

## *In vivo* and *in vitro* Studies of TrpR-DNA Interactions

Jie Yang†, Angelo Gunasekera†, Teresa A. Lavoie, Lihua Jin  
Dale E. A. Lewis and Jannette Carey\*

Chemistry Department  
Princeton University  
Princeton, NJ 08544-1009  
(609) 258-1631, USA

The binding of tryptophan repressor (TrpR) to its operators was examined quantitatively using *in vitro* and *in vivo* methods. DNA sequence requirements for 1:1 and tandem 2:1 (TrpR:DNA) binding in various sequence contexts were studied. The results indicate that the optimal half-site sequence for recognition by one helix-turn-helix motif of one TrpR dimer is  $\frac{5}{3}$  CNACT  $\frac{3}{3}$ , consistent with contacts observed by X-ray diffraction analysis of cocrystalline 1:1 and 2:1 complexes. Half-sites can be paired to form a palindrome either by direct abutment, forming the nucleation site for a tandem 2:1 complex, or with an 8-base-pair spacer, forming a 1:1 target. Dimethylsulfate (DMS) methylation-protection footprinting *in vitro* of 1:1 and 2:1 complexes formed sequentially on the two unequal half-site pairs of the *trp*EDCBA operator from *Serratia marcescens* indicated an obligate hierarchy of site occupancy, with one half-site pair serving as the nucleation site for tandem binding. DMS footprinting of *Escherichia coli* operators *in vivo* showed that, over a wide range of intracellular TrpR concentration, the *trp*EDCBA operator is occupied by three repressor dimers, *aroH* is occupied by two dimers, and the 1:1 binding mode is used on the *trpR* operator. The coexistence of these distinct occupancy states implies that changes in protein concentration affect only the fractional occupancy of each operator rather than the binding mode, which is determined by the number of half-site sequences present in the operator region. Cooperativity of tandem complex formation measured by gel retardation using a symmetrized synthetic operator containing identical, optimal sites spaced as in natural operators was found to be modest, implying a maximum coupling free energy of  $\sim -2$  kcal/mol. On other sequences the apparent degree of cooperativity, as well as the apparent affinity, varied with sequence and sequence context in a manner consistent with the structural models and which suggests compensation between affinity and cooperativity as a mechanism that allows tolerance of operator sequence variation.

© 1996 Academic Press Limited

**Keywords:** affinity; cooperativity; compensation; stoichiometry; footprinting

\*Corresponding author

### Introduction

Much effort has been devoted to elucidating the DNA sequence requirements for TrpR binding over the years. Functional analysis of the sequence by classical (Bennett & Yanofsky, 1978) and molecular (Haydock *et al.*, 1983; Bass *et al.*, 1987, 1988; Czernik *et al.*, 1994) genetics and by biochemical analysis (Gunsalus & Yanofsky, 1980; Kumamoto *et al.*, 1987; Carey, 1989; Staacke *et al.*, 1990; Marmorstein *et al.*,

1991; Carey *et al.*, 1991; Haran *et al.*, 1992; Sutton *et al.*, 1993; LeTilly & Royer, 1993; Liu & Matthews, 1993; Joachimiak *et al.*, 1994) consistently pointed to the self-symmetric sequence 5' CTAG 3' as being an essential part of the recognition motif for TrpR. Yet two X-ray cocrystal structures of TrpR-DNA complexes (Otwinowski *et al.*, 1988; Lawson & Carey, 1993) are in agreement in showing that the helix-turn-helix motif of the repressor contacts only half of the bases of the CTAG sequence, plus additional bases flanking this sequence. Biochemical (Marmorstein *et al.*, 1991) and genetic (Bass *et al.*, 1987, 1988) evidence has not always been in agreement about the importance of bases flanking

† These authors contributed equally, and both should be regarded as first authors.

Abbreviations used: DMS, dimethylsulfate.

the CTAG sequence. Thus, our present clear structural definition of the binding site is not matched by a functional definition of comparable precision.

The contacts observed in both X-ray cocrystal structures to the CTAG sequence and its flanking bases suggest that the half-site for binding of one helix-turn-helix motif of one repressor dimer should be redefined as 5' GNACT 3'. (Underlined bases highlight the relationship of the redefined half-site to the 5' CTAG 3' sequence; only one strand of the duplex half-site is shown, although not all contacts are made to this strand.) Interestingly, theoretical analysis of DNA sequence information content identified these nucleotides as the distinguishing feature of TrpR operators before the first X-ray cocrystal structure was determined (Schneider *et al.*, 1986). Symmetric reflection of this half-site about itself (·) with a spacer of eight base-pairs (5' GNACTNNNN·NNNNAGTNC 3') creates a half-site pair that can be contacted by the two helix-turn-helix motifs of one repressor dimer to form a 1:1 complex; in this configuration no contiguous CTAG sequence is present. Indeed, the DNA sequence used for the original cocrystal structure of the 1:1 complex (Otwinowski *et al.*, 1988) fits this configuration, even though that DNA also contained two contiguous CTAG sequences (indicated in boldface): 5' **G**N**A**C**T**A**G**N**N**·**N**N**C**T**A**G**T**N**C** 3'. Redefinition of the half-site thus clarifies the fact that in the original cocrystallized DNA the inner two bases of each CTAG sequence are not required to form the observed 1:1 complex, and explains why they were not contacted. The structurally defined half-site is a more precise definition of the recognition sequence because it represents only contacts to a single helix-turn-helix motif, and it implies directionality of the bound repressor in a way that the symmetric CTAG sequence cannot.

On the other hand, if two symmetry-reflected half-sites are directly abutted (5' GNACT·AGTNC 3'), this half-site pair can be shared by two helix-turn-helix motifs from two separate repressor dimers bound in a tandem 2:1 complex (Kumamoto *et al.*, 1987; Otwinowski *et al.*, 1988; Carey *et al.*, 1991; Haran *et al.*, 1992). Tandem binding at a central, shared site engages only one of the two helix-turn-helix motifs of each repressor dimer, while the distal motif of each dimer faces the DNA sequence in the next major groove opening approximately one turn (actually eight base-pairs) away. Tandem binding should thus be optimal when the distal DNA sequences also match the structurally defined half-site, and the ideal site for tandem binding of two TrpR dimers is expected to be 5' GNACTNNNGNACTAGTNC·NNNAGTNC 3'. By definition, this sequence contains two perfect 1:1 binding sites (one underlined and one overlined), comprised of a nested set of two half-site pairs. Relative populations of 1:1 and 2:1 occupancy states on such an operator should depend on intrinsic affinity of the

1:1 sites and on cooperativity in forming the 2:1 complex.

On longer but imperfect tandem sequences, such as are found in natural operators, the choice of binding mode should depend on sequence and sequence context through their effects on affinity, stoichiometry, and cooperativity (Liu & Matthews, 1993). However, quantitative data that allow correlation of sequence with binding mode have not been available; thus, it has not been possible to determine by inspection the binding mode used on any given sequence. The tandem binding observed in one of the X-ray cocrystal structures suggests that the apparent importance of the CTAG sequence in functional assays could be related to the fact that this sequence forms the core of a tandem binding site for two repressor dimers, although in many of the functional studies the effect of flanking sequences was not studied systematically, and the binding mode was generally not known.

Natural *trp* operators deviate substantially from each other and contain different numbers of half-site pairs with varying match to the consensus (Figure 1). *In vitro* data indicate that under some conditions multiple TrpR dimers may be bound to some of these operators (Kumamoto *et al.*, 1987; Liu & Matthews, 1993). In particular, Kumamoto *et al.* (1987) proposed that three TrpR dimers bind to the *trp*EDCBA operator, two to *aroH*, and one to *trpR*. The significance of tandem binding for regulation has not been clear, partly because interpretation of mutational results is hampered by the fact that the binding mode *in vivo* is generally not known. In this paper we examine the functional definition of a binding site for TrpR by evaluating binding *in vitro* and *in vivo* on a range of DNAs that differ in sequence and sequence context. The data allow us to address the roles of affinity, stoichiometry, and cooperativity for *in vitro* DNA binding and *in vivo* regulation of gene expression by TrpR.

## Results

### Cooperativity

On the short DNA used for cocrystallization of the 2:1 complex (Figure 2(a)), tandem binding occurs without formation of the 1:1 complex as an intermediate, indicating operationally infinite positive apparent cooperativity (Carey *et al.*, 1991). Presumably, the absence of distal half-sites so destabilizes 1:1 complexes on this DNA that they are not detected by gel retardation (Carey *et al.*, 1991, 1993). This artifactual source of cooperativity might be eliminated by extending the DNA sequence to include complete distal half-sites. In this case, the longer DNA would be expected to display increased affinity in forming the 1:1 complex and decreased apparent cooperativity in forming the 2:1 complex. To test this prediction, an oligonucleotide (Figure 2(b)) containing four optimized, identical half-sites was synthesized. Gel retardation studies

showed sequential formation of two kinds of complexes on this DNA (Figure 2(c)). Detection of the 1:1 complex varied depending on the conditions of the gel, increasing with lower temperature or higher total acrylamide concentration, indicating that these complexes dissociate during running of the gel, as suggested previously (Cann, 1989; Liu & Matthews, 1993). Densitometric scanning showed that the first complex accounts for only ~20% of the total amount of DNA at the point of its maximum accumulation (~1 nM in Figure 2(d)). In the absence of cooperativity or dissociation of the complexes during running of the gel, the 1:1 species is expected to accumulate to 50% of the total DNA (Senear & Brenowitz, 1991). Although dissociation in the gel must contribute to the under-accumulation of 1:1 complex (Cann, 1989), the existence of positive cooperativity in this system is also indicated by the amounts of free DNA and 2:1 complex, which together account for considerably more than half the total DNA.

