The EMBO Journal vol.11 no.8 pp.3021–3030, 1992

The DNA target of the *trp* repressor

Tali E.Haran, Andrzej Joachimiak and Paul B.Sigler

Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University, New Haven, CT 06511, USA

Communicated by A.Klug

Unexpected features seen by high resolution X-ray crystallography at the interface of the *trp* repressor and the 'traditional' *trp* operator provoked the claim that the DNA fragment used in the crystal structure is not the true operator, and therefore that the crystal structure of the *trp* repressor–operator complex does not portray a specific interaction. An alternative sequence was proposed mainly on the basis of mutational studies and gel retardation analysis of short target duplexes (Staacke *et al.*, 1990a,b). We have reexamined the sequence consensus in trpR-repressible promoters and analyzed the mutagenesis experiments of others including Staacke *et al.* (1990a) and found them fully consistent with the interactions of the traditional operator sequence seen in the crystal structure, and stereochemically inconsistent with the above referenced alternative model. Moreover, an *in vitro trp* repressor–DNA binding analysis, employing both novel DNA constructs devised to avoid previously encountered artifacts as well as full-length promoter sequences, indicates that the traditional operator used in the crystal structure is the preferred target of the *trp* repressor.

Key words: DNA hairpins/gel retardation analysis/repressor–operator stoichiometry/*trp* operator

Introduction

It was commonly held that DNA-binding regulatory proteins are targeted to their appropriate sequences primarily by direct contacts between the side chains of the protein and the bases that specify the target (Ptashne, 1986; von Hippel and Berg, 1986). While many structural studies have supported this view, the first description of a specific protein–DNA interface in atomic detail did not (Otwinowski *et al.*, 1988). The 2.4 Å crystal structure of the *trp* repressor–operator complex indicated that direct hydrogen bonds between the protein's side chains and the functional groups in the major groove of the *trp* operator are not the major determinants of specificity. Instead, water molecules mediate hydrogen-bonded contacts between the repressor's helix–turn–helix motif and six of the eight base pairs that contribute most to the operator's identity. The direct hydrogen-bonded contacts are mainly to the phosphate backbone, thereby anchoring the bihelical motif in the major groove. To make the hydrogen-bonded contacts furthest from the central axis of rotational symmetry and to maximize the solvent-excluded interface (2900 $Å^2$), the operator bends towards the major groove by 14° at the T_4A_5 step (see Figure 1 for numbering scheme). In presenting these findings, Otwinowski *et al.* (1988) invoked the possibility of an 'indirect readout' mechanism, that is, sequence-specific deviations (intrinsic and/or induced) from canonical B-DNA are required to make an optimal interface between target DNA and the protein that 'recognizes' it. This suggestion remains to be experimentally confirmed.

The unexpected chemistry of the *trp* repressor–operator interface in the crystalline complex generated some skepticism about the relevance of the crystal structure to *in vivo* function. The criticism fell into two categories. First, there was concern that the conditions of crystallization, in particular the use of organic solvents, might alter the chemistry of the interface (Brennan and Matthews, 1989). The second concern, expressed most forcefully by Staacke *et al.* (1990a,b) states that the operator sequence chosen for the crystallographic experiment was incorrect and, therefore, that conclusions drawn from the crystal structure were not relevant to operator-specific binding. This paper addresses the issue raised by Staacke *et al.* (1990a) regarding the correctness of the operator sequence used in the cocrystalline complex. Experiments validating the chemistry observed at the crystalline repressor–operator interface are addressed elsewhere (A.Joachimiak, T.E.Haran and P.B.Sigler, in preparation). We show here, by *in vitro* binding experiments as well as the analysis of sequence consensus and the mutagenesis experiments of others (Staacke *et al.*, 1990a; Bennett and Yanofsky, 1978; Bass *et al.*, 1987), that the operator sequence used in the crystal structure is, indeed, the preferred target of the *trp* repressor. Furthermore, we complement previous *in vitro* binding studies (Joachimiak *et al.*, 1983; Klig *et al.*, 1987; Carey, 1988, 1989; Marmorstein and Sigler, 1989; Hurlburt and Yanofsky, 1990) by examining the stoichiometry and relative affinity of *trp* repressor binding to the wild-type and 'mutant' *trpEDBCA* promoters of *Escherichia coli*. From these studies we infer patterns of repressor alignment on the promoter. Finally, we suggest how the experiments reported by Staacke *et al.* (1990a), both in their original paper and in their subsequent correction (Staacke *et al.*, 1990b), could lead to the wrong conclusion.

