JOURNAL OF BACTERIOLOGY, Jan. 1988, p. 330–334
0021-9193/88/010330-05$02.00/0

Vol. 170, No. 1

DNA Sequence of the D-Serine Deaminase Activator Gene *dsdC*

S. PALCHAUDHURI,[1] V. PATEL,[1] AND E. McFALL[2]*

Department of Microbiology, Wayne State University School of Medicine, Detroit, Michigan 48201,[1] and Department of Microbiology, New York University School of Medicine, New York, New York 10016[2]

Received 22 June 1987/Accepted 21 October 1987

We have determined the DNA sequence of *dsdC*, the gene that encodes the D-serine deaminase activator protein of *Escherichia coli* K-12. The sequence contains a single open reading frame that terminates in a UGA codon. One the basis of the size of the protein, 33 kilodaltons, and the amino acid sequence encoded by the open reading frame, we identified a likely translation initiation codon 731 base pairs upstream of the translation initiation codon for the divergently transcribed D-serine deaminase gene. There is a broad range of codon usage, not surprising in view of the weak expression of the gene. The N-terminal two-thirds of the activator is arginine-lysine rich and quite polar; the remainder is more neutral. The segment of the protein that seems most likely to have potential to form the helix-turn-helix structure characteristic of DNA-regulatory proteins is located near the end of the polar region. The protein contains a region with significant homology to λ*attB*.

The D-serine deaminase activator, product of the *dsdC* gene, is one of the most efficient procaryotic regulatory proteins yet described. Although present in *Escherichia coli* K-12 at a level of only about eight molecules per induced cell, it represses its own synthesis fivefold in the absence of D-serine and activates transcription of the D-serine deaminase structural gene (*dsdA* gene) 700- to 800-fold in the absence of cyclic AMP-cyclic AMP-binding protein (cAMP-CAP) and 3,000- to 5,000-fold in the presence of cAMP-CAP (10, 13). The *dsdA* and *dsdC* genes are the only genes known to be specific to D-serine deaminase synthesis. They are closely linked and are transcribed with opposite polarity from a central control region of several hundred nucleotide pairs (3; see below).

We have identified the activator monomer on sodium dodecyl sulfate-acrylamide gels as a basic protein of about 33,000 molecular weight and in its active form on sucrose gradients as a dimer of about 66,000 molecular weight (10). Its specificity and efficiency in the activation process, together with the cAMP-CAP effect, suggested that it should be a DNA-binding protein with high affinity—enhanced by cAMP-CAP—for sequences in the *dsdA* promoter.

To determine whether the activator has properties in common with other DNA-regulatory proteins, to set a limit on the size of the intergenic region, and to try to identify an apparent secondary *att*λ site in *dsdC* (2), we sequenced *dsdC*. We found a likely *dsdC* translation initiation sequence, a possible helix-turn-helix sequence (21, 23), and a sequence with significant homology to *att*λ (24), as described below.

MATERIALS AND METHODS

Bacterial strain and its cultivation. Strain AC6083 (10), a *dsdA* and λCI857 derivative of strain C600, which carries plasmid pAC131 (*dsdA*$^+$ *dsdC*$^+$), was used as source of plasmid DNA. It was routinely cultured on LB broth, with kanamycin and tetracycline at final concentrations of 20 μg/ml to maintain selection of the drug resistance markers of the plasmid.

Sequencing of *dsdC*. The products of partial *Hae*III digests of plasmid pAC131 (10), which carries the entire *dsd* region, were cloned into the *Sma*I site of M13mp18 (25). Ligation products were transformed into strain JM103, and progeny bacteriophage were examined for the presence of *dsdC* sequences by Southern blotting (20) to appropriate probes. DNA sequences of cloned *dsd* fragments were determined by the Sanger and Coulson dideoxy method (19). Both strands were sequenced, including all overlaps across restriction sites.

RESULTS

***dsdC* sequence.** By subcloning the *dsdC* gene, we found that it and its promoter are present on a 1,614-base-pair *Sph*I-*Eco*RI fragment (14; S. M. Bornstein-Forst and E. McFall, unpublished data) located between base pairs −211 and −1825 (Fig. 1). (Numbering refers to the transcription start for *dsdA* [3].) Accordingly, partial *Hae*III digests of this fragment were cloned into phage M13mp18 and sequenced as described in Materials and Methods.