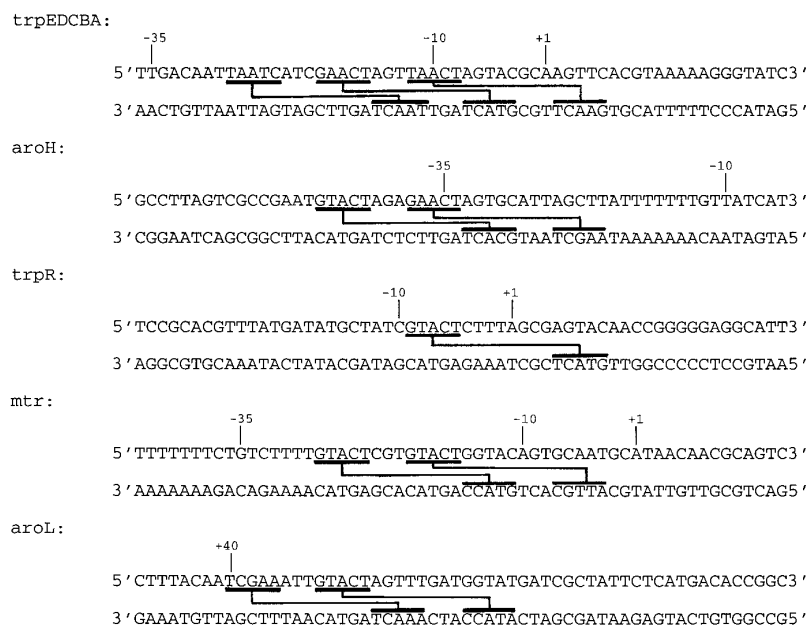
The fractions of each species as a function of TrpR concentration were fit simultaneously to equations describing the cooperativity (Senear & Brenowitz, 1991; see Materials and Methods). Typical curves fitted to data from one such experiment are shown in Figure 2(d)). The average value of  $k_1 = k_2 = 2.2 \times 10^8 \text{ M}^{-1}$  from four independent experiments. The average cooperativity ( $k_{12} = 40 \pm 15$ , from four independent experiments) is equivalent to a coupling free energy of about  $-2 \text{ kcal/mol}$ . This value of the coupling free energy corresponds to a narrowing of the free ligand concentration interval (Weber, 1993) to about 1.65 log units (10% bound to 90% bound over a 45-fold increase in free ligand concentration) compared to 1.91 log units (81-fold) for the non-cooperative case. This narrowing of the interval is probably too small to be detected with confidence directly from the steepness of the

2:1 titration, but is readily revealed by quantifying accumulation of the 1:1 complex (Senear & Brenowitz, 1991).

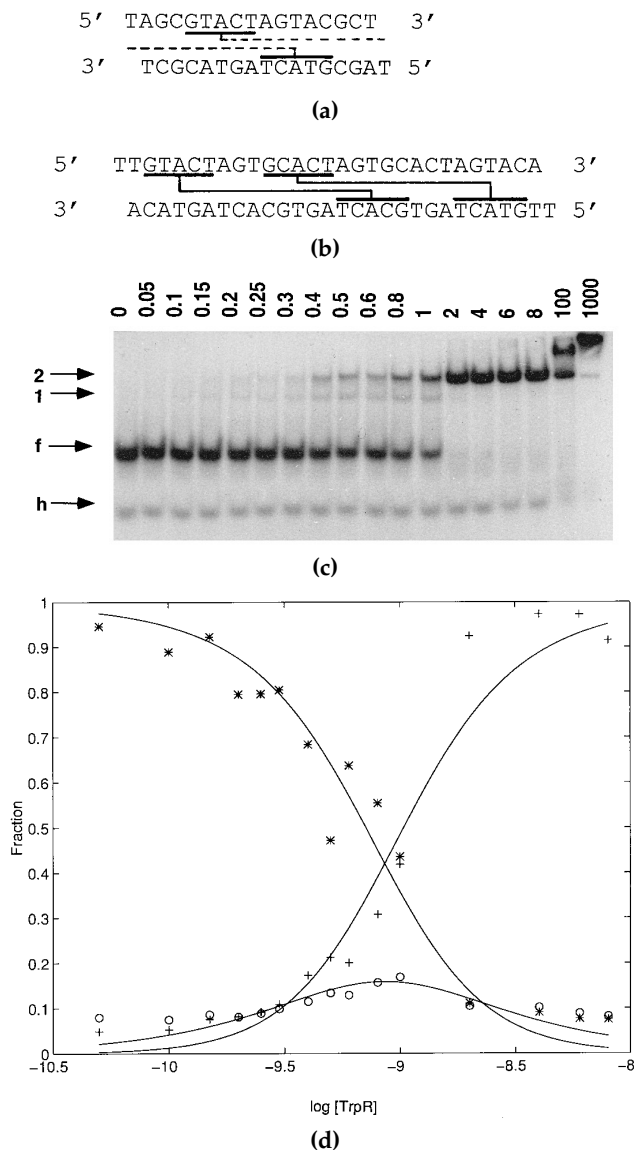
The measured cooperativity would be smaller (coupling free energy less negative) if dissociation of the 1:1 complexes did not occur in the gel. Thus, this value of the coupling free energy almost certainly overestimates the cooperativity to an unknown degree, since it is derived from quantitation of the 1:1 band which is likely to be underrepresented on the gels. As described theoretically for the TrpR case by Cann (1989), the dissociation of these complexes during the gel running time can be significant, due to the slow rate of reassociation at the low concentrations used for gel retardation. We were able to provide indirect support for this expectation using the cocrystallized tandem DNA (Figure 2(a)) by measuring the dissociation rate ( $k_d \sim 10^{-4} \text{ s}^{-1}$ ; data not shown) and apparent affinity ( $K_a \sim 10^{10} \text{ M}$ ; Carey *et al.*, 1991) of TrpR complexes in our gels. The association rate calculated from these measurements ( $k_a \sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$ ) is much slower than that observed experimentally for longer DNAs under similar conditions ( $k_a \sim 10^9 \text{ M}^{-1} \text{ s}^{-1}$ ; Chou & Matthews, 1989; Hurlburt & Yanofsky, 1992). In our hands, the 1:1 complexes are more labile than the 2:1 complexes in all conditions; thus, it is likely that our measurements underestimate the buildup of the 1:1 complexes, and thus overestimate the cooperativity. Nevertheless, the measurements confirm the existence of positive cooperativity in this system, and place an upper limit on its contribution to complex stability.

### DNA sequence requirements for TrpR binding

The principal goal of this part of the work was to test quantitatively the functional importance of bases implicated in the structurally redefined TrpR



**Figure 1.** Sequences of the operator regions of five *E. coli* operons regulated by TrpR. Numbers above sequences indicate positions within each promoter region, defined with respect to +1, the first transcribed residue. One strand of each potential half-site (consensus sequence 5' GNACT 3') is underlined or overlined, and half-site pairs with appropriate spacing to be occupied by individual repressor dimers are connected to show their symmetry. Some potential half-site pairs are marked even if only one half-site matched the consensus sequence; occupancy of these pairs by TrpR is discussed in the text (see Discussion).



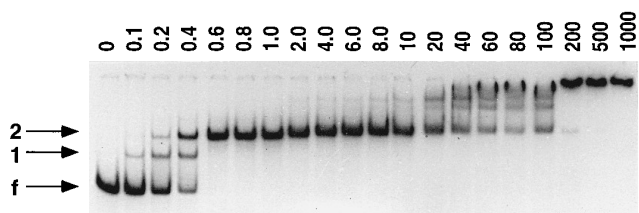
**Figure 2.** Cooperativity of TrpR binding to an optimized DNA target, analyzed by gel retardation titration. (a) The sequence of the short DNA used for cocrystallization of the tandem complex. One strand of each half-site in the central tandem site is underlined or overlined. The connection to the missing half-site is shown as a dotted line to emphasize the fact that this oligonucleotide lacks the distal half-sites. (b) The sequence of the 30-bp idealized, synthetic operator. One strand of each half-site is underlined or overlined, and half-site pairs appropriately spaced for binding by individual repressor dimers are connected to show their symmetry. (c) An autoradiogram of a typical retardation gel is shown. The concentration of TrpR (nM) in each reaction mixture is indicated above the corresponding lane; DNA concentration was  $< 0.01$  nM. Arrows indicate the positions of DNA hairpin (h), free DNA duplex (f), and TrpR-DNA complexes 1 and 2. (d) Plot of fractional species distribution as a function of log (TrpR concentration). Species indicated are free duplex DNA (stars), complex 1 (circles), and complex 2 (crosses). Quantitation of each species was by phosphorimage scanning of the gel in (c), as described in Materials and Methods. Continuous lines represent the best simultaneous fit to equations (1) to (3) of Materials and Methods. The value of the cooperativity parameter determined from the data shown is  $k_{12} = 28$ .

half-site. As we show below, a systematic test of all possible DNA sequence variants is likely to be unnecessary to meet this aim, given the large body of prior data on TrpR binding to many DNAs. Instead, to minimize DNA synthesis we first analyzed binding to existing sequences from various sources to test some of the quantitative effects of sequence variation.

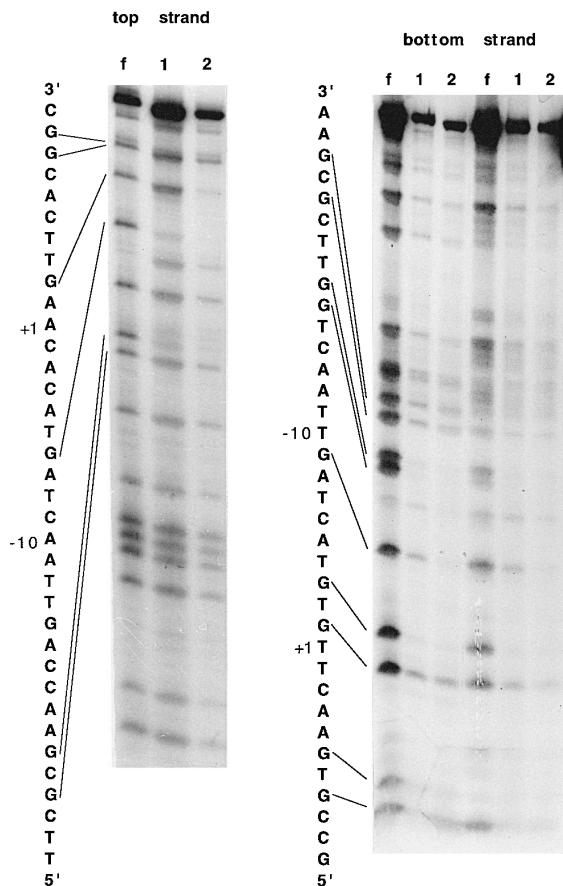
#### Formation of 1:1 and 2:1 complexes on a natural operator in vitro

Liu & Matthews (1993) have shown that two types of complexes can be resolved by gel retardation on a synthetic DNA oligonucleotide representing the natural sequence of the *trp*EDCBA operator from *Escherichia coli*. We had made similar observations on the *Serratia marcescens trp* EDCBA operator (Figure 3(a)). Both operators have two imperfect half-site pairs (Figures 1 and 3(c)), each with reasonable match to the structurally defined consensus, and both operators form first 1:1 and then 2:1 complexes as the TrpR concentration is raised in a titration experiment (Liu & Matthews, 1993, and Figure 3(a)), consistent with the number and spacing of half-sites (Figures 1 and 3(c)). At very high concentrations of TrpR, higher-order complexes on the *S. marcescens trp* EDCBA operator are sometimes resolved in our gels. Their mobilities appear to be consistent with 3:1 and 4:1 complexes (Figure 3(a)), similar to the 3:1 complexes reported by Liu & Matthews (1993) for several superrepressor mutants. Regardless of their stoichiometries, higher-order complexes formed on this operator must use one or more distal flanking half-sites (not shown) with very poor match to the consensus sequence.

