *lac* operator must be between 13

Table 1. Reaction conditions of condensation steps and yields of products

| 5'-Protected component (mmol) | 5'-Hydroxyl component (mmol) | Benzenesulfonyl tetrazole (mmol) | Product yield (%) |
|---|---|---|---|
| | I. 5' HO-T-T-G-T-G-A-G-C-G-G-A-T-A-OH 3' | | |
| [(MeO)$_2$Tr]dibuG ∓ ibuG-ClPh(0.5) | dbzA ∓ T ∓ bzA(OBz)(0.45) | 1.25 | [(MeO)$_2$Tr]dibuG ∓ ibuG ∓ bzA ∓ T ∓ bzA(OBz)(73) |
| [(MeO)$_2$Tr]dibuG ∓ bzA ∓ ibuG ∓ bzC-ClPh(0.125) | dibuG ∓ ibuG ∓ bzA ∓ T ∓ bzA (OBz)(0.1) | 0.37 | [(MeO)$_2$Tr]dibuG ∓ bzA ∓ ibuG ∓ bzC ∓ ibuG ∓ ibuG ∓ bzA ∓ T ∓ A(OBz)(71) |
| [(MeO)$_2$Tr]dT ∓ T ∓ ibuG ∓ T-ClPh (0.065) | dibuG ∓ bzA ∓ bzG ∓ bzC ∓ ibuG ∓ ibuG ∓ bzA ∓ T ∓ bzA ∓ (OBz) (0.05) | 0.20 | [(MeO)$_2$Tr]dT ∓ T ∓ ibuG ∓ T ∓ ibuG ∓ bzA ∓ ibuG ∓ bzC ∓ ibuG ∓ ibuG ∓ bzA ∓ T ∓ bzA(OBz)(50) |
| | II. 5' HO-T-T-G-T-T-A-T-C-C-G-C-T-C-OH 3' | | |
| [(MeO)$_2$Tr]dbzC ∓ bzC ∓ ibuG ∓ bzC-ClPh(0.33) | dT ∓ bzC(OBz)(0.33) | 0.83 | [(MeO)$_2$Tr]dbzC ∓ bzC ∓ ibuG ∓ bzC ∓ T ∓ bzC(OBz)(73) |
| [(MeO)$_2$Tr]dT ∓ bzA ∓ T-ClPh(0.15) | dbzC ∓ bzC ∓ ibuG ∓ bzC ∓ T ∓ bz C(OBz)(0.125) | 0.45 | [(MeO)$_2$Tr]dT ∓ bzA ∓ T ∓ bzC ± bzC ∓ ibuG ∓ bzC ∓ T ∓ bzC(OBz) (69) |
| [(MeO)$_2$Tr]dT ∓ T ∓ ibuG ∓ T-ClPh(0.04) | dT ∓ bzA ∓ T ∓ bzC ∓ bzC ∓ ibuG ∓ bzC ∓ T ∓ bzC(OBz)(0.03) | 0.12 | [(MeO)$_2$Tr]dT ∓ T ∓ ibuG ∓ T ∓ T ∓ bzA ∓ T ∓ bzC ∓ bzC ∓ ibuG ∓ bzC ∓ T ∓ bzC(OBz)(53) |

Abbreviations are as suggested by the IUPAC-IUB [*Biochemistry* 9, 4022 (1970)]. A phosphodiester linkage is represented by a hyphen, and a phosphotriester linkage is represented by (∓). Each internal internucleotidic phosphate is protected with *p*-chlorophenyl group (ClPh). The product yield is based on the limiting component.