The DNA-coding strand of *dsdC*, from the *Sph*I to the *Eco*RI sites, is presented in Fig. 1. The corresponding amino acid sequence, beginning with the first potential translation initiation codon, at base pair −462, of the single open reading frame, is shown above the DNA sequence. Other points of possible interest are also indicated.

***dsdC* initiation and termination signals.** The *dsdC* open reading frame terminates with a TGA codon, 352 base pairs before the *Eco*RI site (Fig. 1), at base pair −1476. This fixes the C terminus of the activator.

There are potential ATG translation initiation codons at −461, −596, and −731 that would yield proteins of molecular weights 44,000, 38,600, and 32,900, respectively (Fig. 1). None has a consensus Shine-Dalgarno sequence (7); with an agreement of five of nine bases, the one for a base pair −731 start is the best. We previously found that the activator monomer has a molecular weight of about 33,000 under both native and denaturing conditions (10), which suggested that −731 is the correct translation start. We therefore cloned the 1,200-base-pair *Bgl*II-*Eco*RI fragment (which spans base pairs −630 to −1830) into the *Bam*HI-*Eco*RI site of plasmid pUC19 (25), downstream of the *lac* promoter, to form plasmid pEM192. When the latter plasmid was transformed into strain EM145-1 (15), which has a *dsdC*::Mu dl(Ap^r *lac*) insertion, inactivating the chromosomal *dsdC*, we were able

* Corresponding author.