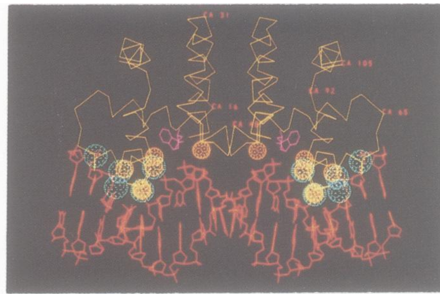
Results

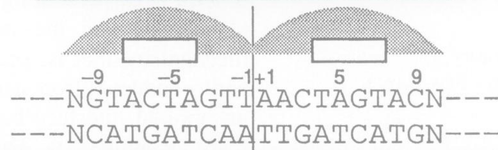
Two distinctly different trp operators have been proposed

Figure 1a introduces the symbol used throughout this paper to represent the dimeric *trp* repressor with its symmetrically disposed recognition surfaces. These recognition surfaces, indicated by the open rectangles, belong to the repressor's flexible 'reading heads', which contain the helix–turn–helix motif and almost all the amino acids shown by saturation (Kelley and Yanofsky, 1982) and site-directed mutagenesis (Bass *et al.*, 1988) to be directly involved in operator binding

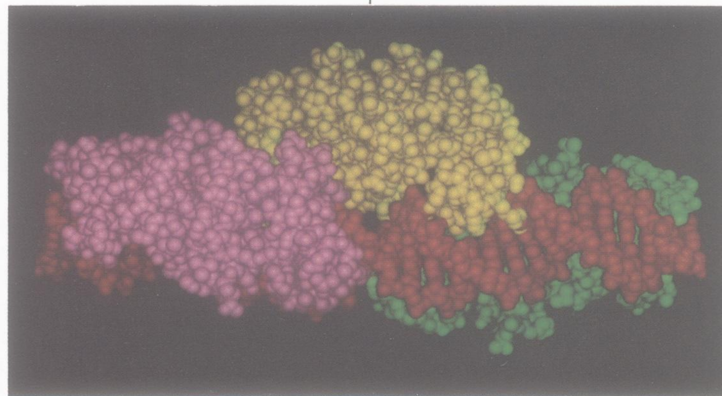
© Oxford University Press

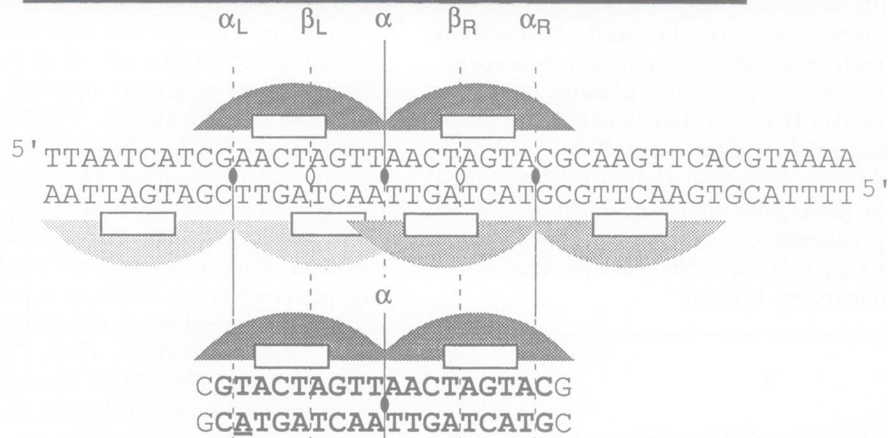
a



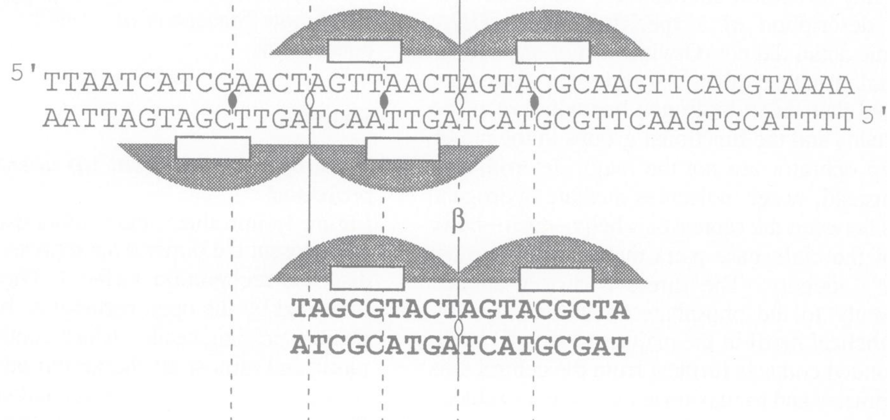


b





c



in vivo (Figure 1a). The orientation of the reading heads has been shown by crystallography (Schevitz *et al.*, 1985; Zhang *et al.*, 1987; Lawson *et al.*, 1988; Otwinowski *et al.*, 1988) and NMR (Arrowsmith *et al.*, 1990, 1991) to account for the tryptophan-induced allosteric modulation of operator affinity—hence their name. The symbols for the repressor and the DNA sequence are drawn closely to scale in order to depict accurately which base pairs would be juxtaposed to the repressor's reading heads by the choice of a particular target sequence.

Two distinctly different alignments have been proposed for the specific interaction of *trp* repressor with the *trpEDCBA* promoter (Figure 1). The 'traditional' mode superimposes the molecular dyad of the dimeric repressor on the DNA sequence dyad called 'α' (Figure 1b). In this arrangement, the repressor contacts an 18 bp α-centered symmetrical target sequence that until recently has been accepted as the *trp* operator (Figure 1b). This is the sequence used in the crystal structure. Figure 1b illustrates an important point made in the report of the crystal structure (Otwinowski *et al.*, 1988) namely, that the repressor bound to the α-centered operator can be flanked—without steric hindrance—by two similarly bound repressors each centered eight base pairs to either side of the α-dyad and bound to the DNA with the same interface. These flanking targets, suggested by Kumamoto *et al.* (1987) on the basis of their alkylation studies, have approximate rotational symmetry about α_R and (less so) about α_L. The inverted repeats of the three overlapping α-centered targets are indicated by the arrows in Figures 2, 3 and 5.

