JOURNAL OF BACTERIOLOGY, Feb. 1997, p. 1051–1058 Vol. 179, No. 4
0021-9193/97/$04.00+0

Critical Base Pairs and Amino Acid Residues for Protein-DNA Interaction between the TyrR Protein and *tyrP* Operator of *Escherichia coli*

J. S. HWANG, J. YANG, AND A. J. PITTARD*

Department of Microbiology, The University of Melbourne, Parkville, Victoria 3052, Australia

Received 8 August 1996/Accepted 26 November 1996

In *Escherichia coli* K-12, the repression of *tyrP* requires the binding of the TyrR protein to the operator in the presence of coeffectors, tyrosine and ATP. This operator contains two 22-bp palindromic sequences which are termed TyrR boxes. Methylation, uracil, and ethylation interference experiments were used to identify the important sites in the TyrR boxes that make contacts with the TyrR protein. Methylation interference studies demonstrated that guanines at positions +8, −5, and −8 of the strong TyrR box and positions +8, −4, and −8 of the weak box are close to the TyrR protein. Uracil interference revealed that strong van der Waals contacts are made by the thymines at position −7 and +5 of the top strands of both strong and weak boxes and that weaker contacts are made by the thymines at positions +7 (strong box) and −5 and +7 (weak box) of the bottom strand. In addition, ethylation interference suggested that the phosphate backbone contacts are located at the end and central regions of the palindrome. These findings are supported by our results derived from studies of symmetrical mutations of the *tyrP* strong box. Overall, the results confirm the critical importance of the invariant (G · C)(C · G)8 base pairs for TyrR recognition and also indicate that interactions with (T · A)(A · T)7 are of major importance. In contrast, mutations in other positions result in weaker effects on the binding affinity of TyrR protein, indicating that these positions play a lesser role in TyrR protein recognition. Alanine scanning of both helices of the putative helix-turn-helix DNA-binding motif of TyrR protein has identified those amino acids whose side chains play an essential role in protein structure and DNA binding.

The *tyrP* gene of *Escherichia coli* K-12, which codes for the tyrosine-specific permease, is transcriptionally regulated by the TyrR protein (27). In *tyrR*$^+$ strains, its expression is repressed by tyrosine and activated by phenylalanine (15, 37). Previous genetic and biochemical studies have identified two adjacent binding sites for the TyrR protein in the regulatory region of *tyrP* (3, 15). These sites, referred to as TyrR boxes, contain imperfect palindromes related to the sequence TGTAAAN$_6$T TTACA (Fig. 1A). Seventeen of these TyrR boxes have been identified in the regulatory regions of the genes of the TyrR regulon (27) and are shown in Fig. 1B. On the basis of the degree of similarity of each box with the consensus palindromic sequence and the binding affinity of TyrR protein for each of the boxes, they have been classified as either strong or weak. In the case of *tyrP*, the strong box lies upstream, just outside the RNA polymerase-binding region, whereas the adjacent weak box overlaps the −35 region of the *tyrP* promoter. Tyrosine-mediated repression has been shown to involve cooperative binding of TyrR molecules to the strong and weak boxes, which are required to be on the same face of the DNA helix (2). This has subsequently been shown to involve the hexamerization of the TyrR protein in the presence of both tyrosine and ATP (38). Phenylalanine-mediated activation requires only a single functional strong box upstream of the promoter. The position of this box is important for activation (2), which involves interaction between the TyrR protein and the α subunit of the RNA polymerase (17).

* Corresponding author. Mailing address: Department of Microbiology, The University of Melbourne, Parkville, Victoria 3052, Australia. Phone: 61 3 9344 5679. Fax: 61 3 9347 1540. E-mail: aj.pittard@microbiology.unimelb.edu.au.

The TyrR polypeptide contains 513 amino acid residues (40), which appear to comprise two structural domains and a less structured carboxyl terminus (11). The carboxyl-terminal region of the TyrR protein contains a sequence of amino acids (from residues 483 to 502) which resembles a typical helix-turn-helix (HTH) DNA-binding motif (Fig. 1C) (7, 13, 31). Mutational studies and negative dominance tests have confirmed the importance of residues of the second α-helix for both TyrR-mediated repression and activation (40). So far, attempts to determine the crystal structure of TyrR protein have not been successful. In the absence of such information, it is nevertheless possible to identify key features of the specific interactions between TyrR protein and DNA targets by in vitro binding studies and different genetic approaches.

In this paper, we report in vitro studies in which methylation of purines, substitution of uracil for thymines, and ethylation of phosphates are examined for their effects on the binding of TyrR protein. In addition, we have used site-directed mutagenesis to construct all possible symmetrical substitutions in the palindromic arms of the strong TyrR box of *tyrP* and have carried out alanine scanning experiments to analyze the importance of amino acid residues in helices 1 and 2 of the putative DNA-binding motif.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophages. All bacterial strains used in this study were derivatives of *E. coli* K-12, and their relevant characteristics are shown in Table 1. The plasmids and bacteriophages used are also shown in Table 1.

Media and chemicals. The minimal medium used was prepared from the 56/2 buffer of Monod et al. (23) and supplemented with 0.2% glucose and appropriate growth factors. The concentrations of various growth factors and antibiotics and the sources of various chemicals used in this study were as described previously (40).