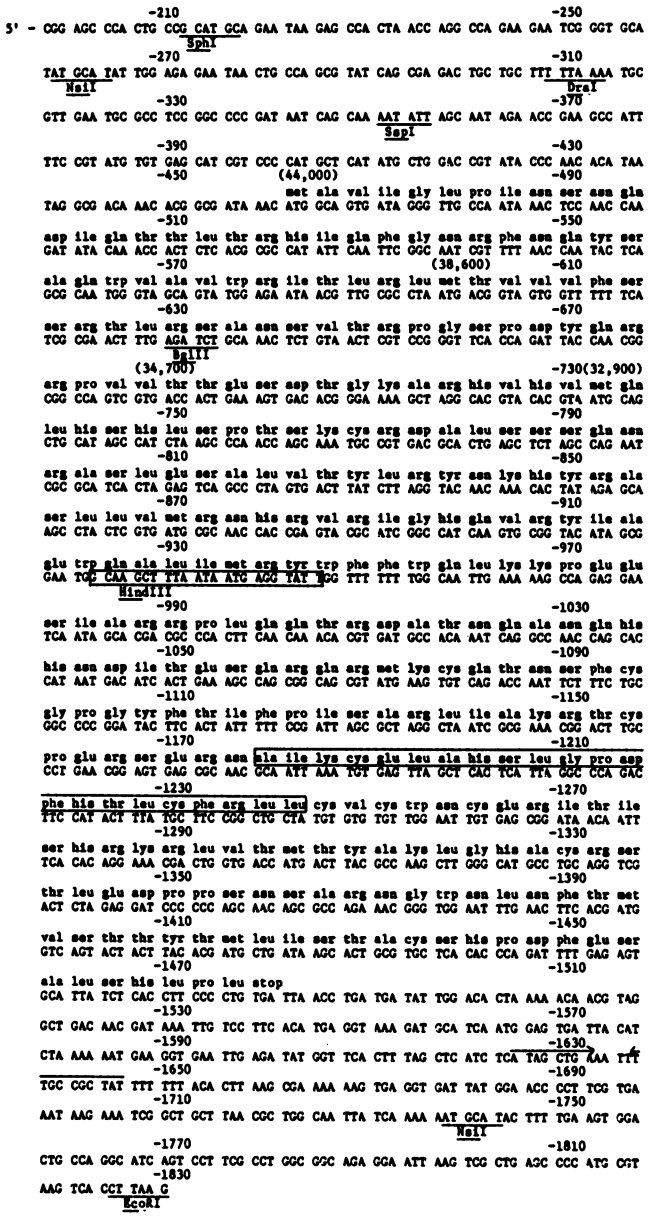

FIG. 1. DNA sequence of the *dsdC* region, with translation of the single open reading frame. Polar amino acids are in heavy type. Numbering is on the basis of the transcription initiation site for *dsdA* being +1. Several restriction sites of interest are noted and underlined. Numbers in parentheses above three ATG and one GTG codon are molecular weights for the *dsdC* activator, if translation was initiated at the respective sites. Potential Shine-Dalgarno sequences are indicated by dots beneath consensus bases. Bases specifying a secondary *attB* site are boxed. Amino acids specifying a possible helix-turn-helix segment are boxed. A possible transcription terminator for *dsdC* is indicated by arrows.

to induce D-serine deaminase synthesis (Table 1). This result means that an intact *dsdC* gene is present on the *Bgl*II-*Eco*RI fragment and rules out a translation start for *dsdC* at −461 or −596. The ATG codon at −731 is thus most probably the *dsdC* translation start site, although the GTG codon at −685 is also a possible site.

We have as yet no physical evidence as to the location of the *dsdC* transcription start. The cloning experiment described above is not helpful, because in that case transcription probably proceeded from *lacP*. The only sequence in the region with good homology to the canonical sequence (18) is located between base pairs −461 and −488 (Fig. 1). It has four of the six bases of the consensus −35 sequence, lacking the highly conserved first T, and four of the six bases of the −10 sequence, lacking two of the less conserved bases. A transcription start at this point would yield an mRNA with a long leader region, beginning 105 base pairs from the first potential translation initiation codon, 235 base pairs from the start at base pair −731.

There is a palindromic sequence following the *dsdC* translation stop codon, extending between base pairs −1624 and −1646 (Fig. 1). It is followed by several T's and might serve as a transcription termination signal (16).

Codon usage in *dsdC*. We have calculated the codon usage for the activator, assuming a translation start at base pair −731 (Table 2). As may be seen, it is relatively random, much more similar to that observed in products of genes that are weakly expressed in *E. coli* than to that observed in products of strongly expressed genes (8). This random usage, as well as the weak Shine-Dalgarno sequences before potential translation initiation codons, is probably largely responsible for the very low level of activator in the cell.

Secondary *att*λ site in *dsdC*. We previously observed that it was relatively easy to obtain λ insertions into *dsdC* (2) and suspected that the gene contained a sequence with homology to *attB* (GCCT GCTTTTTTATACTAA CTTG; 16). We located such a sequence (GCAA GCTTTAATAATGAGG TATT) between base pairs −922 and −945 (Fig. 1). It has 10 bases of 23 in common with *attB*, including 9 of 13 in the first half of the site. We found no other site with *attB* or *attP* homology in *dsdC*.

Physical properties of the activator. We previously observed the activator to be a basic protein (10). In agreement, the sequence shows that the activator has an excess of basic amino acids (arginine and lysine) over acidic ones (glutamic and aspartic acids): 36:17. These are concentrated in the N-terminal two-thirds of the protein, with the C-terminal one-third being relatively neutral. The protein is polar in the N-terminal region, with an excess of acidic, basic, and uncharged polar amino acids over nonpolar amino acids in the ratio 136:109 for the entire protein.

Possible helix-turn-helix structure in the *dsdC* activator. Several transcription-regulatory proteins have been shown to have significant homology in regions that bind to specific promoters. These regions, corresponding to helices two and three of the Cro repressor, consist of two α helices connected by a short hinge segment (21, 23). We examined the amino acid sequence of the *dsdC* activator for evidence of such a region. We found a possible helix-turn-helix near the end of the polar region, between base pairs −1178 and −1244 (Fig. 1 and 2). It conforms to the four criteria of Takeda et al.

TABLE 1. *dsdC* expression in fusion strains

Strain	*dsd* genotype	D-Serine deaminase[a]	
		− D-Serine	+ D-Serine
MC4100	*dsdC*$^+$ *dsdA*$^+$	0.5	23
EM145-1	*dsdC*::Mu dl(Ap^r *lac*) *dsdA*$^+$	0.5	0.5
EM145-4	*dsdC*::Mu dl(Ap^r *lac*) *dsdA*$^+$(pEM192)	0.5	5.6

[a] Specific activity (15).

