

An operator at -280 base pairs that is required for repression of *araBAD* operon promoter: Addition of DNA helical turns between the operator and promoter cyclically hinders repression

(positive regulation/DNA binding/transcription/repression/DNA loops)

TERESA M. DUNN*, STEVEN HAHN*, SHARON OGDEN†, AND ROBERT F. SCHLEIF*

*Department of Biochemistry, Brandeis University, Waltham, MA 02154; and †Department of Microbiology and Immunology, University of Washington, Seattle, WA 98195

Communicated by Charles Yanofsky, May 10, 1984

ABSTRACT A site has been found that is required for repression of the *Escherichia coli* *araBAD* operon. This site was detected by the *in vivo* properties of deletion mutants. *In vitro* protection studies with DNase I and dimethylsulfate showed that araC protein can specifically bind in this area to nucleotides lying at position -265 to -294 with respect to the *araBAD* operon promoter (P_{BAD}) transcription start point. The previously known sites of protein binding in the *ara* operon lie between +20 and -160. Since the properties of deletion strains show that all the sites required for *araBAD* induction lie between +20 and -110, the new site at -280 exerts its repressive action over an unusually large distance along the DNA. Insertions of -16, -8, 0, 5, 11, 15, 24, and 31 base pairs of DNA between the new site and P_{BAD} were constructed. Repression was impaired in those cases in which half-integral turns of the DNA helix were introduced, but repression was nearly normal for the insertions of 0, +11, and +31 base pairs.

The L-arabinose operon in *Escherichia coli* is well documented to be positively regulated by the araC protein. Additionally, the operon is negatively regulated by the same protein (1-4). Paradoxically, the negative regulation appears to involve a site lying upstream of all the sites required for induction. Initially, the site involved in this repression phenomenon appeared to be the araC-binding site, which lies from position -110 to -140 (2, 5-7) (Fig. 1). From this position, the protein could be imagined to make direct contact with the complex of cyclic AMP receptor protein araC protein-RNA polymerase, all of which are involved in induction. Since upstream repression, even from this nearby site, appeared unusual, we examined the question more carefully by using a set of deletions.

As reported here, the deletions and *in vitro* binding experiments revealed the existence of yet another site for araC protein binding in the *ara* regulatory region. This site at position -280 lies too far for any simple direct interaction to exist between it and the complex of proteins on DNA that are required for initiation of transcription, and yet this site is required for repression of transcription. This finding of a second regulatory site located a considerable distance from the promoter is similar to the recent finding in the *gal* operon of a second operator site located downstream from the promoter and lying within the *galE* gene (8).

MATERIALS AND METHODS

Media and Strains and General Methods. Media, strains, and general methods were as described (9-12).

Construction of pTD3 and pTD4. A 440-base-pair fragment containing the *araCBAD* regulatory region (13) was made

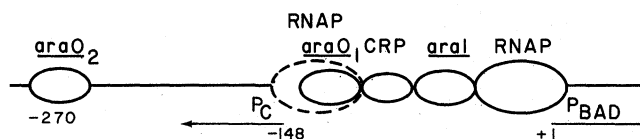


FIG. 1. Location of the protein-binding sites of the L-arabinose *araCBAD* regulatory region, drawn to scale. Transcription of P_{BAD} is rightward from +1, and transcription of P_C is leftward from -148. RNAP, RNA polymerase.

blunt-ended by treatment with S1 nuclease. *Hind*III and *Eco*RI linkers were ligated in 10 M excess to the blunt-ended fragment, followed by codigestion with *Hind*III and *Eco*RI. The *ara* fragments were then ligated to *Hind*III/*Eco*RI-cut pBR322 that had been purified from the 30-base-pair fragment (14) by A-15 agarose chromatography. Plasmids having the inserted *ara* fragment in the desired orientation were selected on the basis that expression of the *araBAD* operon promoter (P_{BAD}), and thus tetracycline resistance (15, 16), becomes arabinose dependent, while expression of the *araC* gene promoter (P_C) is not. The fragments were subsequently recloned between the *Hind*III and *Eco*RI sites of the plasmid pK01 so that the promoters could be assayed by expression of the galactokinase gene to which they were fused (17). pTD3 and pTD4 have P_{BAD} and P_C fused to the *galK* gene, respectively.