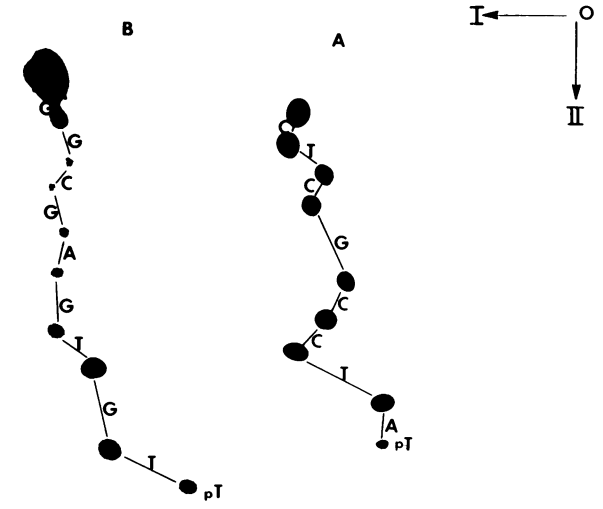


FIG. 2. Two-dimensional map of partial venom phosphodiesterase digests. *A*. Nonanucleotide d(T-A-T-C-C-G-C-T-C). *B*. Tridecanucleotide d(T-T-G-T-G-A-G-C-G-G-A-T-A). Dimension I, electrophoresis on cellulose acetate strip at pH 3.5. Dimension II, homochromatography on DEAE-cellulose thin-layer plates run in Homo-mixture VI (20). The second nonanucleotide d(G-A-G-C-G-G-A-T-A) and tridecanucleotide d(T-T-G-T-T-A-T-C-C-G-C-T-C) also gave expected fingerprints (data not shown).

and 21 nucleotides in length. Patel (28) suggested, on the basis of theoretical considerations, that the sequence A-A-T-T at both ends of the 21-mer duplex (nucleotides 1–4 and 18–21) may be involved in operator-repressor recognition. In the present study, we have defined the minimal recognition length of the operator by synthesizing various lengths of the operator sequences within the chain length of 13 to 21 by a combination of chemical and enzymatic methods (7).

As a first step, a number of duplex DNA molecules were synthesized by repair synthesis with chemically synthesized

Table 2. Repressor binding efficiency of different sizes of operator fragments

| Length of duplex operator | Binding, % |
|---|---|
| 16 | 1.5 ± 0.5 |
| 17 | 12 ± 2 |
| 19 | 40 ± 2 |
| 21 | 40 ± 2 |

The upper 21-mer single strand was annealed to the 15-mer single strand as shown in Fig. 6, structure A-15. Partial repair synthesis (29) was carried out to produce structures with duplex lengths of 16, 17, 19, and 21. The filter binding efficiency was determined by incubating 1 pmol of the operator sequence (approximately 50,000 cpm) with a 10-fold excess of *lac* repressor. Three equal volumes were filtered on three 13-mm filters. The percentage of input counts retained on the filter (% binding) has been corrected for background binding (0.5% of input counts) in the presence of $10^{-3}$ M isopropyl thiogalactoside. The 40% plateau value is almost as high as that with the 50,000 base-pair long λφ80d *lac* DNA (26) which plateaued between 40 and 70% depending on the conditions used.

21-mer templates and a 15-mer or 12-mer primer. AMV reverse transcriptase was used for the repair synthesis rather than *E. coli* DNA polymerase I because the former gave homogeneous extended products whereas the latter produced some partial products. The scheme for the preparation of various products is given in Fig. 6. When the upper 21-mer template and the lower 15-mer primer were extended to duplex lengths of 16, 17, 19, or 21, the 16-mer duplex gave a very low level of *lac* repressor binding (1.5%) (Table 2). The addition of one extra nucleotide to give a 17-mer duplex showed a 8- to 10-fold increase in repressor binding, indicating that the G/C pair at position 17 (see Fig. 1*A*) had an important contribution in the specific recognition by the *lac* repressor. Addition of two more

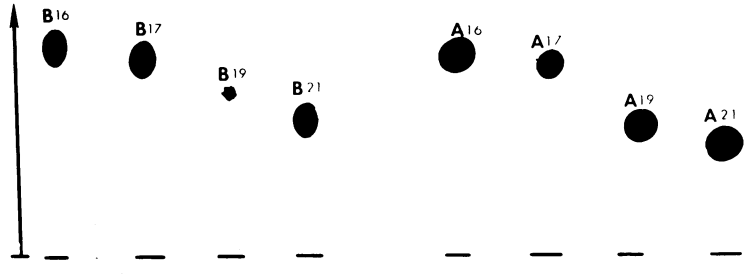


FIG. 3. One-dimensional homochromatography (Homo-mixture III) of different sizes of *lac* operator sequences as described in Fig. 6. The numbers correspond to the numbers in Fig. 6. The lines at the bottom of the figure are the origin and the arrow points to the direction of the homochromatography. Only the primer strands are visible because the 21-nucleotide template strands were unlabeled.