Staacke *et al.* (1990a) challenged the traditional alignment by asserting that specific binding to the *trpEDCBA* promoter entails tandem binding of two dimeric repressors each centered on an approximate dyad spaced eight base pairs apart. The dyads are called β-dyads (Figure 1c). The β_L-dyad is displaced four base pairs to the left, and the β_R-dyad four base pairs to the right, of the α-dyad that previous workers have taken as the center of the operator's sequence and the one chosen for the crystallographic work. Since the β-dyad is four base pairs displaced from the one used in the crystal structure, the target DNA shown in Figure 1c will display a substantially different array of functional groups to the binding surface of the repressor. Moreover, the β-centered target sequence could also exhibit different indirect readout. Note that tandem binding, *per se*, is not a point of contention. The point at issue here is whether the repressor dimer's preferred target is the sequence centered on the α-dyad or the sequence centered on the β-dyad.

Sequence consensus and mutagenesis support the traditional model

Figure 2 depicts the relationship of the repressor's recognition surfaces to the base pairs in the *trpEDCBA* promoter−operator that have been implicated as the operator's identity elements by consensus (Figure 2a−d) and mutagenesis (Figure 2e and f). The traditional α-centered arrangement observed in the crystal structure positions the protein's recognition machinery such that it directly faces the operator's identity elements and, therefore, can form specific interactions (Figure 2a and b). In contrast to the α-centered model which juxtaposes the reading heads and identity elements, the β-centered arrangement (Figure 2c, d and f) places the identity elements of the operator midway between the reading heads. Indeed, the β-centered model of the Köln group (Staacke *et al.*, 1990a) aligns the helix−turn−helix motifs so that they cannot reach most of the bases that define the operator's identity.

Staacke *et al.* (1990a) have also defined the identity elements of the *trp* repressor's target by systematic mutagenesis of a 'symmetrized' operator construction that represses β-galactosidase expression. Figure 2f shows that within the limits arising from differences in the symmetrical design of the experiments, the distribution of mutationally sensitive bases found by Staacke *et al.* (1990a) agrees with that found by the mutagenesis experiments of Bass *et al.* (1987; Figure 2e) and with the identity elements implied by the sequence consensus of three trpR-repressible promoters in *E.coli* (Figure 2a and b).