Construction of the Deletions. Either pTD3 or pTD4 was linearized with *Eco*RI and treated with exonuclease *Bal* 31 to remove 100-600 base pairs. The shortened linear plasmids were ligated to a 10 M excess of *Eco*RI linkers, codigested with *Eco*RI and *Hind*III, purified by electrophoresis, and recloned into *Eco*RI/*Hind*III-cut pK01 that had been purified by electrophoresis through agarose and treated with bacterial alkaline phosphatase (18). After recloning, the plasmids were transformed into TMD29, an *araC* *galK* derivative of C600 (17). Candidates were purified by rapid plasmid isolation (19) and cut with *Eco*RI and *Hind*III. The fragments were labeled with Klenow fragment and [α - 32 P]dATP and sized on a sequencing gel by comparison to a "G's" [dimethylsulfate (guanines)] sequencing reaction size standard. In many cases, the deletions were later sequenced through the deletion end point and linker. With respect to +1, the start of P_{BAD} transcription, the end points of deletions extending from *araC* toward P_{BAD} were as follows: -385, -340, -330, -294, -263, -247, -222, -190, -174, -170, -163, -150, -143, -138, -137, -133, -132, -126, -125, -113, -111, -107, -96, -95, -84, -57, -53, -49, -48, -39, -36, -35, -30, -10; and from *araB* toward P_C they were as follows: -50, -60, -65, -84, -97, -103, -105, -110, -114, -115, -125, -126, -130, -159, -168, -190, -200, -230, -245.

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.

Abbreviations: P_{BAD} , promoter for *araBAD* operon; P_C , promoter for *araC* gene; CRP, cyclic AMP receptor protein.

GalK Assays. Cells were grown in M10 medium/0.2% glycerol or M10 medium/0.2% glycerol/0.2% arabinose. All cells were grown for at least six generations to a density of $1-4 \times 10^8$ cells per ml in the presence of 20 μg of ampicillin per ml. Chloramphenicol was added to 200 $\mu\text{g}/\text{ml}$ before harvesting the cells. The galactokinase assays were as described by McKenney *et al.* (17) except that 25 mM EDTA was added to stop the reactions. Galactokinase units are expressed as nmol of galactose phosphorylated per min per ml of cells at OD_{650} of 1.0.

DNase I and Methylation Protection. The DNase I protection experiments were done by the procedure of Galas and Schmitz (20) with the following modifications. DNase I reaction buffer was 20 mM Tris·HCl, pH 7.6/2.5 mM $\text{MgCl}_2/0.1$ mM EDTA/50 mM KCl/0.5 mM dithioerythritol/25 μg of bovine serum albumin per ml/100 mM arabinose. DNase I was dissolved just prior to use in the reaction buffer, which also contained 0.5 mM CaCl_2 . araC protein was incubated with the DNA (100 μl) for 10–15 min at 37°C prior to the addition of 10 μl of freshly diluted DNase I (0.8 $\mu\text{g}/\text{ml}$) for 45 sec at 20°C. The DNase I reaction was stopped, DNA was precipitated, and the samples were resuspended in 10 mM Tris·HCl, pH 8.0/60% formamide/dyes and evaporated at 90°C prior to loading, until the formamide concentration reached 90% (30 min). The methylation protection studies were done according to the method of Johnsrud (21), with the following modifications. The buffer was as described above for the DNase I experiments, and the protein was incubated with the DNA in a 100- μl vol for 10 min at 37°C prior to the addition of 1 μl of dimethylsulfate for 45 sec at 20°C. Samples were stopped and treated according to the normal "G's" reaction protocol (12).