The two imperfect half-site pairs for 1:1 binding on the *S. marcescens trp*EDCBA operator differ in sequence both from each other and from the structurally defined consensus sequence (see Figure 3(c)). The left-and-middle pair (hereafter called the left pair) has the sequence 5'GNA-CCNNNNNNNAGTNC3' spanning residues -20 to -3 while the middle-and-right pair (right pair) spanning residues -12 to +6 has the sequence 5'TNACTNNNNNNNAGTNC3' (mismatches are underlined; corresponding differences also occur in the opposite strand, which is not shown). Formation of 1:1 complexes on this DNA may occur either by preferential occupancy of one of the two pairs of half-sites, or by binding to the two half-site pairs in statistical distribution according to their affinities. Thus, determining the occupancy of these pairs of sites will give information about sequence preferences for binding. To determine the order in which half-site pairs of this operator become occupied by TrpR in the course of a titration, dimethyl sulfate (DMS) chemical modification-protection was carried out on a series of samples at increasing TrpR concentrations as described in Materials and Methods. Chemically modified samples were resolved by gel retardation (Figure 3(a)) into free

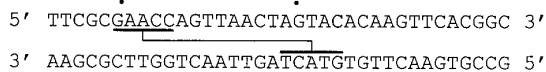


(a)

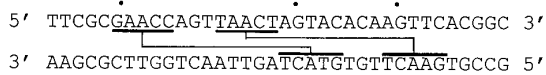


(b)

**Complex 1:**



**Complex 2:**



(c)

**Figure 3.** Sequential formation of 1:1 and 2:1 complexes on the *trpEDCBA* operator of *S. marcescens*, analyzed by gel retardation titration and dimethyl sulfate (DMS) protection. (a) An autoradiogram of a typical retardation gel is shown. A titration experiment similar to that shown

DNA, 1:1, and 2:1 complexes, which were purified from the gels prior to piperidine cleavage and sequencing gel analysis. Sequencing gel autoradiograms (Figure 3(b)) were scanned to determine the protected guanine residues on each strand of the operator in each complex (Figure 3 (c)).

The 1:1 complex resolved by gel retardation shows protection of only the left half-site pair of the operator, and no protections in the right half-site pair. The pattern of protected residues is consistent with previous footprinting results on this DNA at very low TrpR concentrations (Carey, 1989) and with the overall architecture of the 1:1 complex as determined in the X-ray structure (Otwinowski *et al.*, 1988; Yang & Carey, 1995). Thus, these results indicate that TrpR forms a 1:1 complex on this DNA by preferentially using the left half-site pair, while the right half-site pair is largely or completely unoccupied. Both left and right half-site pairs deviate from the consensus sequence, but at different residues. Thus, preferential binding to the left pair could be due either to an improved affinity for that pair, or to a reduced affinity for the right pair. Joachimiak *et al.* (1994) examined the effects of symmetric residue replacements on 1:1 complex formation by short DNAs, and found that DNAs with sequence 5'GNACC3' (identical to the half-site at positions -20 to -16 of the natural operator) bound TrpR almost as well as the perfect consensus site. Because this sequence is not preferred over the consensus sequence, we can conclude that preferential binding to the left half-site pair of the natural operator is more likely to be due to reduced affinity for the right pair. Indeed, Marmorstein *et al.* (1991) found substantially reduced affinity for synthetic symmetric DNAs with the change 5'TNACT3'.

in Figure 2(c) was carried out with end-labeled DNA, but before the reaction mixtures were loaded onto the gel for retardation assay they were probed with DMS. The concentration of TrpR (nM) in each reaction mixture is indicated above the corresponding lane. Arrows indicate the positions of free duplex DNA (f) and TrpR-DNA complexes 1 and 2. (b) An autoradiogram of a typical 6% polyacrylamide sequencing gel for protection analysis is shown. Complexes 1 and 2 and free DNA were eluted from the corresponding bands of the retardation gel, and the purified DNAs were subjected to piperidine treatment to cleave at methylated purines. Cleavage products terminating at purine residues are displayed on the sequencing gel, with the corresponding G residues indicated on the operator-region sequence at the left. Top strand (left panel) is labeled at the *EcoRI* site, and G methylations are shown; bottom strand (right panel) is labeled at the *BamHI* site, and the first group of three lanes represents G methylations only, while the second group of three lanes represents both A and G methylations. (c) Summary of methylation protection results on complexes 1 and 2, displayed on the operator-region DNA sequence. A filled circle above or below indicates a protected guanine on the top or bottom strand, respectively. No protections were observed outside the sequence region shown. Occupancy of half-sites by TrpR is indicated by underlining or overlining as in Figure 1.

The 2:1 complex shows protections on both half-site pairs, and residues protected in the middle half-site in the 1:1 complex show increased protection in the 2:1 complex. This pattern is consistent with previous footprinting results on a DNA sequence that supports 2:1 binding (Staacke *et al.*, 1990; Yang & Carey, 1995) and with the overall architecture of the 2:1 complex from X-ray analysis (Lawson & Carey, 1993). The footprinting results demonstrate that 1:1 binding by TrpR is not statistically distributed over the two unequal half-site pairs of the *trpEDCBA* operator from *S. marcescens*, but rather occurs principally or exclusively on the left half-site pair. Thus, single and tandem binding of TrpR occur sequentially and with an obligate hierarchy of site occupancy.

### Effects of distal and flanking sequence

Several restriction sites in the polylinker region of pBend3, a plasmid used for bending studies with cloned operators, contain the CTAG sequence that forms the core of a tandem TrpR binding target, but with different bases immediately flanking it and in the potential distal half-sites. Cloning of the tandem TrpR target (Figure 2(a) and Materials and Methods) into the *SalI* site of the polylinker permits excision of a 264-bp restriction fragment containing a total of six CTAG sequences in various sequence contexts (Figure 4(a)). Inspection of the surrounding sequence reveals that each CTAG has no distal half-sites with reasonable match to the structurally defined target in a spacing appropriate for forming a 1:1 complex; thus if a CTAG sequence is used, it should nucleate tandem binding. Gel retardation titrations of this fragment with TrpR revealed the formation of at least three discrete complexes (Figure 4(b)). To probe the sites occupied by TrpR in each complex, samples from such titrations were methylated with DMS, resolved by gel retardation, and analyzed by sequencing gel. Typical results are presented in Figure 4(c) and summarized in Figure 4(a). Protected sites were determined by visual inspection using the methylation pattern of the isolated free DNA as a reference for the unprotected state. The first complex, formed at lowest TrpR concentration with fastest mobility (band 1 in Figure 4(b) and lane 1 in Figure 4(c)) has only one protected region centered on the cloned TrpR operator sequence (central unmarked bracket on Figure 4(c)). The second and third complexes (bands 2 and 3 in Figure 4(b) and lanes 2 and 3 in Figure 4(c)) both have two additional protected regions at *SpeI* restriction sites in the polylinker region. In addition to their different gel mobilities, complexes 2 and 3 differ from each other in the extent of protection at both *SpeI* sites. Taken together, these two features suggest that complex 2 has a substoichiometric statistical distribution of TrpR on the two identical *SpeI* sites, while complex 3 has both these sites occupied.

The sequence immediately adjacent to the *SpeI* sites of pBend3 provides a nearly perfect match to

the redefined tandem target, while the surrounding sequences provide little or no match to the consensus sites for contact by the distal helix-turn-helix motifs of repressor dimers tandemly bound at the central shared site. The same is true for the cloned tandem target, reminiscent of the situation in the short DNA target used for cocrystallization of the 2:1 complex (Carey *et al.*, 1991, 1993). That oligonucleotide duplex supports tandem binding despite the complete absence of contacts to the distal helix-turn-helix motif of each repressor. Other restriction sites on the pBend3 fragment that also have a CTAG core sequence but which are flanked by other bases (Figure 4(a)) include *XbaI* and *NheI*. The footprinting results indicate that TrpR does not occupy these sequences at concentrations up to ~150 nM. Taken together, these results affirm the importance of bases immediately flanking the CTAG core while corroborating the inference, based on binding to the short tandem target (Figure 2(a); Carey *et al.*, 1991, 1993), that tandem complex formation is compatible with a wide range of distal sequences.

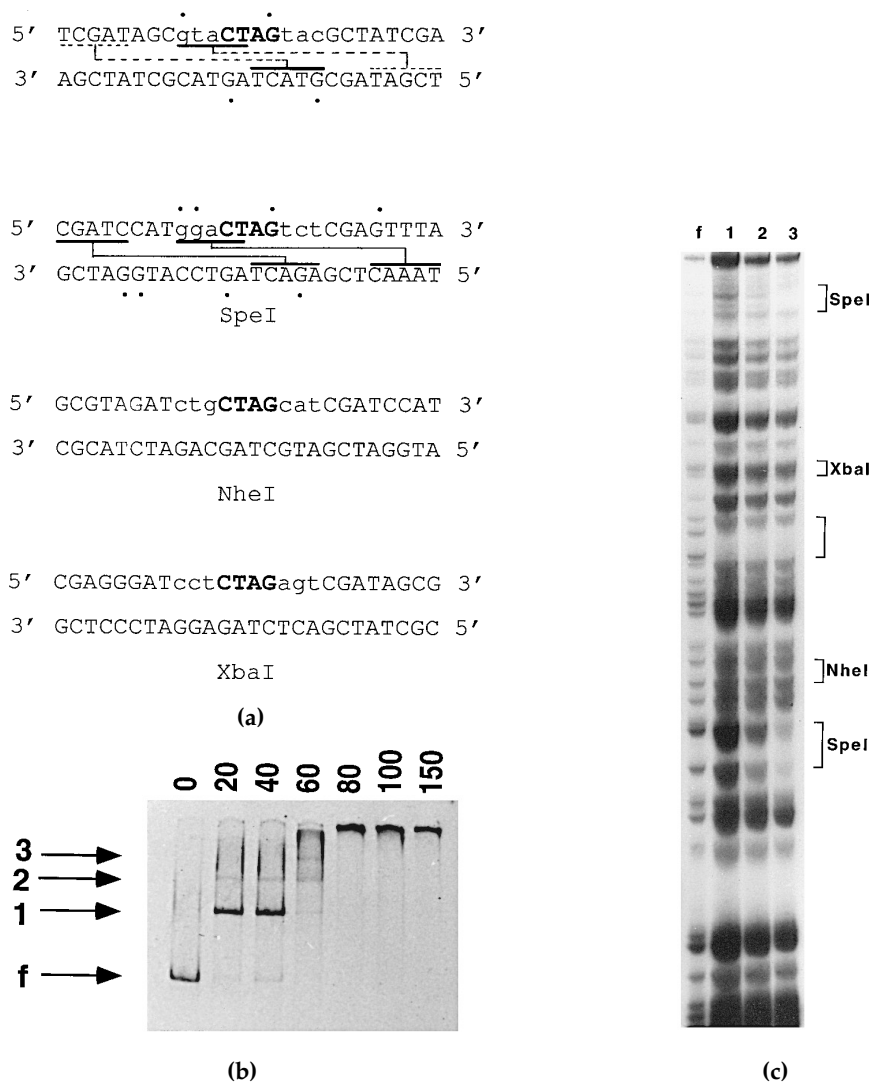
The results of TrpR binding to the cloned tandem target DNA and to the *SpeI* sites suggested the possibility that a minimal target for tandem TrpR binding could be as short as 10 bp containing only two directly abutted ideal half-sites of 5 bp each (5' GNACT·AGTNC 3'). The extreme cooperativity seen on the cocrystallized DNA might permit visualization of a tandem complex even on this much shorter DNA. This 10-bp DNA oligonucleotide was examined for TrpR binding by gel retardation titration at very high DNA concentration (to favor the double-stranded form). A protein-DNA complex was indeed observed at very high TrpR concentration (10 mM; not shown). Although the significance of such a weak complex is not clear, the greatly reduced affinity for the minimal sequence relative to the *SpeI* sites indicates that distal sequences provide some stabilization even when their sequences match only poorly to the consensus.