Gel retardation of short stable hairpin duplexes indicates an α-centered target

Gel retardation experiments were devised to show whether an α-centered or a β-centered sequence is the preferred target of the *trp* repressor. As is customary in such experiments we, like Staacke *et al.* (1990a), initially used 'two-stranded' duplexes that were short enough to limit the alternative modes of binding. The term 'two-stranded' denotes a DNA duplex composed of two noncovalently linked complementary oligonucleotides. The restrictive value of short sequences ($\leq$21 bp) and the potential for ambiguity of larger duplexes ($\leq$26 bp) is illustrated in Figure 1b and c. However, the use of short two-stranded duplexes, especially palindromic ones, introduces an often disregarded complication into affinity measurements that can affect the apparent stability of the retarded protein−DNA complex; that is, two-stranded duplexes can readily disproportionate into 'one-stranded hairpins', particularly at the very low DNA concentrations used to quantify affinity. To avoid uncertainty in the concentration and stability of the duplex target, we used a single continuous strand to create an enlarged hairpin (Chu and Orgel, 1991) whose stem forms the desired duplex irrespective of concentration. We, like Carey (1988, 1991), established the protein−DNA stoichiometry in the retarded complexes with radioactive protein of known specific radioactivity and ^{32}P-labeled DNA. Instead of [^{3}H]leucine-labeled protein, we prepared freshly labeled

Fig. 1. Proposed alignments of the *trp* repressor on the *trpEDCBA* promoter. **(a)** The crystal structure of the *trp* repressor−operator complex where the 'operator' is the DNA sequence chosen by sequence consensus and mutagenesis (Figure 2). The symbol is drawn to indicate the repressor's dimeric character and the open rectangular blocks designate the position of the recognition surfaces as determined by: (i) the negative complementing mutational changes of Kelley and Yanofsky (1982) (orange spheres); (ii) the site-directed mutational studies of Bass *et al.* (1988) (larger blue spheres) and (iii) the regions that are allosterically modulated by L-tryptophan binding (Zhang *et al.*, 1987; Luisi and Sigler, 1990). The symbol for the protein and the DNA sequence below it are drawn to a scale that is preserved throughout this paper. The designations of α- and β-dyads are those of Staacke *et al.* (1990). **(b)** Tandem binding of three *trp* repressors to the *trpEDCBA* promoter. The center of each repressor is eight base pairs from that of its neighbor as suggested by Kumamoto *et al.* (1987). This model was derived by a simple rotation without conformational adjustments. The same arrangement is shown symbolically on the promoter sequence with the α-dyads spaced eight base pairs apart. The bold type denotes the duplex used in the crystal structure and made dyad symmetric around α by replacing the naturally occurring A·T base pair with a T·A base pair at the underlined position. The flanking CG base pair derives from the consensus. **(c)** The alignment of *trp* repressors on the *trpEDBCA* promoter suggested by Staacke *et al.* (1990a). The bold-type sequence is the β-centered *trp* operator used in the band-shift and mutagenesis experiments of Staacke *et al.* (1990a). The sequence was made symmetrical around the β-dyad by replacements described by these authors.