Spacing Mutations. Plasmid pTD3 was cut with *Bst*EII at -208 with respect to P_{BAD} , and digested with S1 nuclease before ligation to reclose the circles to generate the deletions of 8 and 16 base pairs. The single-stranded ends of the *Bst*EII-cut DNA were filled out with DNA pol I before ligation to generate the +5. The filled out DNA was ligated to the *Xba* linker T-C-T-A-G-A to generate the +11. The +11 was cut with *Xba* and filled out with DNA pol I to generate the +15. The +15 was cut, partially filled out with DNA pol I, and the restriction-site mobilizing element from pUC-4K (22) was added using partially filled out *Bam*HI ends, cloned, and the kanamycin-resistance element was eliminated by digesting with *Sal* I and ligating, leaving a total insertion of +31 base pairs compared to wild type. The sequences of the insertions were determined by DNA sequencing. All the insertion-containing plasmids were normally inducible by arabinose.

RESULTS

Construction of the Deletions. Two plasmids, pTD3 and pTD4, were constructed by cloning the 440-base-pair *ara* regulatory region DNA segment (13) with *Eco*RI and *Hind*III linkers in either orientation between the *Eco*RI and *Hind*III sites of the plasmid pK01 (17). Plasmid pTD3 has P_{BAD} fused to the *galK* gene of the pK01 expression vector, and pTD4 has P_{C} fused to *galK* (Fig. 2). Deletions were isolated by linearizing the plasmids with *Eco*RI, digesting with the exonuclease *Bal* 31, replacing the *Eco*RI site with a new linker, and recloning the *ara* fragments between the *Eco*RI and *Hind*III sites of purified pK01. The structure of the plasmid pK01 permits direct comparison of many different promoter activities, and reintroducing the deleted *ara* region into the unmodified vector ensures that the alterations in enzyme level that we measure originate from the *ara* DNA and not the plasmid.

Induction and Repression Requirements of P_{BAD} . The deletions from the P_{C} side of the *ara* regulatory region, which ended before position -107, were about 100-fold inducible by arabinose, the same as the undeleted *ara* regulatory re-

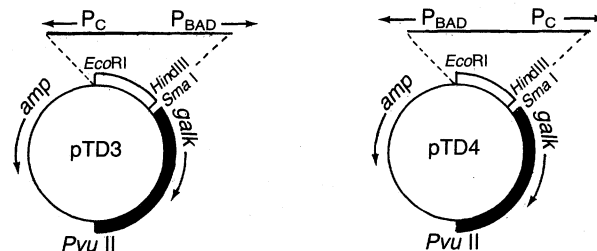


FIG. 2. Structure of the P_{BAD} and P_{C} -*galK* fusion plasmids. The plasmids are identical except for the orientation of the *ara* fragment between the *Eco*RI and *Hind*III sites. The remainder of the plasmid is pK01 (17). The *Pvu* II (destroyed in cloning)-*Eco*RI fragment that encodes β -lactamase (*amp*) is from pBR322 (14). The entire galactokinase gene resides on the *Sma* I-*Pvu* II fragment marked *galK*. There are stop codons in all three reading frames between *Sma* I and the translation start site of *galK*.

gion. This is expected, as such deletions leave the cyclic AMP receptor protein (CRP), I, and RNA polymerase sites of P_{BAD} intact. Similarly, deletions from the P_{BAD} side extending toward P_{C} , which ended before -80, left P_{C} with normal activity. This too is expected as these deletions leave P_{C} intact.

Surprisingly, the basal levels of P_{BAD} expression in the deletion strains reveal that, although a strain containing a deletion extending to nucleotide -294 represses P_{BAD} normally, any deletion that removes additional *ara* DNA expresses a 10-fold increased basal level (Fig. 3), unless the deletion extends into the CRP site. As we show in the next section, araC protein is capable of binding *in vitro* to a 30-base-pair sequence centered at position -280, which we term *araO*₂. While the above results directly demonstrate the involvement of *araO*₂ in the repression of P_{BAD} , these experiments do not speak to the question of whether *araO*₁ is also involved in the repression.