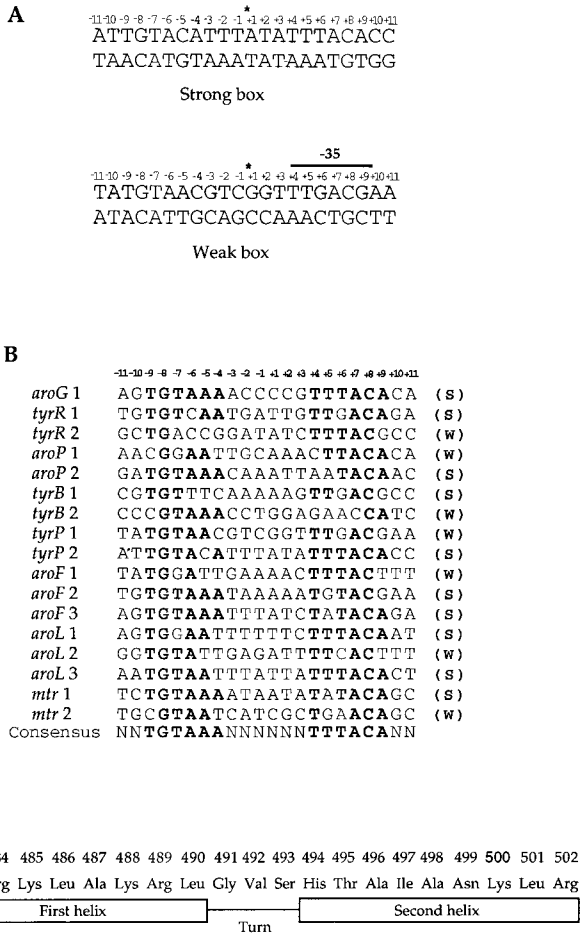

FIG. 1. (A) Nucleotide sequences of the strong and weak TyrR boxes of *tyrP*. The numbers above the sequences indicate the base positions within each box. The center of the palindromic symmetry is marked by an asterisk, and the −35 *tyrP* promoter is overlined. In *tyrP*, the strong and weak boxes are separated by a single base pair. (B) Comparison of the properties of the TyrR boxes. All known TyrR box sequences are aligned according to the central axis of symmetry. The consensus TyrR box sequence is shown, and the positions of the bases are indicated by numbers above the top sequence. Bases corresponding to the consensus sequence are indicated in boldface S, strong TyrR box; W, weak TyrR box. (C) Amino acid sequence of the HTH motif of the TyrR protein. The numbers above the sequence are the residue numbers in the full protein sequence.

Synthesis of a 151-bp *Bam*HI-*Hin*dIII fragment of the *tyrP* operator by PCR. The 151-bp *Bam*HI-*Hin*dIII fragment containing the *tyrP* operator was generated by PCR using mpMU330 as a template. This PCR fragment was cloned into M13mp19, in which its sequence was verified. Following the insertion of the *Bam*HI-*Hin*dIII fragment into pUC19, the resulting construct was used in the methylation and ethylation interference experiments.

Methylation interference. The pUC19 derivative containing the 151-bp *Bam*HI-*Hin*dIII *tyrP* fragment was digested with *Bam*HI (or *Hin*dIII) and then labeled by using Klenow fragment in the presence of [α-^{32}P]dGTP (or [α-^{32}P]dATP). Following ethanol precipitation, the second digestion was done with *Hin*dIII (or *Bam*HI). After gel purification, the end-labeled DNA ($\sim 2 \times 10^6$ cpm) was resuspended in 50 mM sodium cacodylate (pH 8.0)–1 mM EDTA (pH 8.0)–10 mM $MgCl_2$ and modified by the addition of dimethyl sulfate and incubation at 20°C for 10 min. The modified DNA samples were precipitated twice with ethanol, washed, and dried. The mobility gel shift assay was carried out by incubating the modified DNA ($\sim 10^5$ cpm) with TyrR protein in a buffer containing 5 mM Tris-HCl (pH 7.6), 80 mM KCl, 8 mM $MgCl_2$, 1 mM dithiothreitol, 0.2 mM ATP, 1 mM tyrosine, and 4% (vol/vol) glycerol for 20 min at 37°C. The concentration of TyrR protein was determined empirically such that the amounts of bound and unbound fractions of modified DNA were approximately equal. Following the incubation, the mixture was loaded on a preelectrophoresed 5% polyacrylamide gel and electrophoresed in 50 mM Tris-borate (pH 7.0)–1 mM $MgCl_2$–0.2 mM ATP–1 mM tyrosine at 4°C. The gel was autoradiographed for 2 h at 4°C, and bound and unbound DNA fractions were eluted from the gel by a gel eluter (The Australian Chromatography Company-Hoefer). The pellets of both bound and unbound DNA were obtained by ethanol precipitation. Finally, the G or A>G base cleavage reactions were carried out as described by Maxam and Gilbert (19). The products were electrophoresed on a 6% denaturing polyacrylamide sequencing gel against a G+A ladder.

Uracil substitution. The analysis for contacts with the 5-methyl group of thymine was based on the method described by Pu and Struhl (29). To replace the thymine randomly by deoxyuracil, PCR amplification was performed with mpMU330, which contains the *tyrP* regulatory region, as a template and two oligonucleotides which hybridize the regions flanking the *tyrP* regulatory region as primers. The PCR mixture contained 20 pmol of γ-^{32}P-end-labeled forward primer (or 20 pmol of γ-^{32}P-end-labeled reverse primer) and 20 pmol of reverse primer (or 20 pmol forward primer), 0.2 pmol of mpMU330 DNA template, 0.2 mM each dATP, dCTP, dGTP, and dTTP, 0.05 mM dUTP, 15% glycerol, and 2 U of *Taq* polymerase in a total volume of 50 μl. The PCR buffer contained 100 mM Tris-Cl (pH 8.2), 500 mM KCl, and 0.1% gelatin. Amplification was carried out for 15 cycles at 94°C for 1 min, 47°C for 2 min, and 72°C for 2 min. The PCR product was purified on a 5% polyacrylamide gel and precipitated by ethanol. The mobility gel shift assay was then performed on the uracil-substituted DNA exactly as described for the methylation interference experiment. The bound and unbound fractions were subjected to uracil-*N*-glycosylase treatment at 37°C for 60 min and cleavage by 1 M piperidine. The resultant DNA fragments were electrophoresed as described above.

Ethylation interference. The experiment was based on the protocol described by Heuer and Hillen (14). The end labeling of DNA fragments containing the *tyrP* regulatory region was done as described above for the methylation interference experiment. About 2×10^6 cpm of 3′-end-labeled DNA was modified by *N*-ethylnitrosourea (saturated in 95% ethanol; Sigma) in a buffer containing 50 mM sodium cacodylate (pH 8.0) and 1 mM EDTA (pH 8.0). Following incubation at 50°C for 60 min, the ethylated DNA was precipitated with ethanol. A mobility gel shift assay as described for the methylation interference experiment was carried out to separate the protein-DNA complex and free DNA. The bound and unbound DNA fragments were recovered from the gel and resuspended separately in 30 μl of a buffer containing 10 mM sodium phosphate (pH 7.0) and 1 mM EDTA. Following addition of 5 μl of 1 M NaOH, the reaction mixtures were incubated at 90°C for 30 min, after which the reactions were stopped by addition of 5 μl of 1 M HCl. The DNA was precipitated with ethanol and then electrophoresed on a 6% denaturing polyacrylamide gel. The autoradiogram was quantified with a Molecular Dynamics scanning densitometer.

Site-directed mutagenesis. In vitro mutagenesis was performed on mpMU330 containing the *tyrP* promoter-operator region by the method of Vandeyar et al. (35), using a commercially available kit from United State Biochemical Corp. Following the isolation of desired mutations, each *tyrP* fragment was sequenced to ensure that only the desired change was present.

The construction of alanine substitutions in the HTH motif of TyrR was carried out as described by Yang et al. (40).

β-Galactosidase assay. β-Galactosidase activity of mid-log-phase cultures was assayed as described by Miller (21). The data are the means of three independent assays.