### *In vivo* studies of TrpR binding to its operators

The sites occupied by TrpR on several of its natural operators were studied by DMS footprinting *in vivo* (Ephrussi *et al.*, 1985; Sasse-Dwight & Gralla, 1991). Growing cells take up DMS spontaneously; DNA methylation proceeds as in an *in vitro* experiment, and methylation-protection has the same interpretation (Wissman & Hillen, 1992; Yang & Carey, 1995). For TrpR, we have shown that the footprint pattern agrees in detail with the overall architecture of the TrpR-DNA complex, and can be used to distinguish between 1:1 and tandem binding (Carey *et al.*, 1991; Lewis & Carey, 1993; Yang & Carey, 1995). Operator binding as a function of intracellular TrpR concentration in the presence of excess L-tryptophan was assessed by comparing protection patterns in strain CY15071 (Paluh &

Yanofsky, 1986), which has the chromosomal *trpR* gene deleted, with strain DH1, which contains a wild-type, autoregulated chromosomal copy of the *trpR* gene, and with strain DH1 transformed with

plasmids pRPG16 (Gunsalus & Yanofsky, 1980), pRLK13 (Kelley & Yanofsky, 1985), or pJPR2 (Paluh & Yanofsky, 1986), which produce additional TrpR at increasing levels (see Figure 5 legend). Although



**Figure 4.** Effect of distal and flanking sequences on binding-site selection, analyzed by gel retardation titration and DMS protection. The tandem TrpR-binding sequence used for cocrystallization (Figure 2(a)), resynthesized to contain flanking *SalI* sites (see Materials and Methods), was cloned into the *SalI* site in the polylinker region of plasmid pBend3, and excised from the resulting clone on a 260 bp *EcoRI-HindIII* restriction fragment bearing several other potential TrpR binding sites. (a) Summary of potential TrpR-binding sequences on the polylinker fragment. Potential binding sites were considered to be any sequences containing the CTAG palindrome (boldface) that forms the core of a tandem TrpR binding site. Six such sites, with different flanking sequences (lower case) are located on the polylinker fragment: the cloned tandem target (top), and two each of restriction sites for the enzymes *SpeI* and *NheI*, and one for *XbaI*. 26 bp of local sequence around each potential target is shown; a bound TrpR dimer binding on either half of the CTAG core sequence would have no contacts outside this length. Filled circles mark protected guanine residues, and underlining and overlining indicate occupancy by TrpR (see c). (b) Gel retardation titration of TrpR binding to the end-labeled polylinker fragment; an autoradiogram of a typical gel is shown. The reaction mixtures were probed with DMS prior to being loaded onto the gel. The concentration of TrpR (nM) in each reaction mixture is indicated above the corresponding lane; DNA concentration was  $\sim 0.5$  nM. Arrows indicate the positions of free duplex DNA (f) and TrpR-DNA complexes 1, 2, and 3. The bands in the last three lanes mark the position of the wells, where DNA sequences of this length can be retained at very high protein concentration. (c) Protection analysis of TrpR binding to sites on the bottom strand of the polylinker fragment labeled at the *HindIII* site; an autoradiogram of a typical sequencing gel is displayed. Free DNA and complexes 1, 2, and 3 were purified from the gel shown in (b) and cleaved at modified G residues using piperidine. Cleavage products were resolved on a 6% acrylamide sequencing gel. Brackets mark locations of potential TrpR target sequences indicated in (a). The central unmarked bracket identifies the cloned tandem target. DMS methylation protection results on both strands are summarized in (a) by filled circles corresponding to protected guanine residues.

the absolute intracellular TrpR concentrations are unknown for most of these strains, we estimate that they cover a range of TrpR concentrations of at least 150-fold, with the lowest non-zero concentration being approximately 180 dimers per cell (estimated concentration, 270 nM), as determined directly in a strain with only the wild-type chromosomal copy of the *trpR* gene (Gunsalus *et al.*, 1986).

Methylated residues on chromosomal *trpEDCBA* and *aroH* operators (Figure 5(a),(b)) were detected as sites of termination of polymerase chain reactions primed from labeled synthetic oligonucleotides hybridized just outside the operator region on each strand. A site of termination marks a site of modification in the opposite strand; thus, protected purine residues on one strand are detected as bands of reduced intensity that terminate at the pyrimidine residue in the opposite strand. To study protection of the *trpR* operator (Figure 5(c)), we constructed a plasmid bearing only the *trpR* promoter and not the *trpR* structural gene; methylated residues on the *trpR* operator were detected directly on a labeled fragment of plasmid DNA as sites of piperidine cleavage. On both plasmid and chromosomal DNA, most protected

residues were guanines although occasional adenine protections were observed, consistent with the ability of DMS to methylate adenine at a rate slower than that of guanine.

The methylation-protection pattern of the chromosomal *aroH* operator is shown in Figure 5(a). The protected positions (indicated by the filled circles and summarized in Figure 5(d)) are obvious by visual inspection and were confirmed by scanning densitometry (data not shown). On the bottom strand (left panel), reduced intensity is observed for bands corresponding to termination at residues -46, -32, and -24, implying protection of the corresponding purine on the top strand. No other protections are evident from analysis of the remaining termination sites on this strand at positions -53, -50, -40, -38, -30, and -12. On the top strand, bands corresponding to purines -43, -42, -35, -29, and -23 give evidence of protection. No positions beyond -46 or -12 are protected. Two of the purines, at positions -23 and -24, are very weakly protected. The details of the protection pattern, including both positions and extents of protection, are independent of strain and thus of intracellular TrpR concentration. On the *trpEDCBA* operator, differences in

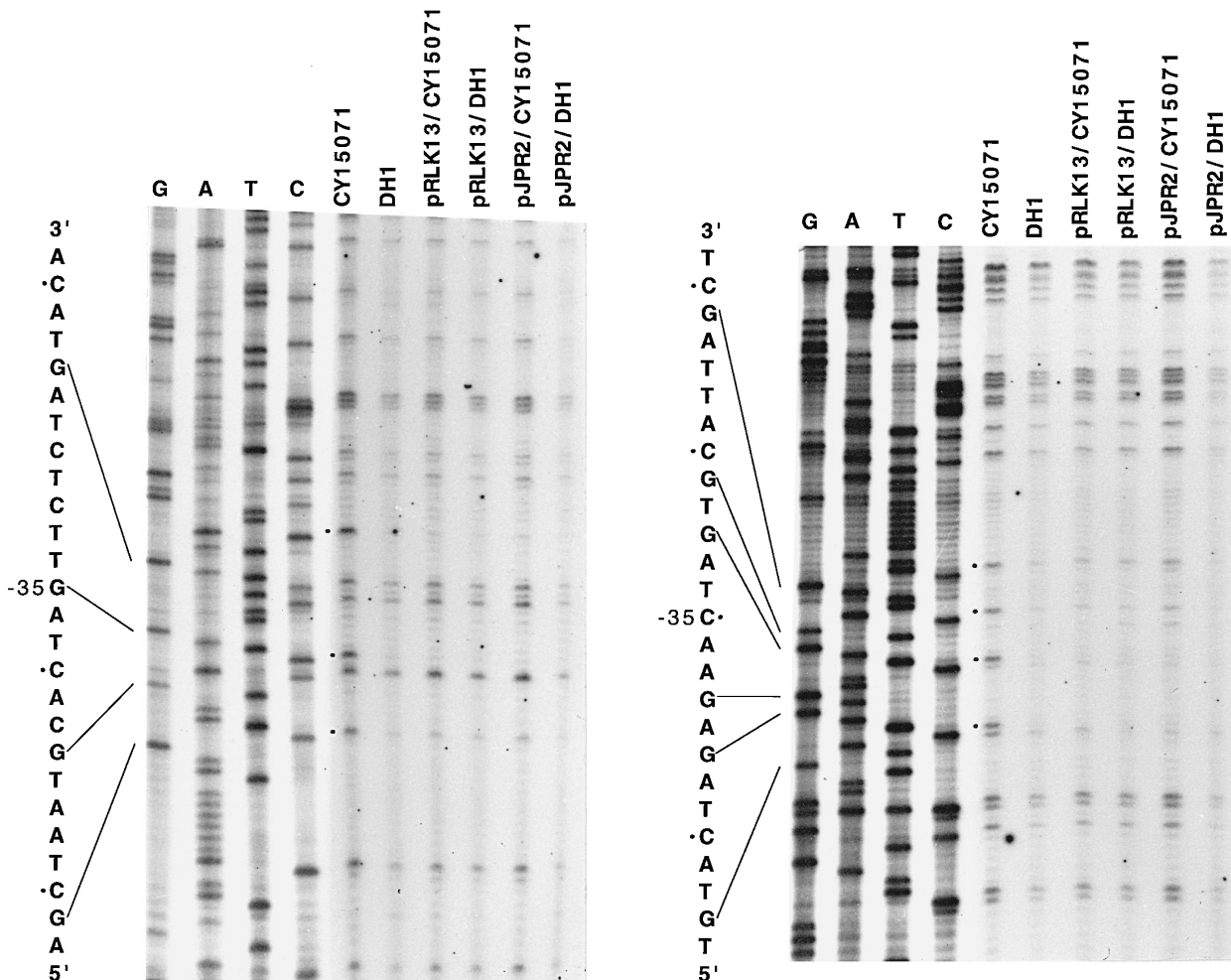


Figure 5(a) (caption on page 10)