The dependence of repression of P_{BAD} on *araO*₂ might be due to a simple titration of araC protein made more extreme by the plasmid copy number (10–30 copies per cell). This possibility can be excluded, because deletions with and without *araO*₂ were transferred to the chromosome by the methods outlined by McKenney *et al.* (17), and in single copy they showed that loss of repression is dependent on deletion of *araO*₂ just as observed for the multicopy plasmids. With *araO*₂ present, the basal level was ≈ 0.03 units of galactokinase; without *araO*₂, it was ≈ 0.28 units; and when induced by arabinose, the level was 2.7 units.

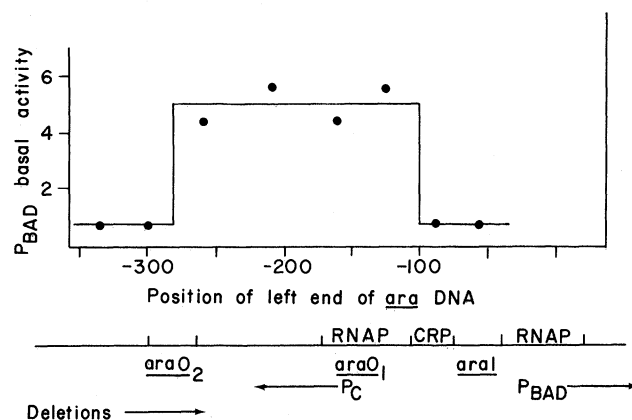


FIG. 3. Basal level *araBAD* promoter activities in galactokinase units from a set of deletion plasmids containing the *ara* P_{BAD} -*galK* fusion in strain SH322 ($F^- \text{leu} \Delta \text{lac74 galK Str}^r \text{thi}$) under noninducing conditions (lacking arabinose) (●). The deletions ended at -330, -294, -263, -222, -163, -126, -94, and -57. RNAP, RNA polymerase.

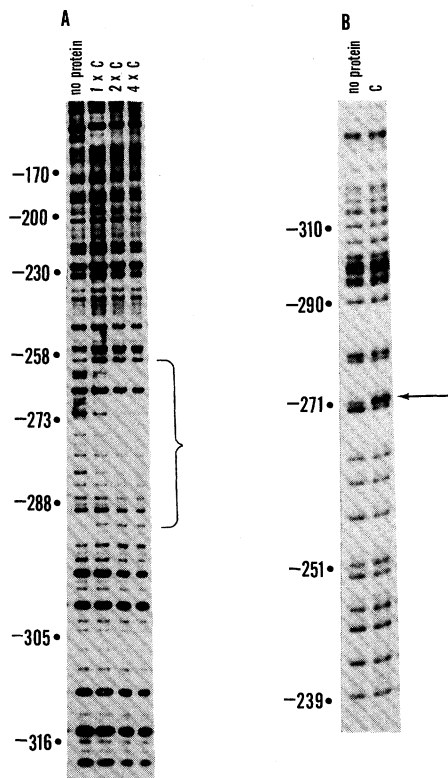


FIG. 4. (A) DNase I protection by araC protein in the presence of arabinose. A deletion plasmid pTD358 was cut with *Eco*RI, 5' end-labeled, and cut with *Hind*III. The resulting fragment (-330 to +47) was used for the DNase I studies. araC protein (C) was added at a level of 0.22 (1 time) to 0.88 (4 times) ng. DNase I was used at 0.08 μ g/ml for 45 sec. The protected region near the top of the lanes to which araC protein was added results from protection of the *araO*₁ site (-110 to -160). When this fragment is run further on the gel binding at both the *araO*₁ and *araI* sites is resolved (not shown). Brace marks the *araO*₂ binding site. (B) Dimethylsulfate protection in the presence of arabinose. Plasmid pTD3 was cut and 5' end-labeled at the unique *Bst*EII site (-200) and cut with *Eco*RI. The resulting fragment (-200 to -407) was used for the methylation studies. The enhanced methylation at the -271 residue is marked with an arrow.