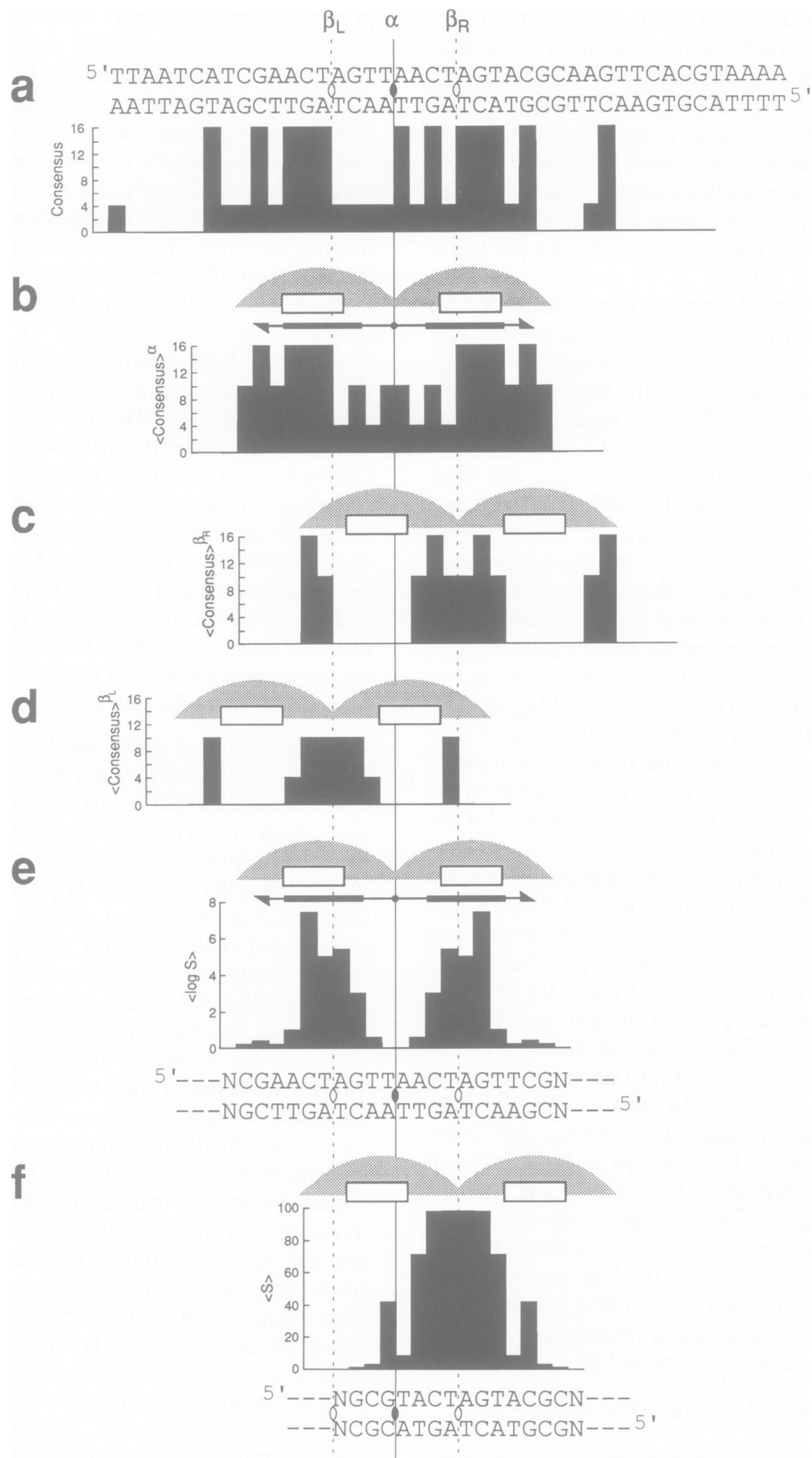


Fig. 2. Identity elements of the *trp* operator. (**a**) Sequence consensus among the three *E.coli* promoters of *trpEDCBA*, *aroH* and *trpR* based on the identity of corresponding sequence positions. The sequence of the *trpEDCBD* promoter is shown at the top for reference. The consensus at each sequence position is scored as: zero for no correspondence; 4 for the occurrence of the same base pair twice (random likelihood is 1/4) and 16 for the occurrence of the same base pair in all three (random likelihood is 1/16). Rotationally averaged consensus around (**b**) the α-dyad, (**c**) the β_R-dyad, and (**d**) the β_L-dyad. The score is zero if symmetrical positions have different base pairs. (**e**) The average log of the sensitivity S of a sequence position to mutations introduced symmetrically around the α-dyad as determined by Bass *et al.* (1987) using the challenge assay of Youderian *et al.* (1983). $S = \mathrm{EOB_0/EOB_i}$ which is the decrement in *in vitro* affinity ('efficiency of binding') caused by one of the three possible symmetrical base pair substitutions at the designated position in an α-centered reference target (5'-CGAACTAGTT.AACTAGTTCG-3'). The thick portions of the horizontal arrows (b and e) signify the traditionally accepted identity elements based on this consensus and mutational studies. (**f**) The average sensitivity S to symmetrical mutations in the β-centered target sequence (Figure 1c) as monitored by Staacke *et al.* (1990a) using β-galactosidase repression. $S = R_0/R_i$ where R_0 and R_i are the repression observed respectively for the reference operator and for one of the three possible symmetrical substitutions (around β) at the indicated site.