RESULTS AND DISCUSSION

Determination of guanine contacts in the major groove. Dimethyl sulfate can methylate the N7 group of guanine and the N3 group of adenine, although the rate of methylation is slightly higher for guanine than for adenine. The N7 of guanine is located in the major groove of the DNA double helix, whereas the N3 of adenine is exposed in the minor groove. To assess the role of the individual guanine residues (in the major groove) and the adenine residues (in the minor groove) in the specific interaction between the *tyrP* operator and the TyrR protein, we performed experiments in which the methylation of the DNA was tested for its effects on the binding of TyrR protein (for details, see Materials and Methods).

As shown in Fig. 2A, methylation of the guanine residues, on the top strand, at positions −8 of both the strong and weak TyrR boxes causes strong interference with binding. Bands corresponding to these two positions are relatively intense in the unbound DNA fraction but are completely depleted in the bound DNA fraction. On the bottom strand, we also observe strong interference at guanines +8 of both TyrR boxes and guanine −4 of the weak TyrR box (Fig. 2B). Methylation of the guanine residue at position −5 of the strong box causes a weak interference, as the band from the unbound fraction is slightly denser than that from the bound fraction (Fig. 2B).

TABLE 1. *E. coli* K-12 strains, plasmids, and phages used in this work

Strain, plasmid, or phage	Relevant characteristics	Reference or source
Strains		
JP3561	*thr-1 leu-1 lacZ*Δ*M15 supE44 tonA2 gyrA379 aroL478*::Tn*10*	15
JP4822	JP3561 *tyr366*	P1 *kc* transduction
JP8042	Δ*lacU169 recA56 tyrR366*	40
Plasmids		
pUC19	Ap^r; pMB1 derivative	25
pMU2385	*galK′-lac′Z* Tp IncW; low-copy-number transcriptional fusion vector derived from pMU575	28
pMU6218	Km^r; pSU39, the *Hin*dIII site in the polylinker cloning region was eliminated	39
pMU2003	Tp^r; 0.25-kb *tyrP* fragment in pMU575 (41); *tyrP-lacZ* transcriptional fusion	40
pMU2493	Tp^r; *aroF* fragment in pMU575 (41); *aroF-lacZ* transcriptional fusion	B. Dickson
Phages		
M13mp19	*lac*POZ′	20
M13tg130	*lac*POZ′	16
M13tg131	*lac*POZ′	16
mpMU36	2.2-kb *tyrR* fragment in M13tg130	39
mpMU330	M13tg131 derivative carrying 0.32-kb *Bam*HI fragment of $tyrP^+$	1

Similar experiments using alkaline cleavage were performed to analyze methylation at both guanine and adenine residues (the A>G reaction). The interference results are essentially the same as those observed with piperidine cleavage, indicating that none of the adenine residues is in close contact with the DNA, via the minor groove (data not shown).

Determination of thymine contacts in the major groove. Thymine bases can interact with proteins via their methyl groups. By substituting uracil for thymine, this contact point is removed. The effects of such substitutions on the binding of TyrR protein were examined in the same general way as described for methylation.

The uracil interference experiment was carried out as described in Materials and Methods, and the results are shown in Fig. 3. Inspection of the gels reveals that the intensities of the bands representing fragments cleaved at positions −7 and +5 of both the strong and weak TyrR boxes in the top strand are increased in the unbound fraction and decreased in the bound fraction (Fig. 3A). This finding indicates that substitution of uracil at these positions weakens the binding of TyrR protein. With reference to the bottom strand, although the effects are weaker, possible interactions are seen at positions +7 of both the strong and weak TyrR boxes and position −5 of the weak box (Fig. 3B).

To highlight these interference effects on the bottom strand, we compared a bound track from an experiment with low ratios of TyrR protein to DNA which produces one part bound DNA to two parts unbound with an unbound track from an experiment where the ratio is such that there are two parts bound to one part unbound. Figure 3C shows that the specific effects at positions +7 and −5 are more evident; in addition, a very weak interference can be seen at position +9 in the bottom strand of the strong box.

Determination of phosphate backbone contacts by TyrR. Ethylation interference was used to identify the region where the phosphate backbone makes close contacts with the TyrR protein. The experiment was performed as described in Materials and Methods, and the samples were analyzed on a sequencing gel. The autoradiogram of the sequencing gel and the densitometric scan of that gel are shown in Fig. 4.

A smear between bands in the bound and unbound lanes of both top and bottom strands (Fig. 4) is generated as a result of the *N*-ethylnitrosourea reacting with the oxygen atoms of thymine, guanine, or cytosine in addition to the phosphate backbone (34). Examination of the gels reveals some regions that show an enhanced intensity of bands in the unbound fraction. In the top strand, ethylation of phosphates 5′ to positions −8,

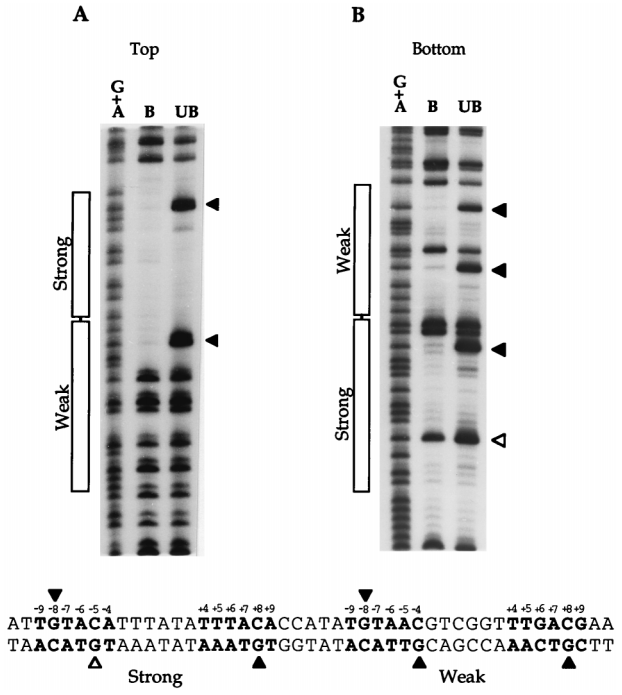

FIG. 2. Methylation interference of the top and bottom strands of the *tyrP* operator. The boxes show the location of strong and weak TyrR boxes. G+A, Maxam-Gilbert sequencing ladder. Bound and unbound DNA fragments which were separated by the mobility gel shift experiment are labeled B and UB, respectively. The sequence given at the bottom corresponds to the sequence shown in the autoradiogram, and the palindromic sequences of both TyrR boxes are shown in boldface. Filled triangles indicate strong interference of methylated guanines on the binding of the TyrR protein; the open triangle defines weak interference that is exerted by the methylation of guanine −5 (strong box, bottom strand).