DNase I and Methylation Protection of *araO*₂ by araC Protein. Since previous studies have implicated araC protein in repression of P_{BAD} (2, 3, 5), it is likely that araC protein itself binds to a previously undetected site in the region of position -280. The DNase I protection experiments revealed specific binding of araC protein in the region from -265 to -294 (Fig. 4A) and the methylation protection experiments showed a single strongly enhanced methylation at -271 and a protection at -270 (Fig. 4B). Similar methylation protection studies of the *araI* and *araO*₁ sites also found only one or two strongly enhanced G residues (23).

The *in vitro* DNA binding experiments presented here to identify the *araO*₂ site were carried out in the presence of arabinose, both to stabilize the highly labile araC protein and because quantitative binding measurements had shown that the affinity of araC for both *araI* and *araO*₁ is greatest in the presence of arabinose (24). Subsequent measurements, however (K. Martin, personal communication), have shown that the affinity of araC protein for *araO*₂ is nearly the same in the presence of the inducer L-arabinose as it is in the presence of the anti-inducer D-fucose.

DNA-Turn Dependence of Repression. To examine the effect of changing the distance and angular orientation of the *araO*₂ site with respect to P_{BAD}, we deleted and inserted 5-

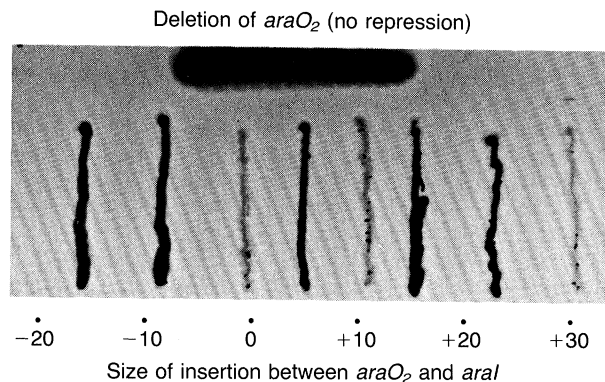


FIG. 5. Photograph of a galactose MacConkey plate streaked with the various insertion strains as indicated. A strain with *araO*₂ deleted was streaked across the top and strains with insertions of -16, -8, 0, +5, +11, +15, +24, and +31 base pairs were streaked as indicated and incubated at 33°C. In one experiment, these same strains grown in M10/glycerol at 34°C contained 4.1, 1.7, 1.7, 0.84, 1.9, 0.72, 0.96, 1.6, and 0.59 units of galactokinase, with the variation between duplicate assays at \approx 0.3 units.

31 base pairs at a nonessential site between *araO*₂ and P_{BAD}. When the change in the number of helical turns was nearly an integral value, assuming \approx 10.5 base pairs per turn, repression was possible, but when the change generated nearly a half-integral number of turns, repression was impaired. These data are shown pictorially by the color generated on the indicating plates in which cells able to repress give white streaks and cells less able to repress possess more galactokinase and yield red streaks (Fig. 5). Direct measurement of the P_{BAD}-driven galactokinase levels in the strains is subject to substantial fluctuation because of its low level, but the values obtained in a typical experiment show the same trend as the indicating plate. Fortunately the MacConkey indicating plates magnify and clearly display the 2- to 5-fold differences in enzyme levels present in the spacing mutation strains.

The spacing mutations eliminate the possibility of one artifact. Conceivably *araO*₂ is an araC protein-dependent attenuator of transcription that originates from a site still further upstream. If so, the spacing mutations would be expected not to alter repression. Since they did, the attenuator mechanism seems unlikely.

DISCUSSION

The properties of deletions extending toward P_{BAD} led to the identification of a site involved in P_{BAD} repression. Remarkably, this site is centered at position -280, far from the sites that are required for the induction of P_{BAD}, which span bases -110 to approximately +10. The newly discovered site is termed *araO*₂, because it functions as an operator in being required for repression of P_{BAD}, and it is the second operator discovered in the P_C-P_{BAD} regulatory region. The first operator to have been discovered is now termed *araO*₁, and its role, if any, in the repression of P_{BAD} has not been established by our experiments. However, if *araO*₁ is involved in repression of P_{BAD}, its involvement is complex, because increased levels of araC protein can generate almost complete repression of P_C, presumably by fully occupying *araO*₁, while P_{BAD} remains fully inducible (25). The operator *araO*₂ was shown to be capable of binding araC protein by DNase I and dimethylsulfate protection studies *in vitro*. *araO*₂ lies in the transcribed but untranslated leader region of the *araC* gene.