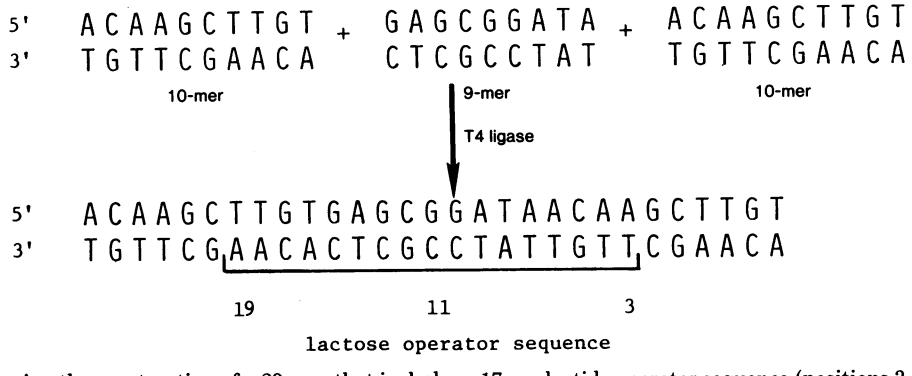


FIG. 4. Scheme showing the construction of a 29-mer that includes a 17-nucleotide operator sequence (positions 3–19). All nucleotides in this figure, as well as in Figs. 6 and 7, are deoxynucleotides.

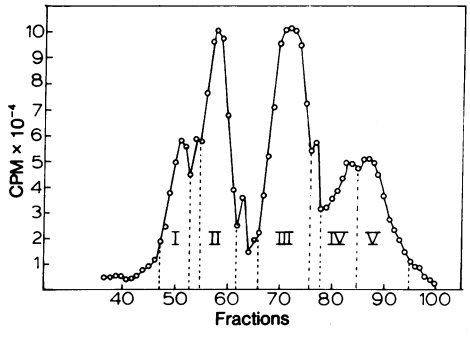


FIG. 5. Sephadex G-75 column fractionation of the blunt-end ligation products (the main reaction is as shown in Fig. 4). Peak I contained duplex fragments 29 or more nucleotides long and all the repressor binding activity. Peak II contained partial reaction products (e.g., the 9-mer dimer) and ligated product of 9-mer and only one 10-mer molecule. Peak III contained 9-mer duplex. Peaks IV and V contained single-stranded 9-mer. The 10-mer was unlabeled and thus undetected.

A residues, to extend the length of the duplex region to position 19, produced another 4-fold increase in repressor binding. This suggests that the two A/T base pairs at positions 18 and 19 provide additional contact points with the repressor molecule. Further additions of two T residues to make a 21-mer duplex