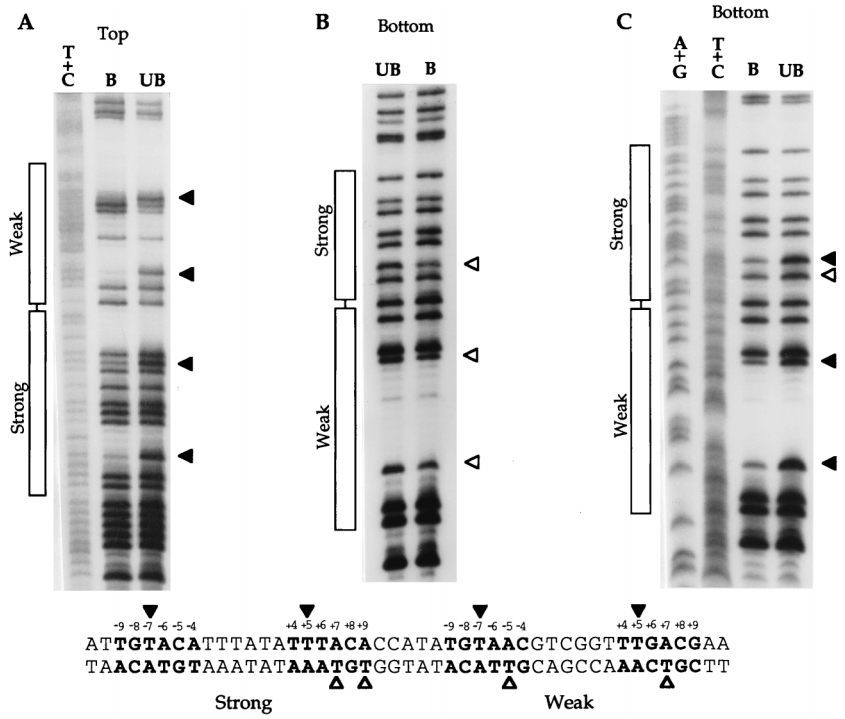

FIG. 3. Uracil interference with binding of TyrR to the top and bottom strands of the *tyrP* operator. A+G and T+C are the Maxam-Gilbert sequencing ladders; B and UB mark the bound and unbound fractions, respectively. The strong thymine contacts are indicated by filled triangles; the weaker thymine contacts are shown by open triangles. In panel C, special conditions for the mobility gel shift assay were applied to achieve maximal thymine contacts of the *tyrP* operator by TyrR (for details, see Results).

−9, and −10 and positions +1, +2, and +3 of the strong box and positions −8, −9, and −10 of the weak box clearly result in the enhancement of fragments in the unbound fraction. In the bottom strand, ethylation of phosphates 5′ to positions +8, +9, and +10 in both the strong and weak boxes shows the same strong positive interference in TyrR binding. Additional but weaker regions showing ethylation interference are in the top strand, 5′ to positions +3 and +4 of the weak box, and in the bottom strand, 5′ to positions −3 and −4 in the weak box and −2 and −3 in the strong box.

Some unexplained observations are regions of apparent enrichment of fragments in the bound fraction caused by ethylation of phosphates in the top strand of the weak box, 5′ to positions −1, −2, and +1 and positions +6, +7, and +8.

As has been reported for other systems such as TetR, GalR, and 434 repressor (8, 14, 18), phosphate contacts flank each recognition sequence of operators. Our ethylation interference results confirm the prediction that the recognition sequence for the TyrR protein is 5′TGTAAA3′. The weakness of the interference effects seen near the center of the palindrome compared with those seen in other systems may reflect some additional structural changes in the case of TyrR-*tyrP* interactions. This may be brought about by the complexity of the cooperative binding involving adjacent strong and weak TyrR boxes, which do not behave in exactly the same way (2).

Determination of the affinity of TyrR protein for the *tyrP* symmetrical mutants. To understand the role played by various base pairs of the strong TyrR box of the *tyrP* operator in specific interactions with TyrR protein and to examine the degree of tolerance to base changes at each of these positions, we used site-directed mutagenesis to construct a complete set of symmetrical mutations involving single-base-pair changes in each of the palindromic arms of the strong box. Although there have been many separate reports of base changes which interfere with the functions of different TyrR boxes (3, 5, 9, 12, 15, 30, 41), there has not been any systematic study of the importance of each of the bases for TyrR binding. To measure the binding affinities of these operator variants in vivo, the mutagenized DNA fragments were cloned separately into plasmid pMU2385 to form *tyrP-lacZ* transcriptional fusions. The resulting plasmids were then introduced individually into strains JP4822 (*tyr366*), JP3561 (haploid *tyrR*$^+$), and JP4822/pMU3325 (a multicopy plasmid carrying the *tyrR*$^+$ gene). The cells were grown under repressing conditions, and β-galactosidase activities were assayed. The results are shown in Table 2, and the repressibility of each *tyrP* operator variant is represented as a repression ratio where the β-galactosidase value in the *tyrR* genetic background (*tyr366*) is divided by that in the *tyrR*$^+$ genetic background (either haploid or multicopy *tyrR*$^+$). The nomenclature that we have adopted to describe the various mutants of the strong box of *tyrP* is as follows: the two letters in the first pair of parentheses denote the base pair in the left arm of the palindrome, and the two letters in the second pair of parentheses describe the symmetrical base pair in the right arm of the palindrome; the number that follows indicates the position. For example, (G · C)(C · G)8 denotes the G · C base pair in the left arm at position −8 and the C · G base pair in the right arm at +8. The first letter of each hydrogen-bonded pair indicates the base in the top strand.

On the basis of these results, it is possible to identify "nonpermissible" symmetrical bases in the arms of the strong TyrR box of *tyrP*, i.e., the symmetrical base pair substitutions which are impaired in repression. Inspection of all of the natural TyrR boxes (Fig. 1B) reveals that none of them contains such nonpermissible symmetrical bases. The specific effects pro-