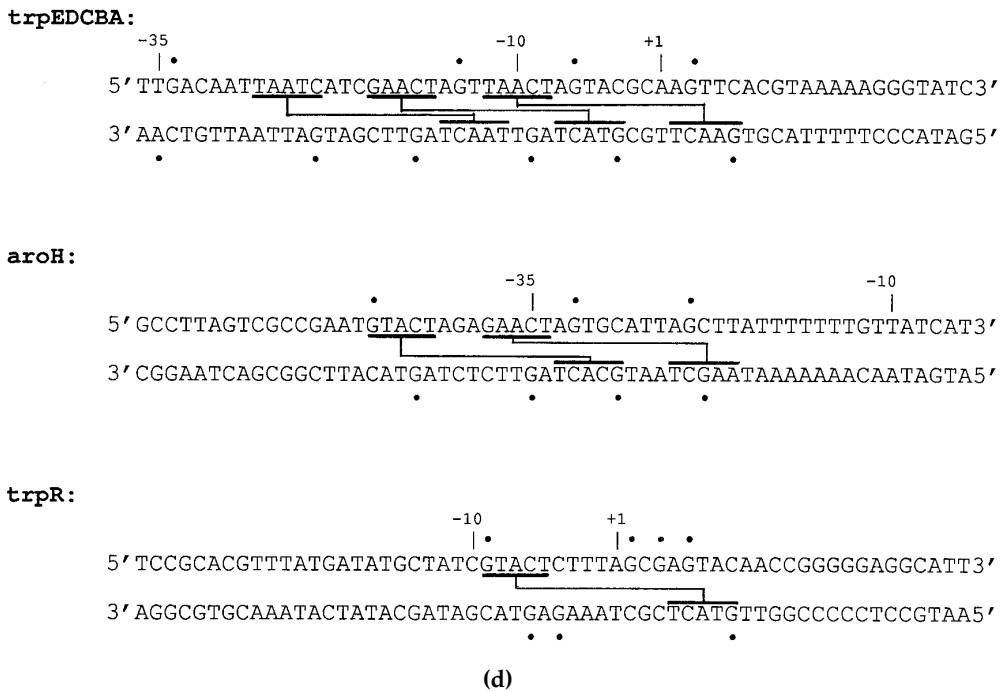
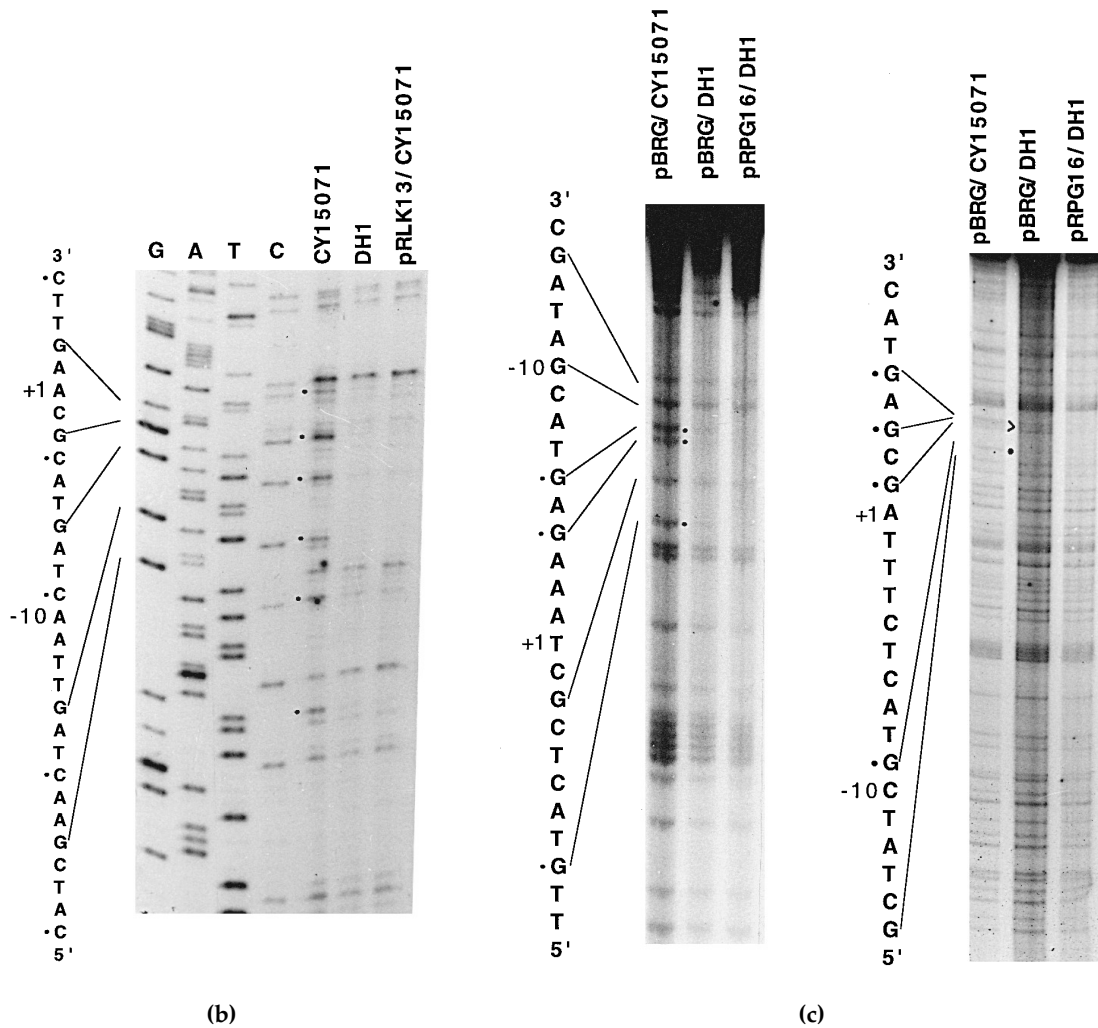


Figure 5 (b)-(d) (legend overleaf)

band intensities could be evaluated by visual inspection (Figure 5(b)) to reveal protection of the following positions on the bottom strand: -35, -24, -17, -9, -3, and +6 (summarized on Figure 5(d)). Protections on the other strand were weaker and not obvious by visual inspection in our experiments, although protections at positions -34, -14, -6, and +3 were observed, consistent with the protected positions on the bottom strand. Positions -32, -2, +3, +8, and +9 were not protected (Figure 5(d)), nor were any protections observed outside of positions -35 and +9. The extent of protection varies at different positions within the operator (e.g. about 50% at position -24, 70% at position -3, and 80% at positions -9 and -17), but neither the sites nor extents of protections vary among the different strains examined. On the *trpR* operator (Figure 5(c)), positions -9, -6, -4, +2, +3, +6, and +9 are protected, while positions -14, -10, and +4, as well as all residues outside of -14 or +9, are not protected. Quantitation indicates a uniform extent of protection of about 40% across the operator, independent of TrpR concentration.

The footprinting data can be superimposed on the binding sites identified in each operator by inspection of the sequences for matches to the structurally defined consensus (Figures 1 and 5(b)). On the *aroH* operator, four half-sites at positions -46 to -42, -38 to -34, -33 to -29, and -25 to -21 are protected. The half-site at positions -25 to -21 has the poorest match to the consensus and contains the residues that are the most weakly protected. On the *trpEDCBA* operator, six half-sites at positions -28 to -24, -20 to -16, -15 to -11, -12 to -8, -7 to -3, and +2 to +6 are protected. Although it appears that base-pairs -11 and -12 are used simultaneously as part of two different half-sites, the X-ray data predict that there is no steric clash between tandemly bound repressors

because of the exact position of the contacts with these base-pairs (Otwiniowski *et al.*, 1988; Lawson & Carey, 1993). Contacts are predicted with T at -11 in the -15 to -11 half-site, and with A at -12 in the -12 to -8 half-site. Two protected residues, at positions -34 and -35, do not fit the pattern expected for TrpR binding (see Discussion). On the *trpR* operator, two half-sites at positions -9 to -5 and +5 to +9 are protected. These results suggest that under repressing conditions *in vivo*, the *trpEDCBA* operator is occupied by three TrpR dimers, the *aroH* operator by two dimers, and the *trpR* operator by one dimer.

## Discussion

The results reported here provide a functional definition of the DNA sequence requirements for binding of *trp* repressor, TrpR, and show how contact residues identified in the two cocrystal structures of TrpR with DNA affect binding affinity, stoichiometry, and apparent cooperativity. The functional and structural results are in good agreement with each other, and allow us to define the binding site sequence precisely. Examining the effects of sequence and sequence context clearly shows that the crystallographically defined half-site including all contacted base-pairs is sufficient for high-affinity binding when duplicated with appropriate spacing. However, minor deviations from the structurally defined half-site are tolerated by TrpR, though not all deviations are equally well-tolerated. This is most dramatically illustrated in the hierarchical binding to the two unequal half-site pairs of the *S. marcescens trpEDCBA* operator, and also in the ability to bind at only some of the closely related sequences in the polylinker site of the pBend3 vector.

**Figure 5.** Binding of TrpR to three of its *E. coli* operators, analyzed by *in vivo* DMS footprinting. DNA methylation was carried out in growing cells, and protections were detected on isolated chromosomal or plasmid DNA by PCR extension of a radiolabeled oligonucleotide primer or by piperidine cleavage of a radiolabeled operator-region restriction fragment, respectively. PCR-extension products terminate at cytosine residues paired with methylated guanine residues in the opposite strand, while directly cleaved products terminate at methylated guanine residues. Labeled products were resolved on 6% acrylamide sequencing gels; typical autoradiograms are shown. The resolved products were correlated with the sequence in each operator region (indicated to the left of each Figure) by comparison with the products of DNA sequencing reactions carried out in parallel by PCR extension or direct chemical sequencing, as appropriate. Extents of protection (normalized % of the intensity for the same band in the minus-TrpR control strain, CY15071) for the indicated bands were determined from scanning densitometry and are given in the text. (a) *aroH* operator, chromosomal DNA; left, bottom strand; right, top strand. Lanes G, A, T, and C: PCR-extension sequencing reaction products terminating at the indicated residues; lanes CY15071: no repressor; lanes DH1: autoregulated chromosomal copy of *trpR* gene; lanes CY15071/pRLK13: modest, unregulated levels of TrpR; lanes DH1/pRLK13: modest, unregulated levels of TrpR plus autoregulated chromosomal copy of *trpR* gene; lanes CY15071/pJPR2: high, unregulated levels of TrpR; lanes DH1/pJPR2: high, unregulated levels of TrpR plus autoregulated chromosomal copy of *trpR* gene. (b) *trpEDCBA* operator, chromosomal DNA, top strand. Lanes are labeled as in (a). (c) *trpR* operator, plasmid DNA; left, bottom strand, labeled at the *SalI* site; right, top strand, labeled at the *BamHI* site. Lanes CY15071/pBRG: no TrpR; lanes DH1/pBRG: autoregulated chromosomal copy of *trpR* gene; lanes DH1/pRPG16: modest, autoregulated levels of TrpR plus autoregulated chromosomal copy of *trpR* gene. (d) Summary of protection results displayed on the operator-region sequences. Protected guanine residues (filled circles except those above the top strand of the *trpEDCBA* operator) were determined by scanning autoradiograms and normalizing band intensity independently to two different control bands well outside each protected region. Differences in normalized intensity of greater than twofold were scored as protections. Occupancy of half-sites by TrpR is indicated by underlining and overlining.