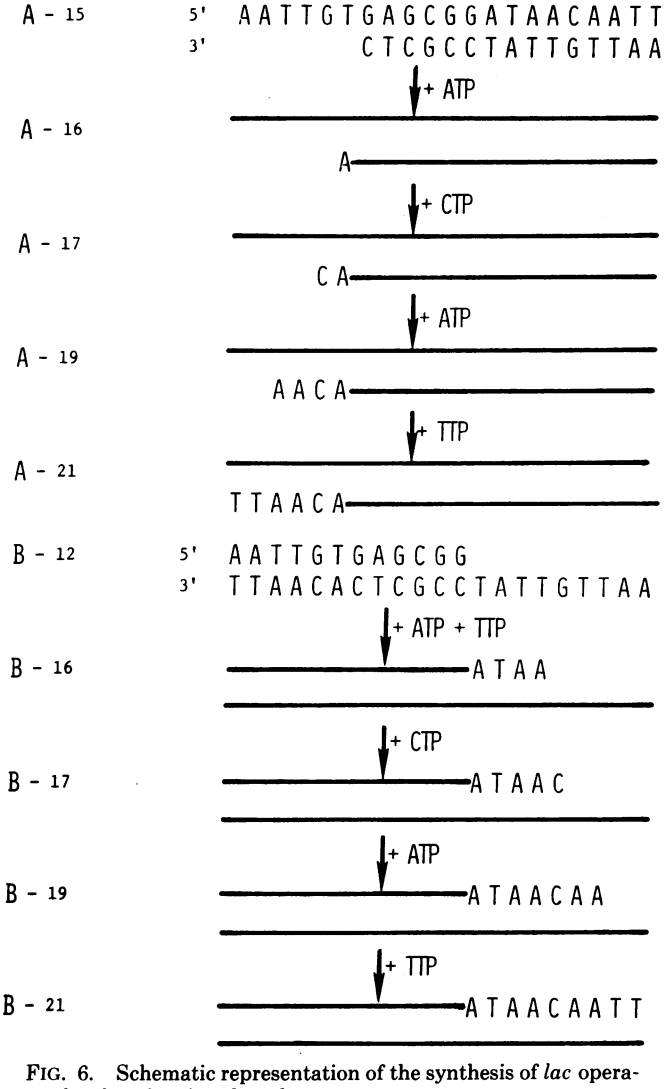


FIG. 6. Schematic representation of the synthesis of *lac* operator molecules of various lengths.

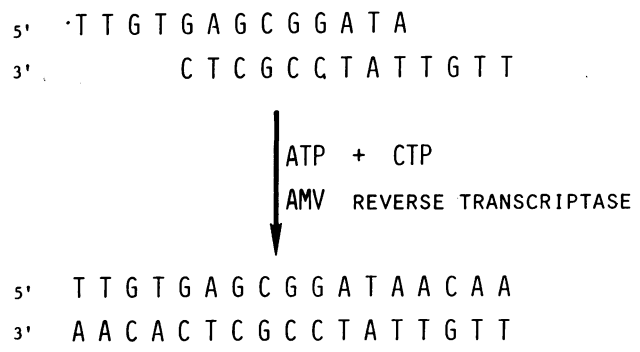


FIG. 7. Synthesis of the 17-mer duplex *lac* operator by repair synthesis.

did not increase the binding, indicating that the A/T pairs numbered 20 and 21 (Fig. 1) were not essential for repressor recognition. Similar results were obtained (not shown) when lower strand 21-mer was used as the template and varying length of duplex DNA were prepared by using a 12-mer as primer (Fig. 6*B*).

From these results, we conclude that the A/T pairs at positions 1, 2 and 20, 21 in the 21-mer duplex DNA are not needed for repressor recognition. All the essential features of the *lac* operator for maximal binding to the repressor are included in nucleotides 3–19 (Fig. 1*D*) which constitute a stretch of 17 nucleotides. Our conclusion is supported by the report of Gilbert *et al.* (30). They found that, when repressor was bound to the *lac* operator, certain A and G residues in the operator region were protected from methylation by dimethyl sulfate, and they concluded that these residues might be contact points between the operator and repressor. It is interesting to note that all their contact points were found to lie within the same 17-nucleotide duplex sequence.

To confirm that the essential features of the *lac* operator sequence lie within this stretch of 17 nucleotides, we prepared the 17-nucleotide sequence, with and without nonspecific nucleotides outside of this region, by two different methods. As shown in Fig. 4, with blunt-end ligation of a 9-mer partial *lac* operator with a 10-mer that contained four nucleotides of the *lac* operator on either end, the products of the reaction included a 29-mer duplex with a 17-mer *lac* operator sequence at the center. A 19-mer duplex with a 13-mer *lac* operator sequence (nucleotides 3–15 or 7–19) was also produced when the 10-mer was added to only one end of the 9-mer. Binding experiments showed that only the sequence that contained the 17-mer *lac* operator sequence (nucleotides 3–19) bound strongly (40% of input counts retained on filter) and specifically (sensitive to isopropyl thiogalactoside) to the *lac* repressor. The sequence that had only a 13-mer *lac* operator sequence did not bind at all (less than 0.4% counts retained) to the *lac* repressor.

A 17-mer *lac* operator sequence was prepared by repair synthesis using two chemically synthesized tridecanucleotides as template and primer (Fig. 7). This 17-mer (nucleotides 3–19) with no additional nonspecific sequences on either end was found to bind specifically to the *lac* repressor and to the same extent as the synthetic 21-mer duplex.

The AMV reverse transcriptase was kindly supplied by the Office of Program Resources and Logistics, Viral Cancer Program, Viral Oncology, Division of Cancer Cause and Prevention, National Cancer Institute, Bethesda, Md. 20014. This work was supported by Research Grants GM 18887 from the National Institutes of Health, U.S. Public Health Service, and No. 15700 from the National Research Council of Canada.