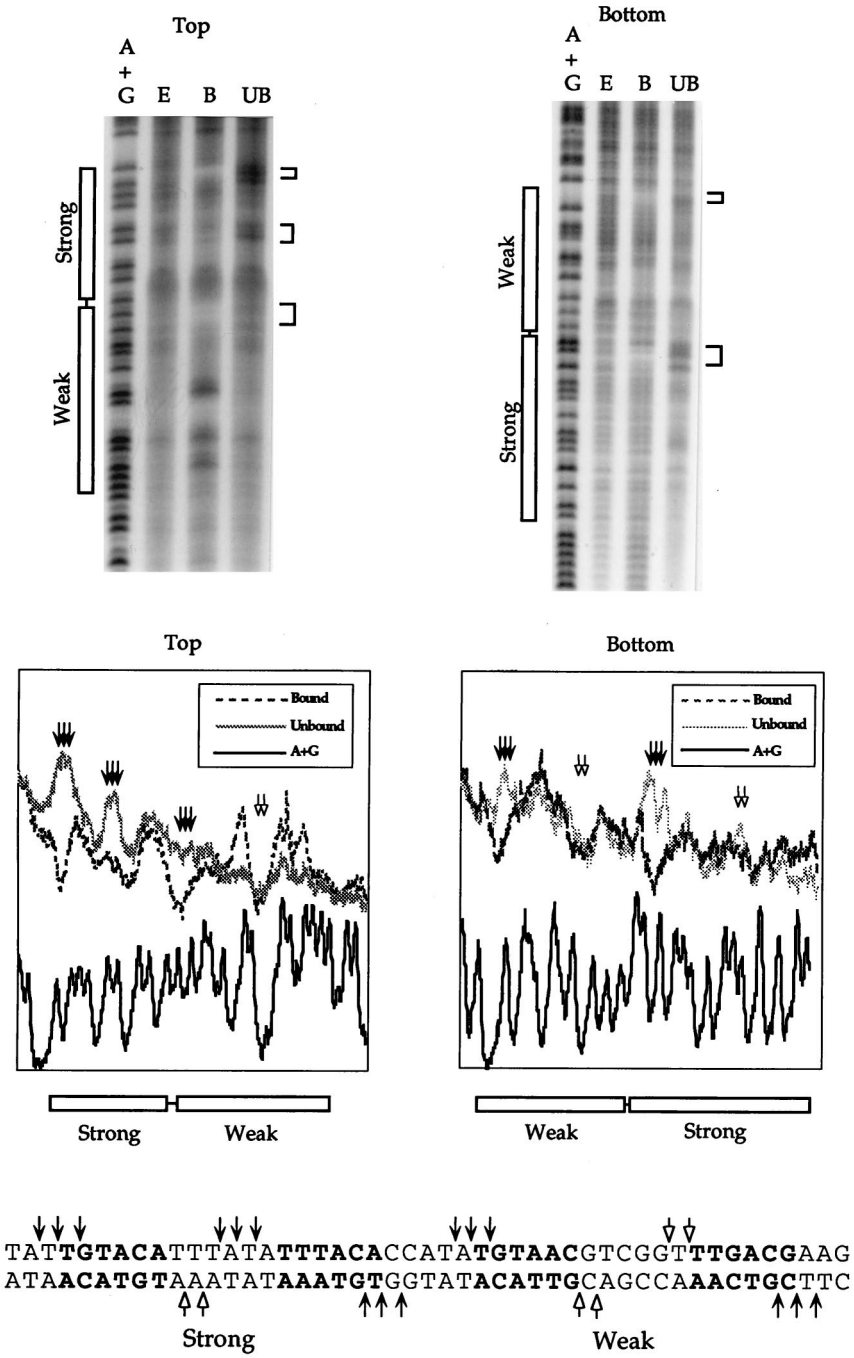

FIG. 4. Ethylation interference with TyrR protein-*tyrP* operator interaction. A+G, Maxam-Gilbert sequencing ladder; E, cleavage of ethylated DNA which has not undergone the mobility gel shift assay; B, ethylated DNA bound by TyrR; UB, ethylated DNA not bound by TyrR. The strong and weak TyrR boxes are shown in boxes. The small brackets indicate the regions containing the modified phosphate groups that strongly interfere with TyrR protein binding. Below each autoradiogram is the densitometer scan for the A+G, bound, and unbound fractions. The arrows above the peaks represent the base positions whose 5′ ethylated phosphate group causes interference to the TyrR binding. Filled arrows show strong interference; open arrows define weaker interference of ethylated phosphate backbones with the binding of TyrR.

duced by mutations are shown in Table 2 and are discussed below along with the other results reported above.

Attributes of the base pairs comprising the arms of the palindrome. (i) Position 9. Although the (T · A)(A · T)9 arrangement is well conserved among the various TyrR boxes of the TyrR regulon, mutational studies showed that substitution by (C · G)(G · C) at this position improves repressor binding, (G · C)(C · G) has little effect, and only (A · T)(T · A) causes a significant reduction in repression. The uracil interference experiments reported in this paper showed only a very weak interaction at position +9 of the bottom strand. Recently, Bailey et al. (4) studied the interactions between TyrR protein and the strong TyrR box of the *tyrR* gene by using fluorescence footprinting and reported interactions between TyrR protein and the DNA at position −9 of the top strand.

(ii) Position 8. (G · C)(C · G)8 is invariant in all of the TyrR boxes of the TyrR regulon. Any base substitution at this position totally destroys repressor binding, and repression of op-

TABLE 2. β-Galactosidase assay of strains carrying a plasmid containing symmetrical mutations in the *tyrP* strong box

Mutation[b]															Tyrosine-mediated repression ratio[a]		
															tyr366	*tyrR*$^+$ (haploid)	*tyrR*$^+$ (multicopy)
−9	−8	−7	−6	−5	−4		+4	+5	+6	+7	+8	+9					
T	G	T	A	C	A		T	T	T	A	C	A			75	7.1	78.0
A												T			99	2.4	38.3
C												G			87	11.3	102.0
G												C			100	5.2	51.5
•	A										T	•			89	1.3	1.1
•	C										G	•			92	1.0	1.3
•	T										A	•			111	1.2	1.3
•	•	A								T	•	•			93	1.0	1.2
•	•	C								G	•	•			90	1.7	8.0
•	•	G								C	•	•			90	3.8	56.3
•	•	•	C						G	•	•	•			114	10.9	60.0
•	•	•	G						C	•	•	•			110	1.9	14.5
•	•	•	T						A	•	•	•			98	3.4	27.5
•	•	•	•	A				T	•	•	•	•			71	5.5	78.0
•	•	•	•	C				G	•	•	•	•			86	11.8	94.0
•	•	•	•	G				C	•	•	•	•			92	1.4	2.8
•	•	•	•	T				A	•	•	•	•			99	3.4	53.0
•	•	•	•	•	C		G	•	•	•	•	•			106	2.1	37.0
•	•	•	•	•	G		C	•	•	•	•	•			110	2.3	31.0
•	•	•	•	•	T		A	•	•	•	•	•			99	1.9	57.0

[a] Defined as specific activity of β-galactosidase in the *tyrR366* strain divided by the specific activity in *tyrR*$^+$ cells grown in minimal medium supplemented with 1 mM tyrosine.