[^{125}I]repressor before each experiment and showed by isotope dilution that the labeled protein was unchanged in its affinity for DNA (T.E.Haran, A.Joachimiak and P.B. Sigler, in preparation; see Materials and methods and the legend of Figure 6).

Figure 3a shows that the traditional model predicts that the stem of a hairpin duplex containing only the traditional α-centered operator sequence will bind one dimeric repressor with high affinity. In contrast, the model of the Köln group predicts that this sequence will bind two dimeric repressors. The result is clear—as predicted by the traditional model, the *trp* repressor binds a single α-centered target strongly with 1:1 stoichiometry even at very high concentrations of repressor. Figure 3b shows that the traditional repressor–operator alignment will cause the β-centered hairpin stem to bind two overlapping dimeric repressors, each centered on an α-dyad eight base pairs apart (see Figure 1b). The binding would be weakened because the flanking half site of each target is incomplete (broken arrow). The model of the Köln group predicts strong binding of a single dimeric repressor. Here again, the crystallographic/traditional model correctly predicts the outcome of the experiment—two repressors bind more weakly.

Gel retardation experiments with short unstable two-stranded duplexes

Figure 4 demonstrates that the relative instability of a short two-stranded duplex can have a significant effect upon the appearance and interpretation of gel retardation experiments. Panel (i) in Figure 4a shows that an 18 bp two-stranded α-centered duplex forms a complex with repressor, as indicated by the loss of the unshifted duplex band. However, the complex is unstable as it migrates in the gel and produces a poorly defined smear. In contrast, panel (iii) shows that the repressor clearly binds to the same α-centered target sequences in the 21 bp stem of a large hairpin and gives a sharp band comprised of a 1:1 complex. This result was also seen in Figure 3. It is important to note than in panels (i) and (ii) there is a substantial amount of unshifted DNA migrating ahead of the unshifted duplex. The rapid mobility and the inverse concentration dependence of this band indicate the existence of small hairpins formed from one strand of the two-stranded palindromic target duplex. When the concentrations of the target sequences are raised 10-fold (Figure 4b), we see a lower amount of rapidly migrating small hairpins and a sharpening of the shifted bands. Thus, the apparently 'poor shift' of the α-centered target is related