1. Marians, K. J. & Wu, R. (1976) *Nature* **260,** 360–363.
2. Marians, K. J., Wu, R., Stawinski, J., Hozumi, T. & Narang, S. A. (1976) *Nature* **263,** 744–748.
3. Bahl, C. P., Marians, K. J., Wu, R., Stawinski, J. & Narang, S. A. (1976) *Gene* **1,** 81–92.
4. Jacob, F. & Monod, J. (1961) *J. Mol. Biol.* **3,** 318–356.
5. Gilbert W. & Müller-Hill, B. (1966) *Proc. Natl. Acad. Sci. USA* **56,** 1891–1898.
6. Dickson, R., Abelson, J., Barnes, W. M. & Reznikoff, W. S. (1975) *Science* **187,** 27–35.
7. Bahl, C. P., Wu, R., Itakura, K., Katagiri, N. & Narang, S. A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 91–94.
8. Gilbert W. & Maxam, A. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 3581–3584.
9. Dykes, G., Bambara, R. A., Marians, K. J. & Wu, R. (1975) *Nucleic Acid Res.* **2,** 327–345.
10. Heyneker, H. L., Shine, J., Goodman, H. M., Boyer, H. W., Rosenberg, J., Dickerson, R. E., Narang, S. A., Itakura, K., Lin, S. & Riggs, A. D. (1976) *Nature* **263,** 748–752.
11. Itakura, K., Katagiri, N., Narang, S. A., Bahl, C. P., Marians, K. J. & Wu, R. (1975) *J. Biol. Chem.* **250,** 4592–4600.
12. Sulkowski, E. & Laskowski, M., Sr. (1971) *Biochim. Biophys. Acta* **240,** 443–447.
13. Platt, T., Files, J. G. & Weber, L. (1973) *J. Biol. Chem.* **248,** 110–121.
14. Weiss, B., Jacquemin-Sablon, A., Live, T. R., Fareed, G. & Richardson, C. C. (1968) *J. Biol. Chem.* **243,** 4543–4555.
15. Panet A., van de Sande, J. H., Loewen, P. C., Khorana, H. G., Raae, A. J., Lillehauge, J. R. & Kleppe, K. (1973) *Biochemistry* **12,** 5045–5050.
16. Wu, R., Jay, E. & Roychoudhury, R. (1976) *Methods Cancer Res.* **12,** 88–176.
17. Itakura, K., Bahl, C. P., Katagiri, N., Michniewicz, J. J., Wightman, R. H. & Narang, S. A. (1973) *Can. J. Chem.* **51,** 3649–3651.
18. Itakura, K., Katagiri, N., Bahl, C. P., Wightman, R. H. & Narang, S. A. (1975) *J. Am. Chem. Soc.* **97,** 7327–7332.
19. Katagiri, N., Itakura, K. & Narang, S. A. (1975) *J. Am. Chem. Soc.* **97,** 7332–7337.
20. Jay, E., Bambara, R., Padmanabhan, R. & Wu, R., (1974) *Nucleic Acid Res.* **1,** 331–353.
21. Tu, C. D., Jay, E., Bahl, C. P. & Wu, R. (1976) *Anal. Biochem.* **74,** 73–93.
22. Temin, H. & Baltimore, D. (1972) *Adv. Virus Res.* **17,** 129–186.
23. Leis, J. & Hurwitz, J. (1974) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, pp. 143–150.
24. Kacian, D. L. & Spiegelman, S. (1974) in *Methods in Enzymology,* eds. Grossman, L. & Moldave, K. (Academic Press, New York), Vol. 29, pp. 150–173.
25. Sgarmella, V., van de Sande, J. H. & Khorana, H. G. (1970) *Proc. Natl. Acad. Sci. USA* **67,** 1468–1475.
26. Riggs, A. D., Suzuki, H. & Bourgeois, S. (1970) *J. Mol. Biol.* **48,** 67–83.
27. Gilbert, W., Gralla, J., Majors, J. & Maxam, A. (1975) *Protein Ligand Interactions* (Walter de Gruyter, Berlin and New York), pp. 193–206.
28. Patel, D. (1975) *Biochemistry* **14,** 1057–1059.
29. Wu, R. (1970) *J. Mol. Biol.* **51,** 501–521.
30. Gilbert, W., Maxam, A. & Mirzabekov, A. (1976) "Control of ribosome synthesis, in *Alfred Benzon Symposium IX,* eds. Maaløe, O. & Kjelgaard, N. (Munksgaard, Copenhagen), pp. 139–148.