[b] Substituted bases are shown in boldface. The base positions are numbered from the central base pair to the outside of the box; bases to the right of the central position are indicated with +, whereas bases to the left are indicated with −.

erator variants at this position is not restored by increasing the level of TyrR protein. This is in contrast to that observed at position 9, where (G · C)(C · G) can substitute for (T · A)(A · T), and at positions 5 and 6, where (C · G)(G · C) and (A · T)(T · A) appear to function equally well. Whereas the O6 of guanine and the N4 of cytosine offer hydrogen bonding in the major groove equivalent to that of the O4 of thymine and the N6 of adenine, the N7 hydrogen acceptor of a G · C pair and the N7 hydrogen acceptor of a T · A pair are not equivalent. At positions 9, 5, and 6, it would appear that any critical hydrogen bonds are restricted to guanine O6 and/or cytosine N4 and thymine O4 and/or adenine N6. The failure of any other base pair to substitute for G · C at position 8 indicates that critical hydrogen bonds involve the N7 of guanine with at least one other hydrogen bonding group of the G · C pair. The involvement of the N7 of guanine is confirmed by the very strong methylation interference reaction involving guanine bases at −8 in the top strand and +8 in the bottom strand.

(iii) Position 7. (T · A)(A · T)7 is conserved in all of the strong TyrR boxes. These base pairs can be substituted by (G · C)(C · G) with only a 50% reduction in repression, but substitution by (A · T)(T · A) or by (C · G)(G · C) almost completely destroys the binding of TyrR. These results indicate (i) that (T · A)(A · T)7 is probably involved in important hydrogen bonds with the amino acids of the TyrR protein and (ii) that these bonds probably involve either the N6 of adenine or the O4 of thymine. Uracil interference experiments show strong interactions involving thymines at position −7 in the top strand and +7 in the bottom strand. This would suggest that the methyl group of thymine interacts with the protein and may explain the 50% reduction in repression when (T · A)(A · T) is substituted by (G · C)(C · G). Bailey et al. (4) also identified position −7 in the top strand as an interacting base in their fluorescence studies.

(iv) Position 6. Although (A · T)(T · A)6 is well conserved among the TyrR boxes, substitution by (T · A)(A · T) reduces repression only by about half, and substitution by (C · G)(G · C) results in increased repression. However, substitution by (G · C)(C · G) at this position is not tolerated, as it almost completely destroys repression. This latter substitution would create the sequence TGTG, which in the cyclic AMP receptor protein site has been associated with DNA bending (32), and this might account for its dramatic effects.

(v) Position 5. Although (A · T)(T · A)5 is well conserved in TyrR boxes, the left arm of the strong box contains a C · G substitution at this position. Changing this to the consensus (A · T)(T · A)5 barely alters repression, and changing it to the symmetrical (C · G)(G · C)5 causes increased repression. Substituting a (T · A)(A · T)5 results in only a slight decrease in repression, whereas again, substitution by (G · C)(C · G)5 almost completely destroys repression. In this case, such a substitution does not create a TGTG sequence as before. Uracil interference experiments indicate that thymines at position +5 in the top strand and −5 in the bottom strand play a significant role in protein-DNA interactions. Bailey et al. also have reported that thymine at +5 in the top strand of the TyrR box of the *tyrR* gene is important for TyrR protein-DNA interaction (4). The finding that the (C · G)(G · C)5 substitution results in stronger binding even though the methyl group of thymine is no longer available to interact with the protein may imply that a new and additional hydrogen bond can be formed with the protein by the (C · G)(G · C) pair. The guanine residue which occurs at position −5 in the bottom strand of the strong box gives only a very weak interaction in the methylation interference test.

(vi) Position 4. (A · T)(T · A)4 is well conserved in the strong boxes of the TyrR regulon, and the T · A base pair at position +4 is moderately well conserved in the weak boxes (five of seven). However, the A · T base pair at −4 is poorly conserved in the weak boxes (two of seven). The thymines at position 4 did not show any uracil interference, and the mutational studies produced unusual results. Substitution of (A · T)(T · A) by any other symmetrical base pairs at this position reduces repression to about one-third, but repression was restored to 50% or more if the levels of TyrR protein were increased. If the base pairs at position 4 form specific bonds with the TyrR protein, these bonds would appear not to be essential but to contribute to the overall affinity between the protein and the DNA.

Alanine scanning mutagenesis of the HTH region of TyrR protein. A putative DNA-binding domain with the characteristics of a classical HTH motif has been identified in the carboxyl-terminal region of the TyrR protein, and the activities of a number of mutants with amino acid changes in this region have been described (40). To investigate the contributions that the various side chains of these amino acids make to protein structure or to protein-DNA interactions, we used site-directed mutagenesis to change each of the amino acids in helix 1 and helix 2 to alanine (Materials and Methods). These various *tyrR* alleles were then introduced into a *tyrR366* strain on plasmid pMU6218, and the effectiveness of the mutant proteins as repressors was measured by using *tyrP-lacZ* and *aroF-lacZ* as reporters. The results are shown in Table 3. The alanine substitutions at positions 484, 489, 494, 500, and 501 had a major effect on repression, and the alanine substitutions at positions 483, 485, 486, 490, 495, 497, 499, and 502 had a lesser but significant effect on repression. By analogy with other well-

TABLE 3. Specific activities of β-galactosidase expressed from *tyrP-lac* and *aroF-lac* fusions in host strain JP8042 (*tyrR366*) and derivatives which contain various mutant *tyrR* alleles

tyrR allele on pSU39	β-Galactosidase sp act (U)[a]		Repression ratio[b]	
	tyrP-lacZ	*aroF-lacZ*	*tyrP-lacZ*	*aroF-lacZ*
None	271	5,708	1	1
tyrR$^+$	7	11	39	519
TA483	50	128	5	45
RA484	251	4,695	1	1
KA485	46	217	6	26
LA486	32	305	8	19
KA488	8	13	34	439
RA489	229	1,307	1	4
LA490	26	133	10	43
HA494	303	5,529	0.9	1
TA495	31	386	9	15
IA497	65	316	4	18
NA499	30	56	9	102
KA500	232	5,586	1	1
LA501	133	1,755	2	3
RA502	36	251	8	23

[a] Cells were grown in minimal medium supplemented with 1 mM tyrosine.
[b] Specific activity of β-galactosidase of the *tyrR366* strain divided by that of *tyrR*$^+$ or *tyrR* mutant strain.

characterized regulatory proteins which contain an HTH motif (13, 26), T483, R484, H494, T495, N499, and R502 are predicted to face the DNA helix, and it is therefore plausible to assume that the side chains of these amino acids are directly involved in interaction with DNA. Other amino acids, namely, K485, L486, L490, I497, K500, and L501, which are predicted to face away from the helix, may play a role in maintaining the structure of the HTH DNA-binding motif. Among all of the amino acids, only the side chain of K488 appears to play no role in either DNA binding or protein structure. The side chains of R484 and H494 clearly play a critical role in binding of the TyrR protein to DNA.