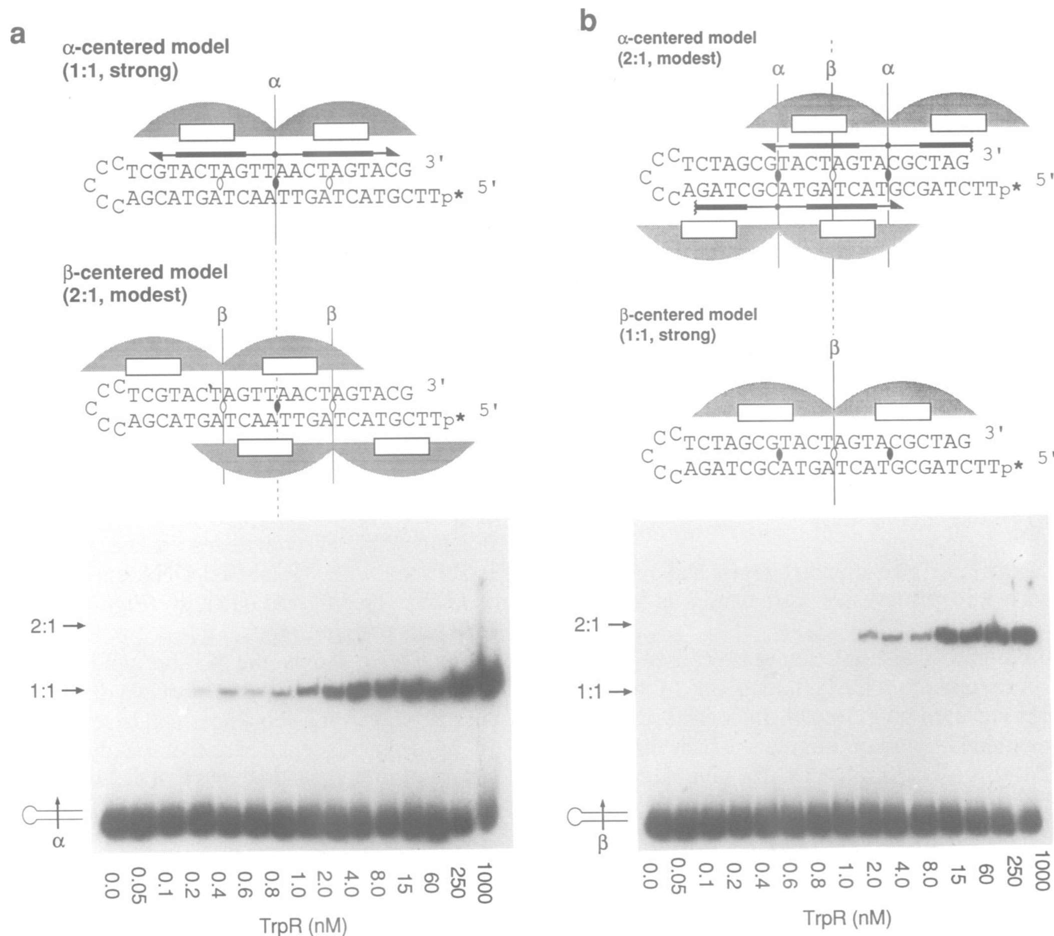


Fig. 3. Gel retardation analysis of alternative operator sequences. (**a**) The hairpin whose stem contains the symmetrized α-centered duplex shown in Figure 1a, and (**b**) the symmetrized β-centered duplex shown in Figure 1c. DNA concentration is <0.2 nM in both experiments. In each case, two alternative modes of repressor binding are diagrammed; one predicted by the traditional alignment scheme (above) and the other by the alignment suggested by Staacke *et al.* (1990a) (below). Autoradiographs of the gel retardation analysis are below. The designation of the stoichiometry as '1:1' or '2:1' indicates the number of dimeric repressors bound to one DNA fragment.

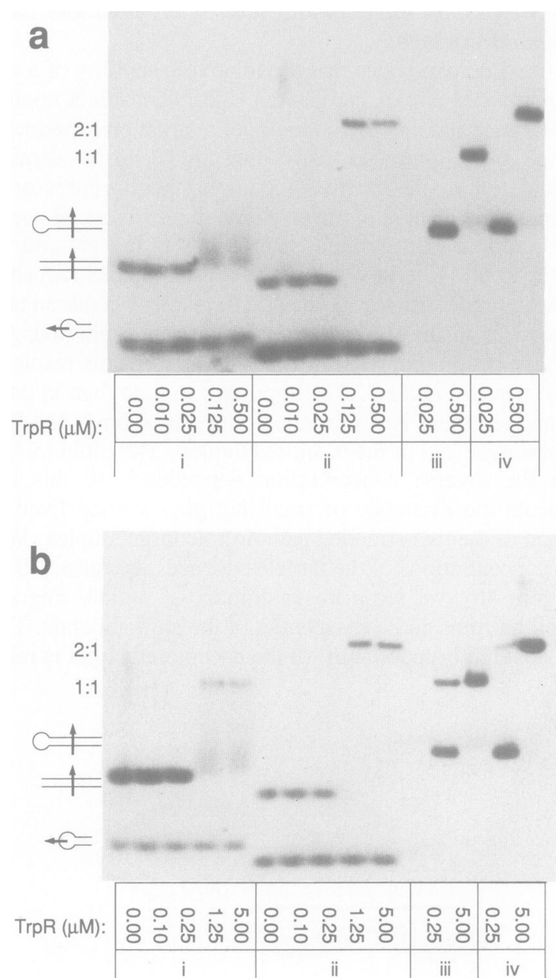


Fig. 4. Gel retardation analysis of short two-stranded duplexes. **(a)** 10 μM DNA **(b)** 100 μM DNA. Panels labeled (i) show gel retardation of the 18 bp *trp* operator used in the crystal structure. Panels labeled (ii) show gel retardation of the same β-centered 14mer used by Staacke *et al.* (1990a). Panels (iii) and (iv) show control experiments with 21 bp, hairpin stems (described in Figure 3) containing α-centered and β-centered targets respectively.

mainly to the use of short two-stranded targets and the failure of the complex with *trp* repressor to maintain its integrity in the gel.