The involvement of *araO*₂ in repression of P_{BAD} is consistent with published deletion data. The Englesberg deletion Δ 719, which originally revealed the phenomenon of repres-

sion from an upstream site, removes *araO*₂ in addition to entering the site we now name *araO*₁, whereas the shorter deletion Δ 766, which did not interfere with repression, does not enter the *ara* regulatory region (2, 7).

Three general mechanisms appear possible for repression generated by the binding of a protein at a substantial distance from the transcription start point. An alteration in the structure of the DNA could be propagated between the repression site and the transcription site. Alternatively, the protein could polymerize from the repression site along the DNA and interfere with transcription. The third possibility is the generation of a loop in the DNA, which brings the protein bound at the distant repression site near the DNA or proteins on the DNA at the transcription start site. For example, the DNA could bend so that araC protein bound at *araO*₂ could bind to araC protein bound at *araI*. This could leave araC protein in a state incapable of activating transcription from P_{BAD} and would therefore mean that binding at *araI* is not tantamount to stimulating transcription.

The spacing mutations that alter the distance between *araO*₂ and the induction region of the *araBAD* also alter the number of DNA helical turns between the induction region and *araO*₂. Introduction of 11 and 31 base pairs, both near integral multiples of the 10.5 base pairs per turn of DNA *in vitro* (26), permit repression, but alterations in the number of turns by half-integral values interfered with repression of *araBAD*. These data are most easily compatible with a DNA-looping mechanism. Repression could be hindered in the strains with insertions generating half-integral turns because araC protein then is on the wrong side of the DNA to form the loop required for repression (Fig. 6).

If a loop is to form when araC protein is on the wrong side of the DNA, the DNA between P_{BAD} and *araO*₂ must be twisted by half a turn. *In vitro*, \approx 4 kcal/mol (1 cal = 4.184 J) are required to twist linear DNA of 200 base pairs by half a turn (27, 28). Although the supercoiled state of DNA *in vivo* as well as the presence of other DNA-binding proteins undoubtedly will alter this value, this order of energy is likely to be involved. Typical protein-DNA dissociation constants indicate that 10–15 kcal per mol of energy are available from binding. Therefore, the consumption of several kcal/mol in twisting the DNA up to half a turn could substantially alter a delicately balanced interaction required for repression.

Upon locating the binding sites of the proteins involved in regulation of the arabinose operon by DNase I protection, we proposed a model that accounted for much of the operon's known behavior (6). Part of this model was that under inducing conditions araC protein selectively bound at the *araI* site to stimulate P_{BAD}, and under repressing conditions it did not bind to *araI*, but instead, it selectively bound to *araO*₁ to repress P_{BAD} as well as P_C. A second model, also derived from DNase I protection data, for *ara* regulation also incorporated the simple mechanism of selective binding (29). DNase I protection is not a good method for measuring binding constants of a tight-binding protein, and repeated experiments (23) failed to reproduce the tendency of selective binding shown by *araC* in the initial experiments (6). Similarly, the data of Lee *et al.* (29) show no clear selectivity in bind-

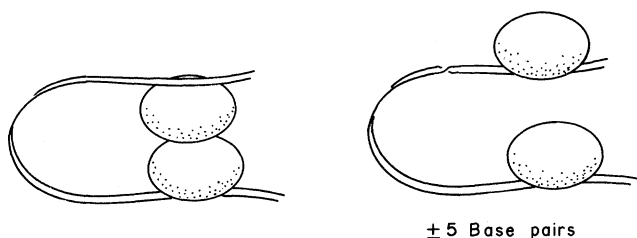


FIG. 6. Half a helical twist can hinder loop formation.

ing. More definitive experiments by the gel electrophoresis DNA binding assay clearly show that the affinities of araC protein for *araO*₁ and *araI* vary in parallel (24). That is, there is no apparent selectivity in binding by araC protein, at least not *in vitro* on linear DNA.