If the amino acids of the TyrR protein HTH are substituted for the equivalent amino acids of Cro repressor or catabolite gene activator protein (CAP) by using the published coordinates (22, 36), it is possible to produce a hypothetical representation of the TyrR protein HTH showing the disposition of the various side chains. The models obtained with Cro or CAP are virtually superimposable. The CAP-substituted model is shown in Fig. 5, with the side chains of R484, H494, T495, N499, and R502 highlighted. As can be seen, the side chains of R484 and H494 lie close to each other; the side chains of T495, N499, and R502 can also be seen directed toward the same face. If the TyrR-CAP model is positioned next to the DNA of a TyrR box, hydrogen bonds can be formed between the epsilon amino group of R484 and the N7 of guanine and between a ring nitrogen of H494 and the O6 of the same guanine.

Some interesting parallels can be drawn with the published structure of the PurR repressor, where it has been shown that a histidine in helix 2 and an arginine in the loop following that helix can interact in an interesting way (33). Van der Waals forces between the histidine imidazole ring and the side chain of arginine are postulated to position the guanidino nitrogen atoms to donate hydrogen bonds to the N7 and O6 acceptors of the guanine at position 4 (33). Our own results suggest that N7 of the invariant guanine at position 8 of the *tyrP* operator must play an important role in any reactions with TyrR. In the case of the *pur* regulon, the guanine at position 4 is absolutely conserved in the various *pur* operators and is the only one to show methylation interference sensitivity.

The first helix of the DNA-binding motif of TyrR is highly homologous to that of TetR, and residues at positions 1, 2, 3, 4, 5, and 8 are identical between the two proteins. The arginine residue at position 2 of the first helix of TetR has been postulated to form two hydrogen bonds with a guanine at position 2 of the TetR-binding site (6). The N7 of this guanine is also protected against methylation by the binding of TetR (14). The previously reported failure of TyrR mutants in which H494 had been replaced by phenylalanine, tyrosine, or leucine to repress genes of the TyrR regulon (40) supports the contention that the nitrogen in the imidazole ring of histidine plays an important role in protein-DNA interactions. We believe that the data in this paper support the hypothesis that the guanidino groups of arginine 484 and the imidazole ring of histidine 494 form critical bonds with the invariant bases (G · C)(C · G)8. Changing any of those three components completely destroys binding of repressor.

The amino acid in the lambda repressor which corresponds to R484 of TyrR is also an arginine. It has been shown that when this residue is replaced by lysine, it interacts with the phosphate backbone (24). We have substituted lysine for R484, but the mutant protein has lost all repressor activity (data not shown). This result would favor the hypothesis of a specific interaction between R484 and guanine over a more nonspecific interaction with the phosphate backbone.

The mutants TA495, NA499, and RA502 show a 5- to 10-fold reduction in their ability to repress *tyrP* when the *tyrR* alleles (wild type and mutant) are present on plasmid pMU6218. It is possible that greater fold reduction may be seen at lower levels of TyrR protein, but these strains have not yet been constructed. Since we have previously shown that T495 can be replaced by serine without any reduction of repressor activity (40), it seems likely that the hydroxyl of threonine is involved in the interaction with DNA. Other amino acids whose side chains appear to be vital for the binding of the TyrR protein to DNA such as I497, K500, and L501 are postulated to make vital contributions to the hydrophobic brace which positions the two helices of the HTH.

On the basis of work presented in this report, we propose that in binding to the strong box of *tyrP*, TyrR protein, as a dimer, forms essential and multiple hydrogen bonds between

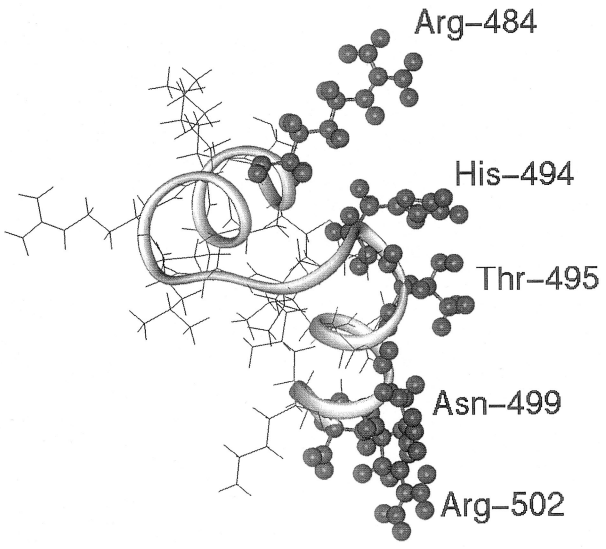

FIG. 5. Hypothetical model of the TyrR HTH motif, showing that the side chains of R484, H494, T495, N499, and R502 protrude on the same face.

R484 of helix 1, H494 of helix 2, and (G · C)(C · G)8 in both arms of the TyrR box. The distance between these G · C and C · G base pairs (14 bp) is also critical, as it has been frequently demonstrated that increasing or decreasing this distance destroys binding (5, 10). Helix 2 of the HTH is predicted to lie along the major groove of the DNA so that the more distal amino acids T495 and N499 can interact with bases closer to the center of the palindrome.

ACKNOWLEDGMENTS

We thank Frank Gibson for helpful discussion and computer modeling of the HTH motif of TyrR. We are grateful to Thu Betteridge for the construction of some of the mutant *tyrR* alleles and to Jing Hong An and Yan Jiang for technical assistance. We are also grateful to Helen Camakaris and Judyta Praszkier for their comments on the manuscript.

This work was supported by a grant from the Australian Research Council. J. S. Hwang is the recipient of an Overseas Postgraduate Research Scholarship from the Australian Government and a Malaysian Alumni Melbourne University Postgraduate Scholarship.