The choice of 1:1 or tandem 2:1 binding modes also depends on sequence and sequence context, and represents a balance between intrinsic affinity for each 1:1 site and cooperativity in forming the 2:1 complex. The competing effects of affinity and cooperativity are most clearly seen by comparing binding to idealized sequences with two perfect half-site pairs, where 1:1 binding can occur in two equivalent orientations, with binding to sequences having one or more defective half-sites such as the *trpEDCBA* operator of *S. marcescens*. In the latter case one defective half-site pair has the effect of apparently enhancing binding at the other half-site pair. In the limit of completely absent distal half-sites, as for example on the DNA used for cocrystallization of the tandem complex (Carey *et al.*, 1991, 1993; Lawson & Carey, 1993), 2:1 complex formation occurs with lower apparent affinity, and no 1:1 complex is detected on gels. These results illustrate the compensation that occurs between intrinsic affinity and cooperativity in choice of binding mode. Thus, on natural sequences, which typically have at least one imperfect half-site, we can expect some compensation even if it cannot be resolved experimentally.

The measured value of the coupling free energy for TrpR represents the maximum cooperativity due to all non-artifactual sources, including the protein-protein interactions observed in the tandem cocrystal structure (Lawson & Carey, 1993). The measured value of the coupling free energy can be compared with the free energy contribution from tandem protein-protein interactions expected on the basis of the structure of the tandem complex, which shows that approximately 500 Å<sup>2</sup> of surface area are buried at the contact region between adjacent dimers, about two-thirds of which is hydrophobic (Lawson & Carey, 1993). Model compound studies indicate a linear relationship between burial of hydrophobic surface area and free energy of transfer to water, with a proportionality constant of -21 cal/mol Å<sup>2</sup> (Reynolds *et al.*, 1974). The validity of this relationship for proteins has been discussed thoroughly (for a recent review, see Spolar & Record, 1994). Thus, the hydrophobic contribution to the cooperative free energy for tandem TrpR binding is calculated to be approximately -7 kcal/mol. Offsetting contributions from burial of polar surface area and from loss of conformational and configurational entropy must therefore be at work to reduce the net magnitude of the cooperative free energy by at least 5 kcal/mol, although these contributions cannot be quantified separately for lack of adequate model systems.

The results of *in vivo* footprinting indicate that the *trpEDCBA* operator is occupied by three repressor dimers under regulatory conditions *in vivo*, the *aroH* operator by two dimers, and the *trpR* operator by one dimer. These results are in agreement with those determined by Kumamoto *et al.* (1987) using DMS and DNaseI footprinting *in vitro* at high TrpR concentrations. The binding of a third dimer to the *trpEDCBA* operator at the half-site pair at positions

-28 to -24 and -15 to -11 is also consistent with an operator-constitutive mutant, O<sup>c</sup>6, isolated by Bennett & Yanofsky (1978), that reduced repressibility without affecting promoter strength. This mutation was shown to be a change from an A·T to a G·C base-pair at position -15, and gave the strongest effect on repressibility among the mutants in their collection. The structurally defined consensus sequence predicts no role for position -15 unless a third dimer is bound to this half-site pair. The observation that the left half-site of this pair is the most weakly protected region of the operator is consistent with the poor match of this half-site to the consensus sequence.

On the other hand, the two central tandem half-sites are strongly protected despite each having a sequence mismatch (T at positions -11 and -12 instead of the 5' G of the optimal half-site 5' GNACT 3'). The most frequent mismatch among natural half-sites is a replacement of this G residue. The equivalent position is also degenerate in the tandem cocrystal structure (Lawson & Carey, 1993), and there is some hint of alternative electron density for the side-chain of Arg69, which contacts this base (C. L. Lawson & J.C., unpublished observations). It seems likely that the observed strong protection of imperfect central sites is related to the fact that they are bound by adjacent tandem repressors, whereas imperfect sites at the edge of the operator may be more weakly protected in part because they are not shared by an adjacent tandem dimer. Interestingly, the degree of protection at the edge sites in the *trpEDCBA* and *aroH* operators (~50%) is similar to that observed for the singly bound *trpR* operator (40%).

The protection results suggest that the nested 1:1 sites comprising a tandem binding site are occupied simultaneously by tandemly bound repressors, rather than statistically by singly bound repressors, *in vivo*. This conclusion is consistent with that of *in vitro* footprinting which showed that the weaker binding site is occupied only as part of a tandem complex. For all three operators, the insensitivity of the protection pattern to a wide range of TrpR concentrations, taken together with the fact that the three operators have different numbers of repressor dimers bound at equivalent intracellular TrpR concentrations, suggests that the occupancy state (i.e. binding mode) is determined by DNA sequence and not protein concentration over the range studied. As well, positive cooperativity narrows the range of protein concentration over which intermediate occupancy states are populated. Changes in the total extent of occupancy at a given operator (i.e. fraction bound) may occur at lower concentrations of active holorepressor than those we have studied or, alternatively, these changes may be manifested as only minor differences in band intensity. Thus, modulation of transcription from a given promoter in response to changes in active repressor concentration is likely to result from changes in the overall extent of operator occupancy rather than from changes in the binding mode.

It is not clear whether to expect some correlation between the degree of protection and the extent of repression. In fact, the weakest protections are observed on the *trpR* operator. This result could be consistent with the low residual expression of the *trpR* gene observed under repressing conditions when TrpR is produced using its chromosomal signals for transcription and translation (Kelley & Yanofsky, 1982). This gene is regulated over a three fold range (Klig *et al.*, 1988). The strongest protections are observed in the central tandem sites of the *aroH* and *trpEDCBA* operators, which are regulated by sixfold and 70-fold, respectively (Klig *et al.*, 1988).

Why does TrpR use 1:1 or tandem binding modes on different operators? The functional definition of the half-site allows us to rule out the simple explanation that binding stoichiometry compensates for deviations of the operator sequences from the consensus. Rather, the answer may be related to the function of TrpR in differentially modulating the activity of RNA polymerase at each promoter, as proposed originally by Kumamoto *et al.* (1987), since the three regulated operons studied here by *in vivo* footprinting also differ in their degree of repression in response to TrpR (Klig *et al.*, 1988). However, a given repression factor represents the ratio of maximum expression level upon derepression to minimum expression level under repression, and these promoters differ in strength (Cho & Yanofsky, 1988) as well as in their locations relative to the operators, a factor that has also been implicated in control of transcription (Elledge & Davis, 1989). Thus, the correlation between repression and TrpR binding may not be a simple one.

Interestingly, the *trpEDCBA* operator has two TrpR-dependent protections at positions -35 and -34 that do not lie in any identifiable TrpR binding site. Because of their locations, it is tempting to speculate that these sites might be bound by RNA polymerase. *In vitro* footprinting of RNA polymerase on the *S. typhimurium trpEDCBA* promoter (Oppenheim *et al.*, 1980) shows a similar small, isolated cluster of effects at these same bases in the -35 sequence region, as well as another cluster in the -10 region. *In vitro* transcription experiments (Squires *et al.*, 1975) suggest that binding of TrpR and RNA polymerase is mutually exclusive, but footprinting has apparently not been carried out in the presence of both proteins. It seems reasonable to expect that an available -35 region of a strong promoter like that of the *trpEDCBA* operon (Cho & Yanofsky, 1988) might be occupied by polymerase independently of the availability of the -10 region, particularly if repressor and polymerase bind with positive cooperativity. Because cooperativity must act reflexively, it could also help to explain the large repression factor through stabilization of TrpR binding to the adjacent weak half-site. Cooperative binding of TrpR and polymerase would also have the effect of increasing the otherwise relatively low specificity of TrpR (Carey, 1988), by enhancing operator but not non-operator binding. Alterna-

tively, the additional protections in the *trpEDCBA* operator could be related to the WrbA protein, proposed to affect TrpR regulation in stationary phase (Yang *et al.*, 1993), although the footprinting was carried out on cells in the log phase of growth.

The functional definition of the half-site presented here permits a self-consistent interpretation of a large body of previous work on DNA sequence requirements for TrpR binding (Bennett & Yanofsky, 1978; Gunsalus & Yanofsky, 1980; Haydock *et al.*, 1983; Kumamoto *et al.*, 1987; Bass *et al.*, 1987, 1988; Carey, 1989; Staacke *et al.*, 1990; Marmorstein *et al.*, 1991; Carey *et al.*, 1991; Haran *et al.*, 1992; Sutton *et al.*, 1993; LeTilly & Royer, 1993; Liu & Matthews, 1993; Joachimiak *et al.*, 1994; Czernik *et al.*, 1994). In most cases the binding mode was not known in those studies, but where known it agrees with the expectations of the functionally defined binding site presented here. For other cases, the results presented here can be used to predict the binding mode. For example, in both the *mtr* and *aroL* operators, it seems likely that TrpR will bind in tandem to the sites indicated in Figure 1, despite the fact that in each case one half-site (the right-most of *mtr* and the left-most of *aroL*) provides rather poor match to the consensus sequence. Note that the left-most half-site of the *mtr* operator extends to position -28, as does the *trpEDCBA* operator, offering the potential for interaction with RNA polymerase bound at the -35 site of the P3 promoter (Heatwole & Somerville, 1991a,b, 1992; Sarsero *et al.*, 1991). Like the *trpEDCBA*, the *mtr* operator has a very high repression factor (~100-fold; Heatwole & Somerville, 1991b). Both *mtr* and *aroL* are also co-regulated by the tyrosine repressor (Heatwole & Somerville, 1991a,b, 1992; Sarsero *et al.*, 1991; Lawley & Pittard, 1994). In the *aroL* regulatory region, the TyrR binding sites are immediately adjacent to the left-most half-site of the TrpR operator, which has the poorest match to the consensus sequence. In the *mtr* operator the TyrR binding sites are ~90 base-pairs upstream.