In light of the absence of selective binding and the findings presented in this paper, we now modify the model for regulation of the arabinose operon. In the absence of arabinose, araC protein bound at *araO*₂ could interact via a DNA loop with a component of the induction complex, CRP, araC protein, or RNA polymerase, and the loop's formation could prevent induction by holding one of the components in an inactive state. However, in inducing conditions, presence of arabinose and functional CRP, the loop could open and RNA polymerase would be permitted to initiate transcription from P_{BAD}. Until a mutational analysis is completed, we must leave open any role for *araO*₁ in repression of P_{BAD}.

We thank Keith McKenney for providing the plasmid pK01, Carol Stoner and William Hendrickson for many helpful discussions, and Michael Wormington and Pieter Wensink for suggestions on the manuscript. This work was supported by Grant GM18277 from the National Institutes of Health.

- Sheppard, D. & Englesberg, E. (1967) *J. Mol. Biol.* **25**, 443–454.
- Englesberg, E., Squires, C. & Meronk, F. (1969) *Proc. Natl. Acad. Sci. USA* **62**, 1100–1107.
- Greenblatt, J. & Schleif, R. (1971) *Nature (London) New Biol.* **233**, 166–170.
- Wilcox, G., Meuris, P., Bass, R. & Englesberg, E. (1974) *J. Biol. Chem.* **249**, 2946–2952.
- Schleif, R. & Lis, J. (1975) *J. Mol. Biol.* **95**, 417–431.
- Ogden, S., Haggerty, D., Stoner, C., Kolodrubetz, D. & Schleif, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3346–3350.
- Wilcox, G., Al-Zarban, S., Cass, L. G., Clarke, P., Heffernan, L., Horowitz, A. H. & Miyada, C. G. (1982) in *Promoters: Structure and Function*, eds. Rodriguez, R. & Chamberlain, M. (Prager, New York), pp. 183–194.
- Irani, M. H., Orosz, L. & Adhya, S. (1983) *Cell* **32**, 783–788.
- Stoner, C. & Schleif, R. (1983) *J. Mol. Biol.* **171**, 369–381.
- Schleif, R. & Wensink, P. (1981) *Practical Methods in Molecular Biology* (Springer, New York).
- Kosiba, B. & Schleif, R. (1981) *J. Mol. Biol.* **156**, 53–66.
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- Smith, B. R. & Schleif, R. (1978) *J. Biol. Chem.* **253**, 6931–6933.
- Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
- Rodriguez, R. L., West, R. W., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1979) *Nucleic Acids Res.* **6**, 3267–3287.
- Widera, G., Gautier, F., Lindenmaier, W. & Collins, J. (1978) *Mol. Gen. Genet.* **163**, 301–305.
- McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. & Rosenberg, M. (1981) in *Gene Amplification and Analysis*, eds. Chirikjian, J. & Papas, T. S. (Elsevier/North-Holland, Amsterdam), Vol. 2, pp. 383–415.
- Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. (1977) *Science* **196**, 1313–1316.
- Klein, R. D., Selsing, E. & Wells, R. D. (1980) *Plasmid* **3**, 88–91.
- Galas, D. J. & Schmitz, A. (1978) *Nucleic Acids Res.* **5**, 3157–3170.
- Johnsrud, L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5314–5318.
- Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
- Ogden, S. (1981) Dissertation (Brandeis University, Waltham, MA).
- Hendrickson, W. & Schleif, R. (1984) *J. Mol. Biol.*, in press.
- Hahn, S. & Schleif, R. (1983) *J. Bacteriol.* **155**, 593–600.
- Wang, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 200–203.
- Shore, D. & Baldwin, R. (1983) *J. Mol. Biol.* **170**, 957–981.
- Horowitz, D. & Wang, J. (1984) *J. Mol. Biol.* **173**, 75–91.
- Lee, N., Gielow, W. & Wallace, R. G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 752–756.