REFERENCES

1. **Andrews, A. E.** 1988. Ph.D. The University of Melbourne, Parkville, Victoria, Australia.
2. **Andrews, A. E., B. Dickson, B. Lawley, C. Cobbett, and A. J. Pittard.** 1991. Importance of the position of TyrR boxes for repression and activation of the *tyrP* and *aroF* genes in *Escherichia coli*. J. Bacteriol. **173:**5079–5085.
3. **Andrews, A. E., B. Lawley, and A. J. Pittard.** 1991. Mutational analysis of repression and activation of the *tyrP* gene in *Escherichia coli*. J. Bacteriol. **173:**5068–5078.
4. **Bailey, M., P. Hagmar, D. P. Millar, B. E. Davidson, G. Tong, J. Haralambidis, and W. H. Sawyer.** 1995. Interaction between the *Escherichia coli* regulatory protein TyrR and DNA: a fluorescence footprinting study. Biochemistry **34:**15802–15812.
5. **Baseggio, N., W. D. Davies, and B. E. Davidson.** 1990. Identification of the promoter, operator and 5′ and 3′ ends of the mRNA of the *Escherichia coli* K-12 gene *aroG*. J. Bacteriol. **172:**2547–2557.
6. **Baumeister, R., V. Helbl, and W. Hillan.** 1992. Contacts between Tet repressor and *tet* operator revealed by new recognition specificities of single amino acid replacement mutants. J. Mol. Biol. **226:**1257–1270.
7. **Brennan, R. G., and B. W. Matthews.** 1989. The helix-turn-helix DNA binding motif. J. Biol. Chem. **264:**1903–1906.
8. **Bushman, F. D., J. E. Anderson, S. C. Harrison, and M. Ptashne.** 1985. Ethylation interference and X-ray crystallography identify similar interactions between 434 repressor and operator. Nature **316:**651–653.
9. **Chye, M.-L., and J. Pittard.** 1987. Transcription control of the *aroP* gene in *Escherichia coli* K-12: analysis of operator mutants. J. Bacteriol. **169:**386–393.
10. **Cobbett, C. S., and M. L. Delbridge.** 1987. Regulatory mutants of the *aroF-tyrA* operon of *Escherichia coli* K-12. J. Bacteriol. **169:**2500–2506.
11. **Cui, J., and R. L. Somerville.** 1993. The TyrR protein of *Escherichia coli*, analysis by limited proteolysis of domain structure and ligand-mediated conformational changes. J. Biol. Chem. **268:**5040–5047.
12. **Garner, C. C., and K. M. Herrmann.** 1985. Operator mutations of the *Escherichia coli aroF* gene. J. Biol. Chem. **260:**3820–3825.
13. **Harrisson, S. C., and A. K. Aggarwal.** 1990. DNA recognition by protein with the helix-turn-helix motif. Annu. Rev. Biochem. **59:**933–969.
14. **Heuer, C., and W. Hillen.** 1988. Tet repressor-*tet* operator contacts probed by operator DNA-modification interference studies. J. Mol. Biol. **202:**407–417.
15. **Kasian, P. A., B. E. Davidson, and J. Pittard.** 1986. Molecular analysis of the promoter operator region of the *Escherichia coli* K-12 *tyrP* gene. J. Bacteriol. **167:**556–561.
16. **Kieny, M. P., R. Lathe, and J. P. Lecocq.** 1983. New versatile cloning and sequencing vectors based on bacteriophage M13. Gene **26:**91–99.
17. **Lawley, B., N. Fujita, A. Ishihama, and A. J. Pittard.** 1994. The TyrR protein of *Escherichia coli* is a class I transcription activator. J. Bacteriol. **177:**238–241.
18. **Majumdar, A., and S. Adhya.** 1989. Effect of ethylation of operator-phosphates on Gal repressor binding. DNA contortion by repressor. J. Mol. Biol. **208:**217–223.
19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
20. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101:**20–78.
21. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Mondragon, A., and S. C. Harrison.** 1991. The phage 434 Cro/OR1 complex at 2.5 Å resolution. J. Mol. Biol. **219:**321–334.
23. **Monod, J., G. Cohen-Bazire, and M. Cohen.** 1951. Sur la biosynthesis de la β-galactosidase (lactase) chez *Escherichia coli*. La specificite de l'induction. Biochim. Biophys. Acta **7:**585–599.
24. **Nelson, H. C., and R. T. Sauer.** 1985. Lambda repressor mutations that increase the affinity and specificity of operator binding. Cell **42:**549–558.
25. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. Gene **26:**101–106.
26. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095.
27. **Pittard, A. J., and B. E. Davidson.** 1991. TyrR protein of *Escherichia coli* and its role as repressor and activator. Mol. Microbiol. **5:**1585–1592.
28. **Praszkier, J., I. Wilson, and A. J. Pittard.** 1992. Mutations affecting translational coupling between the *rep* genes of an IncB miniplasmid. J. Bacteriol. **174:**2376–2383.
29. **Pu, W. T., and K. Struhl.** 1992. Uracil interference, a rapid and general method for defining protein-DNA interactions involving the 5-methyl group of thymines: the GCN4-DNA complex. Nucleic Acids Res. **20:**771–775.
30. **Sarsero, J. P., P. J. Wookey, and A. J. Pittard.** 1991. Regulation of expression of the *Escherichia coli* K-12 *mtr* gene by TyrR protein and Trp repressor. J. Bacteriol. **173:**4133–4143.
31. **Sauer, R. T., R. R. Yocum, R. F. Doolittle, M. Lewis, and C. O. Pabo.** 1982. Homology among DNA-binding proteins suggests use of a conserved supersecondary structure. Nature **298:**447–451.
32. **Schultz, S. C., G. C. Smelds, and T. A. Steitz.** 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°C. Science **253:**1001–1007.
33. **Schumacher, M. A., K. Y. Choi, H. Zalkin, and R. G. Brennan.** 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by a helices. Science **266:**763–770.
34. **Singer, B.** 1976. All oxygen in nucleic acids react with carcinogenic ethylating agents. Nature **264:**333–339.
35. **Vandeyar, M. A., M. P. Weiner, C. J. Hutton, and C. A. Batt.** 1988. A simple and rapid method for the selection of oligodeoxynucleotide-directed mutants. Gene **65:**129–133.
36. **Weber, I. T., and T. A. Steitz.** 1987. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. J. Mol. Biol. **198:**311–326.
37. **Whipp, M. J., and A. J. Pittard.** 1977. Regulation of aromatic amino acid transport systems in *Escherichia coli*. J. Bacteriol. **132:**453–461.
38. **Wilson, T. J., P. Maroudas, G. J. Howlett, and B. E. Davidson.** 1994. Ligand-induced self-association of the *Escherichia coli* regulatory protein TyrR. J. Mol. Biol. **238:**309–318.
39. **Yang, J., H. Camakaris, and A. J. Pittard.** 1996. Further genetic analysis of the activation function of the TyrR regulatory protein of *Escherichia coli*. J. Bacteriol. **178:**1120–1125.
40. **Yang, J., S. Ganesan, J. Sarsero, and A. J. Pittard.** 1993. A genetic analysis of various functions of the TyrR protein of *Escherichia coli*. J. Bacteriol. **175:**1767–1776.
41. **Yang, J., and A. J. Pittard.** 1987. Molecular analysis of the regulatory region of the *Escherichia coli* K-12 *tyrB* gene. J. Bacteriol. **169:**4710–4715.