TABLE 2. Codon usage in *dsdC*[a]

Amino acid	Codon	*dsdC*	S	W	Amino acid	Codon	*dsdC*	S	W
Ala	GCU	3	33	17	Lys	AAA	6	49	31
	GCC	6	9	34		AAG	3	20	8
	GCA	6	23	20	Met	AUG	7	27	25
	GCG	3	25	28	Phe	UUU	5	7	29
Arg	CGU	3	42	19		UUC	5	22	19
	CGC	5	19	25	Pro	CCU	1	4	6
	CGA	3	1	5		CCC	4	0.4	9
	CGG	6	0.2	8					
	AGA	2	1	5		CCA	5	5	9
	AGG	5	0.2	3		CCG	1	31	19
Asn	AAU	6	2	19	Ser	UCU	3	18	7
	AAC	7	30	19		UCC	0	17	9
Asp	GAU	3	22	35		UCA	6	1	7
	GAC	3	39	20		UCG	1	2	12
Cys	UGU	5	2	6		AGU	3	2	11
	UGC	6	4	7		AGC	11	9	12
Gln	CAA	5	7	17	Thr	ACU	8	20	9
	CAG	7	32	32		ACC	3	26	23
Glu	GAA	4	63	40		ACA	3	3	6
	GAG	7	20	19		ACG	2	5	15
Gly	GGU	0	43	24	Trp	UGG	5	5	13
	GGC	3	33	27	Tyr	UAU	3	6	18
	GGA	1	1	8		UAC	5	19	12
	GGG	2	3	13	Val	GUU	0	37	21
His	CAU	6	4	18		GUC	1	8	13
	CAC	7	14	11		GUA	0	23	9
Ile	AUU	4	13	30		GUG	5	16	24
	AUC	3	15	23					
	AUA	5	0.4	5					
Leu	UUA	5	2	14					
	UUG	2	3	12					
	CUU	4	5	14					
	CUC	1	6	13					
	CUA	7	1	4					
	CUG	6	66	56					

[a] Codon usage in *dsdC* (assuming a translation start at −731) is compared with that in strongly (S) and weakly (W) expressed *E. coli* genes (tabulated by Grosjean and Fiers [8]). The *dsdC* data are for a total of 245 codons; the Grosjean and Fiers data are for relative codon usage per 1,000 codons.

(21): there are no prolines in the helical regions; the first hinge amino acid is glycine; amino acids 6, 10, 12, and 17, expected to be buried, are hydrophobic; and amino acid 7 is alanine.

DISCUSSION

The *dsdC* region contains a single open reading frame between base pairs −461 and −1476 (Fig. 1). The TGA stop codon at base pair −1476 defines the C-terminal end of the *dsdC* coding sequence. The most likely N terminus is the ATG codon at base pair −731, which would yield a protein of molecular weight 32,900, containing 249 amino acids, as compared with the experimental molecular weight of 33,000 (10). The −731 start has a recognizable Shine-Dalgarno sequence (7), in agreement with the consensus in five of nine bases. There is also a GTG codon at base pair −686, with a reasonable (six of nine bases) Shine-Dalgarno sequence. If it were used as a start codon, a protein of molecular weight 34,700 would result, also quite close to the experimental molecular weight. The distance between the translation starts for *dsdC* and *dsdA* (3) is thus surprisingly large, at least 760 base pairs.

The only recognizable transcription initiation sequence preceding *dsdC* spans base pairs −461 to −488 (Fig. 1). *dsdCp* serves as a readily identifiable promoter, with about 1/200 the strength of the fully expressed *dsdAp* (15), and should have some homology to the consensus transcription initiation sequence. The −461 to −488 sequence lacks the first T of the −35 sequence but otherwise contains the strongly conserved bases (18). A palindromic sequence between base pairs −1624 and −1646 (Fig. 1), followed by several T's, might serve as a transcription terminator for *dsdC* (16).