The functional definition can also be used to evaluate the likelihood that TrpR will bind at potential sites identified in the available *E. coli* genomic sequence database. A Genbank search for the half-site sequence (5' GNACT 3') using FASTA (Pearson & Lipman, 1988) identified several perfect matches in diverse genes. Visual inspection of the flanking sequences suggests the presence of candidate binding sites in some of these genes, although in no case is a perfect match to consensus found, nor even any sequence with only mismatches that are known to be functional. Furthermore, none of these sites meets the logical, independent criterion for consideration as a new potential binding site, namely, localization in or near the promoter region of a gene whose physiological role places it credibly in the *trp* regulon, but in no case has TrpR binding been tested. Nevertheless, this result suggests that there are probably no cryptic members of the *trp* regulon among already sequenced *E. coli* genes.

Typical prokaryotic repressor-operator systems are examples of mutually embedded sequences that carry out different functions. Operator sequences presumably evolve to permit the level of repressor occupancy required to effect the desired level of control. However, they must do so within the constraints of the promoter function carried out by the overlapping sequence region. A unique operator target within the *E. coli* genome requires specification of 12 base-pairs (von Hippel & Berg, 1986), a number that could be prohibitive in a sequence region with dual function. A minimal site for TrpR binding requires specification of 8 bp with defined spacing among 18 contiguous bp, but specification of only 8 bp has the potential to bring too many sites in the genome (estimated at about 100; von Hippel & Berg, 1986) under the control of TrpR. The likelihood that far fewer than 100 *trp* operators are present in the *E. coli* genome suggests that some counterselection may operate against evolution of incipient TrpR binding sites. The adaptability of binding mode exhibited by TrpR permits this protein to use a wider range of operator sequences because half-site modules can be combined to yield different occupancy states, and the positive cooperativity in tandem binding can compensate for lower affinity of sites with imperfect match to the consensus sequence, further extending the range of sequences compatible with operator function. The tolerance of TrpR for deviations from the consensus sequence and for different occupancy states can reduce the number of base-pairs that must be specified to define a TrpR operator, and thus extends the range of sequence compatibility for operator and promoter function. All these sources of adaptability by TrpR may also provide additional opportunities for fine-tuning in the *trp* regulon.

## Materials and Methods

### Strains and plasmids

Plasmid pRK9 (Nichols & Yanofsky, 1983) was the source of the *trpEDCBA* operator of *S. marcescens* used for the *in vitro* DMS modification-protection experiment. *In vivo* DMS footprinting of TrpR on chromosomal *trpEDCBA* and *aroH* operators was carried out in *E. coli* strains CY15071, a W3110 derivative with the chromosomal *trpR* gene deleted (Paluh & Yanofsky, 1986), and DH1, a wild-type strain. Plasmids pRLK13 (Kelley & Yanofsky, 1985) and pJPR2 (Paluh & Yanofsky, 1986) were used to transform these strains, as indicated in the Figure legends. *In vivo* DMS footprinting of the *trpR* operator was performed on a plasmid, pBRG, constructed to contain the *trpR* operator but not the *trpR* structural gene by cloning the 450 bp *Bam*HI-*Sal*I fragment of pRPG16 (Gunsalus *et al.*, 1979) into pBN60 (Joachimiak *et al.*, 1983) in place of its 760 bp *Bam*HI-*Sal*I fragment. pBRG was then transformed into CY15071 and DH1 separately.

### Oligonucleotides

All oligonucleotides were synthesized in the synthesis/sequencing facility in the Molecular Biology Department at Princeton University. Self-complementary oligonucleo-

tides were annealed by heating the single strands in annealing buffer (0.1 mM EDTA, 0.1 M NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> (pH 7.0)) at 90°C for five minutes and cooling slowly to room temperature. Oligonucleotides were labeled by end-filling with the large fragment of DNA polymerase I, or by polynucleotide kinase, as appropriate. For evaluation of the effects of distal and flanking sequences, an oligonucleotide (5'TCGATAGC-GTACTAGTACGCTA3'), which contains an operator target for 2:1 binding, was synthesized. After self-annealing, the operator was cloned into the unique *Sal*I restriction site in the polylinker region of pBend3 (Zwieb & Adhya, 1994; kindly provided by Dr Richard Ebright, Rutgers University). A 30-bp (after end-filling) symmetrized synthetic operator containing three optimized half-sites was used to measure the cooperativity: 5' TTGTACTAGT-GCACT AGTGCACTAGTACA 3'. The following 10-bp (after end-filling) sequence was synthesized to test the minimal DNA length for tandem binding: 5'GTAC-TAGTA3'.

### Chemicals and enzymes

Dimethylsulfate (DMS; Sigma), L-tryptophan (L-trp; Sigma) and piperidine (Fisher) were used without further purification. Restriction enzymes were purchased from New England Biolabs. T4 polynucleotide kinase and shrimp alkaline phosphatase were purchased from USB. [ $\gamma$ -<sup>32</sup>P]ATP (6000 Ci/mmol) was purchased from NEN (DuPont).

### Protein and DNA purification

All plasmid DNAs, including those methylated, were purified with the Wizard mini-prep kit (Promega). All DNA fragments were purified from agarose gels using the GeneClean kit (Bio101, Inc.). DNA concentrations were determined either by UV absorption at 260 nm or on agarose gel by ethidium bromide staining. Single-end-labeled restriction fragments were prepared from plasmids by T4 polynucleotide kinase labeling at the first restriction digestion site followed by recutting at the second restriction site. Synthetic oligonucleotides were generally labeled at both 5' ends by T4 polynucleotide kinase.

*trp* aporepressor was overexpressed in strain CY15071/pJPR2 and purified to homogeneity as described (Joachimiak *et al.*, 1983; Paluh & Yanofsky, 1986), with the exception that a French press was used instead of lysozyme to open the cells. Protein concentrations were determined spectrophotometrically using an extinction coefficient of 1.2 cm<sup>-1</sup> mg<sup>-1</sup> ml at 280 nm (Joachimiak *et al.*, 1983).

### Assays

#### Gel retardation assay

This assay was performed as described (Carey, 1988). A typical titration experiment used a fixed, low concentration of labeled DNA ( $\leq 0.01$  nM unless otherwise stated) mixed with a range of TrpR concentrations; samples were incubated and then run at room temperature on a 10% (w/v) polyacrylamide gel (acrylamide:bis = 29:1). Dried gels were exposed to Kodak X-ray film or, for quantitative analysis, in phosphorimager cassettes and scanned with a Molecular Dynamics PhosphorImager. Bands were analyzed densitometrically

with the ImageQuant program from Molecular Dynamics. Fractions of each component determined from band intensities were fitted (Yang, 1995) to equations describing the cooperativity (Senear & Brenowitz, 1991):

$$q_0 = 1/[1 + (k_1 + k_2)L + k_1k_2k_{12}L^2] \quad (1)$$

$$q_1 = (k_1 + k_2)L/[1 + (k_1 + k_2)L + k_1k_2k_{12}L^2] \quad (2)$$

$$q_2 = k_1k_2k_{12}L^2/[1 + (k_1 + k_2)L + k_1k_2k_{12}L^2] \quad (3)$$

$$k_1 = k_2$$

where  $k_1$  and  $k_2$  are the microscopic equilibrium association constants for intrinsic binding to sites 1 and 2, respectively;  $k_{12}$  is the constant describing cooperative interactions when both sites are liganded; and  $q_i$  ( $i = 0, 1, 2$ ) is the fraction of DNA molecules with exactly  $i$  ligands bound. Optimal gel conditions used in cooperativity studies were 4°C, 12.5% acrylamide.

### DMS footprinting

*In vitro*: Protein-DNA complexes were probed with DMS before being resolved by gel retardation. To a 30 µl reaction mixture, 1 µl of a 1:10 dilution of DMS (in 0.2 M NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, pH 6.6) was added and methylation was allowed to proceed for six minutes at room temperature. The reaction was then loaded directly onto a retardation gel. Free DNA and protein-DNA complexes were located by exposing wet gels to X-ray film, and were isolated by the "crush and soak" method (Maniatis *et al.*, 1982). Modified guanine residues were cleaved by piperidine treatment as described by Maxam & Gilbert (1980). Cleavage products were analyzed in a 6% sequencing gel containing 8 M urea.

*In vivo*: DNA methylation was carried out in a 50 ml cell culture grown in LB medium at 37°C to  $A_{600} = 0.6$  to 0.8 by adding neat DMS at room temperature for 30 seconds prior to harvesting the cells; 100 µl DMS was used for chromosomal DNA methylation, and 15 µl was used for plasmid methylation. Methylated chromosomal DNAs were purified with a Qiagen chromosomal purification kit according to manufacturer's instructions. Chromosomal DNAs (10 µg) were treated with piperidine (final concentration 0.01 M), and/or heated at 90°C (Brewer *et al.*, 1990), and purified with a G-50 spin column essentially as described (Sasse-Dwight & Gralla, 1991). Modified sites were identified on each strand by using the polymerase chain reaction (PCR) to extend 5'-end-labeled primers that anneal near each operator:

trpEDCBA (top): 5' CTCCCGTTCTGGATAATG 3'

trpEDCBA (bottom): 5' GCCCGTTTCAGGAAGTGC 3'

aroH (top): 5' CCGTGGATTAACAGTACC 3'

aroH (bottom): 5' GCTGCTGATTTCAAGTGC 3'

(Top and bottom are arbitrary designations that refer to the sequences in each operator region as illustrated in the figures of this paper.) For DNA footprinting on the *trpR* operator of plasmid pBRG, one-end-labeled methylated DNA fragments (*Sal*I-*Aat*II for bottom strand and *Bam*HI-*Sal*I for top strand) were purified, and the modified G residues were cleaved with piperidine as described by Maxam & Gilbert (1980).

Quantitation of protected bands on both *in vitro* and *in vivo* footprints was accomplished using scanning densitometry of the autoradiograms and processing by NIH Image Analysis. Normalization used two different bands from outside the protected region for each

operator. The extent of protection is reported as the normalized percent intensity of a band relative to the intensity of the same band in the control lane.

## Acknowledgements

The authors thank Michael Beck and Kathy Tuohy for their contributions to related experiments; Bin Gao for help with data analysis; Sancar Adhya and Richard Ebricht for gifts of pBend3; Shirley Tilghman for the use of phosphorimager cassettes; and the members of our laboratory for discussion and comments on the manuscript.