The cellular level of activator is extremely low, about eight molecules (dimers) per fully induced cell, compared with about 90,000 molecules of D-serine deaminase (monomers) (10, 13). The relatively random codon usage in the *dsdC* gene and the weak transcription initiation and Shine-Dalgarno sequences are very likely the basis for the difference. *dsdC* expression is not activated, and the three fixed parameters that determine its efficiency are relatively ineffectual.

The structure of the activator protein must provide for at least three functions: interaction with *dsdAp* DNA, interaction with the inducer D-serine, and formation of active dimers. We do not yet have any genetic evidence as to which parts of the protein are responsible for these functions. Most transcription-regulatory proteins examined so far interact with their respective promoter DNAs via a two-α-helix structure (12, 23), and we expected to find such a structure in the *dsdC* activator. There is a segment of the protein that conforms to the criteria of Takeda et al. (21) for a helix-turn-helix structure. It also has the appropriate amino acids in

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
dsdC	ala	ile	lys	cys	glu	leu	ala	his	ser	leu	gly	pro	asp	phe	his	thr	leu	cys	phe	arg	leu	leu
lacI	val	thr	leu	tyr	asp	val	ala	glu	tyr	ala	gly	val	ser	tyr	glu	thr	val	ser	arg	val	val	asn
trpR	met	ser	gln	arg	glu	leu	lys	asn	glu	leu	gly	ala	gly	ile	ala	thr	ile	thr	arg	gly	ser	asn
galR	ala	thr	ile	lys	asp	val	ala	arg	leu	ala	gly	val	ser	val	ala	thr	val	ser	arg	val	ile	asn
CAP	ile	thr	arg	gln	glu	ile	gly	gln	ile	val	gly	cys	ser	arg	glu	thr	val	gly	arg	ile	leu	lys
malT	leu	asp	ala	leu	lys	leu	ala	asn	arg	thr	gly	phe	ile	ser	his	phe	val	ile	glu	gly	glu	ala
cro	phe	gly	glu	thr	lys	thr	ala	lys	asp	leu	gly	val	tyr	glu	ser	ala	ile	asn	lys	ala	ile	his
cro 434	met	thr	gln	thr	glu	leu	ala	thr	lys	ala	gly	val	lys	gln	gln	ser	ile	gln	leu	ile	glu	ala
c_I	leu	ser	gln	gln	ser	val	ala	asp	lys	met	gly	met	gly	gln	ser	gly	val	gly	ala	leu	phe	asn
c_I 434	leu	asn	gln	ala	gln	leu	ala	gln	lys	val	gly	thr	thr	gln	gln	ser	ile	gln	gln	leu	gln	asn
c_{II}	leu	gly	thr	glu	lys	thr	ala	gln	ala	val	gly	val	asp	lys	ser	glu	ile	ser	arg	trp	lys	arg
	Helix 2													Helix 3								

FIG. 2. Possible helix-turn-helix segment of the *dsdC* activator compared with those observed in other regulatory proteins. The helix 2-helix 3 nomenclature refers to the corresponding region of Cro protein (23). The amino acids are arbitarily numbered 1 to 22. The *malT* sequence is from Cole and Raibaud (4); the *trpR* sequence is from Kelley and Yanofsky (12) and Gunsalus and Yanofsky (9). Data for the other proteins were compiled by Takeda et al. (21) and von Wilcken-Bergman and Müller-Hill (22).

several of the most highly conserved positions (Fig. 2): amino acid 6 is leucine, 7 is alanine, 10 is leucine, and 16 is threonine (12, 21). However, there is little analogy in the remainder of the segments, and it is not clear whether the proline at position 12 and the arginine at position 20 would fit properly. The *araC* activator is apparently an exception to the helix-turn-helix rule (11), and perhaps the *dsdC* activator is also. The latter is more basic and polar than the *lac*, *gal*, λ, and *trp* repressors (1, 8, 17, 21) and the CAP, *araC*, and *malT* activators (4, 5, 20), resembling more the N antitermination proteins of the lambdoid phages (6).