## References

- Bass, S., Sugiono, P., Arvidson, D. N., Gunsalus, R. P. & Youderian, P. (1987). DNA specificity determinants of *Escherichia coli* tryptophan repressor binding. *Genes Dev.* **1**, 565–572.
- Bass, S., Sorrells, V. & Youderian, P. (1988). Mutant *trp* repressors with new DNA-binding specificities. *Science*, **242**, 240–245.
- Bennett, G. N. & Yanofsky, C. (1978). Sequence analysis of operator constitutive mutants of the tryptophan operon of *Escherichia coli*. *J. Mol. Biol.* **121**, 179–192.
- Brewer, A. C., Mason, P. J. & Patient, R. F. (1990). A simplified method for *in vivo* footprinting using DMS. *Nucl. Acids Res.* **18**, 5574–5574.
- Cann, J. R. (1989). Phenomenological theory of gel electrophoresis of protein-nucleic acid complexes. *J. Biol. Chem.* **264**, 17032–17040.
- Carey, J. (1988). Gel retardation at low pH resolves *trp* repressor-DNA complexes for quantitative study. *Proc. Natl Acad. Sci. USA*, **85**, 975–979.
- Carey, J. (1989). *trp* Repressor arms contribute binding energy without occupying unique locations on DNA. *J. Biol. Chem.* **264**, 1941–1945.
- Carey, J., Lewis, D. E. A., Lavoie, T. A. & Yang, J. (1991). How does *trp* repressor bind to its operator? *J. Biol. Chem.* **266**, 24509–24513.
- Carey, J., Combatti, N., Lewis, D. & Lawson, C. L. (1993). Cocrystals of *Escherichia coli trp* repressor bound to an alternative operator DNA sequence. *J. Mol. Biol.* **234**, 496–498.
- Cho, K.-O. & Yanofsky, C. (1988). Development of a *trp* promoter-strength measuring system and its use in comparison of the *trp* EDCBA, *trp* R, and *aro* H promoters. *J. Mol. Biol.* **204**, 41–50.
- Chou, W.-Y. & Matthews, K. S. (1989). Serine to cysteine mutations in *trp* repressor protein alter tryptophan and operator binding. *J. Biol. Chem.* **264**, 18314–18319.
- Czernick, P., Shin, D. S. & Hurlburt, B. K. (1994). Functional selection and characterization of DNA binding sites for *trp* repressor of *Escherichia coli*. *J. Biol. Chem.* **269**, 27869–27875.
- Elledge, S. J. & Davis, R. W. (1989). Position and density effects on repression by stationary and mobile DNA-binding proteins. *Genes Dev.* **3**, 185–197.
- Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. (1985). Lineage-specific interactions of an immunoglobulin enhancer with cellular factors *in vivo*. *Science*, **227**, 134–140.
- Gunsalus, R. P. & Yanofsky, C. (1980). Nucleotide sequence and expression of *Escherichia coli trpR*, the structural gene for the *trp* aporepressor. *Proc. Natl Acad. Sci. USA*, **77**, 7117–7121.

- Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1979). Structural and functional analysis of cloned deoxyribonucleic acid containing the trpR-thr region of the *E. coli* chromosome. *J. Bacteriol.* **140**, 106–113.
- Gunsalus, R. P., Miguel, A. G. & Gunsalus, G. L. (1986). Intracellular Trp repressor levels in *Escherichia coli*. *J. Bacteriol.* **167**, 272–278.
- Haran, T. E., Joachimiak, A. & Sigler, P. B. (1992). The DNA target of the *trp* repressor. *EMBO J.* **11**, 3021–3030.
- Haydock, P. V., Bogosian, G., Brechling, K. & Somerville, R. L. (1983). Studies on the interaction of *trp* holorepressor with several operators. *J. Mol. Biol.* **170**, 1019–1030.
- Heatwole, V. M. & Somerville, R. L. (1991a). Cloning, nucleotide sequence, and characterization of *mtr*, the structural gene for a tryptophan-specific permease of *Escherichia coli* K-12. *J. Bacteriol.* **173**, 108–115.
- Heatwole, V. M. & Somerville, R. L. (1991b). The tryptophan-specific permease gene, *mtr*, is differentially regulated by the tryptophan and tyrosine repressors in *E. coli* K12. *J. Bacteriol.* **173**, 3601–3604.
- Heatwole, V. M. & Somerville, R. L. (1992). Synergism between the *trp* repressor and *tyr* repressor in repression of the *aroL* promoter of *Escherichia coli* K-12. *J. Bacteriol.* **174**, 331–335.
- Hurlburt, B. K. & Yanofsky, C. (1992). The NH<sub>2</sub>-terminal arms of *trp* repressor participate in repressor/operator association. *Nucl. Acids Res.* **20**, 337–341.
- Joachimiak, A., Kelley, R. L., Gunsalus, R. P., Yanofsky, C. & Sigler, P. (1983). Purification and characterization of *trp* aporepressor. *Proc. Natl Acad. Sci. USA*, **80**, 668–672.
- Joachimiak, A., Haran, T. E. & Sigler, P. B. (1994). Mutagenesis supports water-mediated recognition in the *trp* repressor-operator system. *EMBO J.* **13**, 367–372.
- Kelley, R. L. & Yanofsky, C. (1982). *trp* aporepressor production is controlled by autogenous regulation and inefficient translation. *Proc. Natl Acad. Sci. USA*, **79**, 3120–3124.
- Kelley, R. L. & Yanofsky, C. (1985). Mutational studies with the *trp* repressor of *E. coli* support the helix-turn-helix model of repressor recognition of operator DNA. *Proc. Natl Acad. Sci. USA*, **82**, 483–487.
- Klig, L. S., Carey, J. & Yanofsky, C. (1988). *trp* repressor interactions with the *trp*, *aroH*, and *trpR* operators. *J. Mol. Biol.* **202**, 769–777.
- Kumamoto, A., Miller, W. & Gunsalus, R. P. (1987). *E. coli* tryptophan repressor binds multiple sites within the *aroH* and *trp* operators. *Genes Dev.* **1**, 556–564.
- Lawley, B. & Pittard, A. J. (1994). Regulation of *aroL* expression by TyrR protein and Trp repressor in *E. coli* K12. *J. Bacteriol.* **176**, 6921–6930.
- Lawson, C. L. & Carey, J. (1993). Tandem binding in crystals of a *trp* repressor/operator half-site complex. *Nature*, **366**, 178–182.
- LeTilly, V. & Royer, C. A. (1993). Fluorescence anisotropy assays implicate protein-protein interactions in regulating *trp* repressor DNA binding. *Biochemistry* **32**, 7753–7758.
- Lewis, D. E. A. & Carey, J. (1993). Analysis of *trp* repressor-DNA interactions using gel electrophoresis. *Electrophoresis*, **14**, 713–719.
- Liu, Y.-C. & Matthews, K. S. (1993). Dependence of *trp* repressor-operator affinity, stoichiometry, and apparent cooperativity on DNA sequence and size. *J. Biol. Chem.* **268**, 23239–23249.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Marmorstein, R. Q., Sprinzl, M. & Sigler, P. B. (1991). An alkaline phosphatase protection assay to investigate *trp* repressor/operator interactions. *Biochemistry*, **30**, 1141–1148.
- Maxam, A. M. & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 499–560.
- Nichols, B. P. & Yanofsky, C. (1983). Plasmids containing the *trp* promoters of *Escherichia coli* and *Serratia marcescens* and their use in expressing cloned genes. *Methods Enzymol.* **101**, 155–164.
- Oppenheim, D. S., Bennett, G. N. & Yanofsky, C. (1980). *Escherichia coli* RNA polymerase and *trp* repressor interaction with the promoter-operator region of the tryptophan operon of *Salmonella typhimurium*. *J. Mol. Biol.* **144**, 133–142.
- Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature*, **335**, 321–329.
- Paluh, J. L. & Yanofsky, C. (1986). High-level production and rapid purification of the *E. coli trp* repressor. *Nucl. Acids Res.* **14**, 7851–7860.
- Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
- Reynolds, J. A., Gilbert, D. B. & Tanford, C. (1974). Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Natl Acad. Sci. USA*, **71**, 2925–2927.
- Sarsero, J. P., Wookey, P. J. & Pittard, A. J. (1991). Regulation of expression of *Escherichia coli* K-12 *mtr* gene by TyrR protein and *trp* repressor. *J. Bacteriol.* **173**, 4133–4143.
- Sasse-Dwight, S. & Gralla, J. D. (1991). Footprinting protein-DNA complexes *in vivo*. *Methods Enzymol.* **208**, 146–168.
- Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188**, 415–431.
- Senear, D. F. & Brenowitz, M. (1991). Determination of binding constants for cooperative site-specific protein-DNA interactions using the gel mobility-shift assay. *J. Biol. Chem.* **266**, 13661–13671.
- Spolar, R. S. & Record, M. T. (1994). Coupling of local folding to site specific binding of proteins to DNA. *Science*, **263**, 777–784.
- Squires, C. L., Lee, F. D. & Yanofsky, C. (1975). Interaction of the *trp* repressor and RNA polymerase with the *trp* operon. *J. Mol. Biol.* **92**, 93–111.
- Staacke, D., Walter, B., Kisters-Woike, B., Wilcken-Bergmann, B. V. & Muller-Hill, B. (1990). How *trp* repressor binds to its operator. *EMBO J.* **9**, 1963–1967.
- Sutton, C. L., Mazumder, A., Chen, C. H. & Sigman, D. S. (1993). Transforming the *E. coli trp* repressor into a site-specific nuclease. *Biochemistry*, **32**, 4225–4230.
- von Hippel, P. H. & Berg, O. G. (1986). On the specificity of DNA-protein interactions. *Proc. Natl Acad. Sci. USA*, **83**, 1608–1612.
- Weber, G. (1993). *Protein Interactions*. Chapman & Hall, New York.

- Wissman, A. & Hillen, W. (1992). DNA contacts probed by modification protection and interference studies. *Methods Enzymol.* **208**, 365–379.
- Yang, J. (1995). *In vivo* and *in vitro* studies of TrpR-DNA interactions. PhD. Thesis, Princeton University, Princeton, N.J.
- Yang, J. & Carey, J. (1995). Footprint phenotypes: structural models of DNA-binding proteins from chemical modification analysis of DNA. *Methods Enzymol.* **259**, 452–468.
- Yang, W., Ni, L. & Somerville, R. L. (1993). A stationary-phase protein of *Escherichia coli* that affects the mode of association between the *trp* repressor protein and operator-bearing DNA. *Proc. Natl Acad. Sci. USA*, **90**, 5796–5800.
- Zurawski, G., Gunsalus, R. P., Brown, K. D. & Yanofsky, C. (1981). Structure and regulation of *aroH*, the structural gene for the tryptophan-repressible 3-deoxy-D-arabino-heptulosonic acid-7-phosphate synthetase of *Escherichia coli*. *J. Mol. Biol.* **145**, 47–53.
- Zwieb, C. & Adhya, S. (1994). Improved plasmid vectors for the analysis of protein-induced DNA bending. *Methods. Mol. Biol.* **30**, 281–294.

*Edited by P. E. Wright*

(Received 6 June 1995; received in revised form 11 January 1996; accepted 30 January 1996)