ACKNOWLEDGMENTS

This work was supported by American Cancer Society grant NP-557, and by Public Health Service grant GM11899 from the National Institutes of Health.

LITERATURE CITED

1. **Beyreuther, K.** 1980. Chemical structure and functional organization of *lac* repressor from *Escherichia coli*, p. 123–154. *In* J. H. Miller and W. S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Bloom, F. R., E. McFall, M. C. Young, and A. M. Carothers.** 1975. Positive control in the D-serine deaminase system of *Escherichia coli* K-12. J. Bacteriol. **121**:1092–1101.
3. **Bornstein-Forst, S. M., E. McFall, and S. Palchaudhuri.** 1987. In vivo D-serine deaminase transcription start sites in wild-type *Escherichia coli* and in *dsdA* promoter mutants. J. Bacteriol. **169**:1055–1060.
4. **Cole, S. T., and D. Raibaud.** 1986. The nucleotide sequence of the *malT* gene encoding the positive regulator of the *Escherichia coli* maltose regulon. Gene **42**:201–208.
5. **Cossart, P., and B. Gicquel-Sanzey.** 1982. Cloning and sequence of the *crp* gene of *Escherichia coli* K-12. Nucleic Acids Res. **10**:1363–1378.
6. **Franklin, N. C.** 1985. "N" transcription antitermination proteins of bacteriophage λ, Q21 and P22. J. Mol. Biol. **181**:85–91.
7. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and B. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
8. **Grosjean, H., and W. Fiers.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
9. **Gunsalus, R. P., and C. Yanofsky.** 1980. Nucleotide sequence and expression of *Escherichia coli trpR*, the structural gene for the *trp* aporexpressor. Proc. Natl. Acad. Sci. USA **77**: 7117–7121.
10. **Heincz, M. C., S. M. Bornstein, and E. McFall.** 1984. Purification and characterization of D-serine deaminase activator protein. J. Bacteriol. **160**:42–49.
11. **Hendrickson, W., and R. Schleif.** 1985. A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. Proc. Natl. Acad. Sci. USA **82**:3129–3133.
12. **Kelley, R. L., and C. Yanofsky.** 1985. Mutational studies with the *trp* repressor of *Escherichia coli* support the helix-turn-helix model of repressor recognition of operator DNA. Proc. Natl. Acad. Sci. USA **82**:483–487.
13. **McFall, E.** 1973. Role of adenosine 3′,5′-cyclic monophosphate and its specific binding protein in the regulation of D-serine deaminase synthesis. J. Bacteriol. **113**:781–785.
14. **McFall, E.** 1986. *cis*-acting proteins. J. Bacteriol. **167**:429–432.
15. **McFall, E., and M. C. Heincz.** 1983. Identification and control of synthesis of the *dsdC* activator protein. J. Bacteriol. **153**: 872–877.
16. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
17. **Ptashne, M.** 1980. λ repressor function and structure, p. 325–343. *In* J. H. Miller and W. S. Reznikoff (ed.), The operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of transcription. Annu. Rev. Genet. **13**:319–353.
19. **Sanger, F., and A. R. Coulson.** 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. **94**:441–448.
20. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol.

98:503–517.
21. **Takeda, Y., D. H. Ohlendorf, W. F. Anderson, and B. W. Matthews.** 1985. The structure of *cro* repressor protein, p. 233–263. *In* F. A. Jurnak and A. McPherson (ed.), Biological macromolecules and assemblies, vol. 2. Nucleic acids and interactive proteins. John Wiley & Sons, Inc., New York.
22. **von Wilcken-Bergman, B., and B. Müller-Hill.** 1982. Sequence of *galR* gene indicates a common evolutionary origin of *lac* and *gal* repressor in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **79:** 2427–2431.
23. **Weber, I. T., D. B. McKay, and T. A. Steitz.** 1982. Two helix DNA binding motif of CAP found in *lac* repressor and *gal* repressor. Nucleic Acids Res. **10:**5085–5102.
24. **Weisberg, R. A., and A. Landy.** 1983. Site-specific recombination in phage lambda, p. 211–250. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Yanisch-Penon, C., J. Vieira, and J. Messing.** 1983. Improved m13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.