Panel (ii) in Figure 4a shows the shift caused by the binding of *trp* repressor to the symmetrical 14 bp two-stranded duplex centered on a β-dyad. In contrast to the α-centered two-stranded operator fragment which binds one repressor and forms a kinetically unstable complex, the β-centered 14 bp target of Staacke *et al.* (1990a) binds two repressors. The binding of two repressors to β-centered sequences was noted earlier by Haran *et al.* (1991) and more recently by Carey *et al.* (1991). In addition to producing a dramatic shift, the binding of two repressors gives a relatively sharp band, indicating greater kinetic stability during electrophoretic migration of the complex. The binding of two dimeric *trp* repressors to a 14 bp target seriously calls into question a model in which the repressor straddles a β-dyad when it binds to a single β-centered operator. If one were unaware of the 2:1 stoichiometry of the β-centered complexes [e.g. panel (ii) of Figure 4a and b] and the kinetic instability of the α-centered complexes [panel (i) of Figure 4a,b] one would reasonably but mistakenly conclude that the preferred target was the β-centered sequence. The results of Figure 4 demonstrate that a gel retardation analysis of site-specific complexes containing the *trp* repressor and short two-stranded DNA targets may exhibit anomalies of stability and stoichiometry that can compromise its value in determining target preferences for the *trp* repressor (Figure 3 of Staacke *et al.*, 1990a; Figure 2 of Staacke *et al.*, 1990b).

Why is the 1:1 complex formed with an 18 bp, two-stranded complex less stable during electrophoresis than the 2:1 complex containing an even smaller two-stranded duplex? Our understanding of the kinetic behavior of protein–DNA complexes during gel electrophoresis is limited at best (Garner and Revzin, 1986; Crothers, 1987), but a plausible reason for this seeming paradox can be derived from the crystal structure of the complex (Otwinowski *et al.*, 1988) namely, that there are 10 extra protein–phosphate group contacts provided by the 'two-repressor' binding mode. These interactions may slow the dissociation rate of the complex in the gel. Thus, short two-stranded targets especially symmetrical ones, introduce multicomponent equilibria (before loading and during electrophoresis) that can render their use in band-shift analyses ineffective and even misleading as a means of defining high affinity targets for site-specific DNA-binding proteins. This topic is treated in more detail elsewhere (T.E.Haran *et al.*, in preparation) in connection with the use of hairpin constructions such as those in Figure 3.

The alignment of trp repressor on the trpEDCBA promoter

Gel retardation and filter-binding techniques were used to see if the *trp* repressor's preferential affinity for the traditional *trp* operator sequence in short duplex constructions was consistent with the pattern of *in vitro* binding to the *trpEDCBA* promoter–operator fragment. The radioactive probe was the same one introduced for filter-binding analysis by Hurlburt and Yanofsky (1990), namely, a 43 bp *trpEDCBA* promoter–operator fragment whose 3′-terminal residues are filled in at both ends by the Klenow fragment of *E.coli* DNA polymerase I and [α-^{32}P]dATP. The stoichiometry of complexes in the retarded bands was established with ^{32}P-labeled DNA and ^{125}I-labeled protein of known specific radioactivity (Figure 6, and T.E.Haran *et al.*, in preparation).

Figure 5a shows the 43 bp wild-type *trp* repressor fragment and two sequence variants designed to abolish all alternative binding sequences and to leave unaltered only a 18 bp segment centered either on dyads β_L or β_R. The pattern of shifted bands shows that the wild-type *trp* promoter fragment has two high affinity sites (Figure 5b). One site binds on repressor (1.02 $\pm$ 0.04) at very low repressor concentrations and at slightly higher concentrations the very stable 1:1 complex is replaced by an equally, if not more, stable 2:1 (2.00 $\pm$ 0.06) complex. At much higher protein concentration, a third repressor (3.1 $\pm$ 0.4) is bound. The promoter with the single intact β_R-centered sequence does not show the 1:1 complex that would be expected if it were the preferred target (Figure 5c). Like the short β-centered duplexes, the intact β_R sequence binds two repressors (2.02 $\pm$ 0.06) with lower affinity and higher cooperativity. The unaltered β_L segment forms an even weaker 2:1

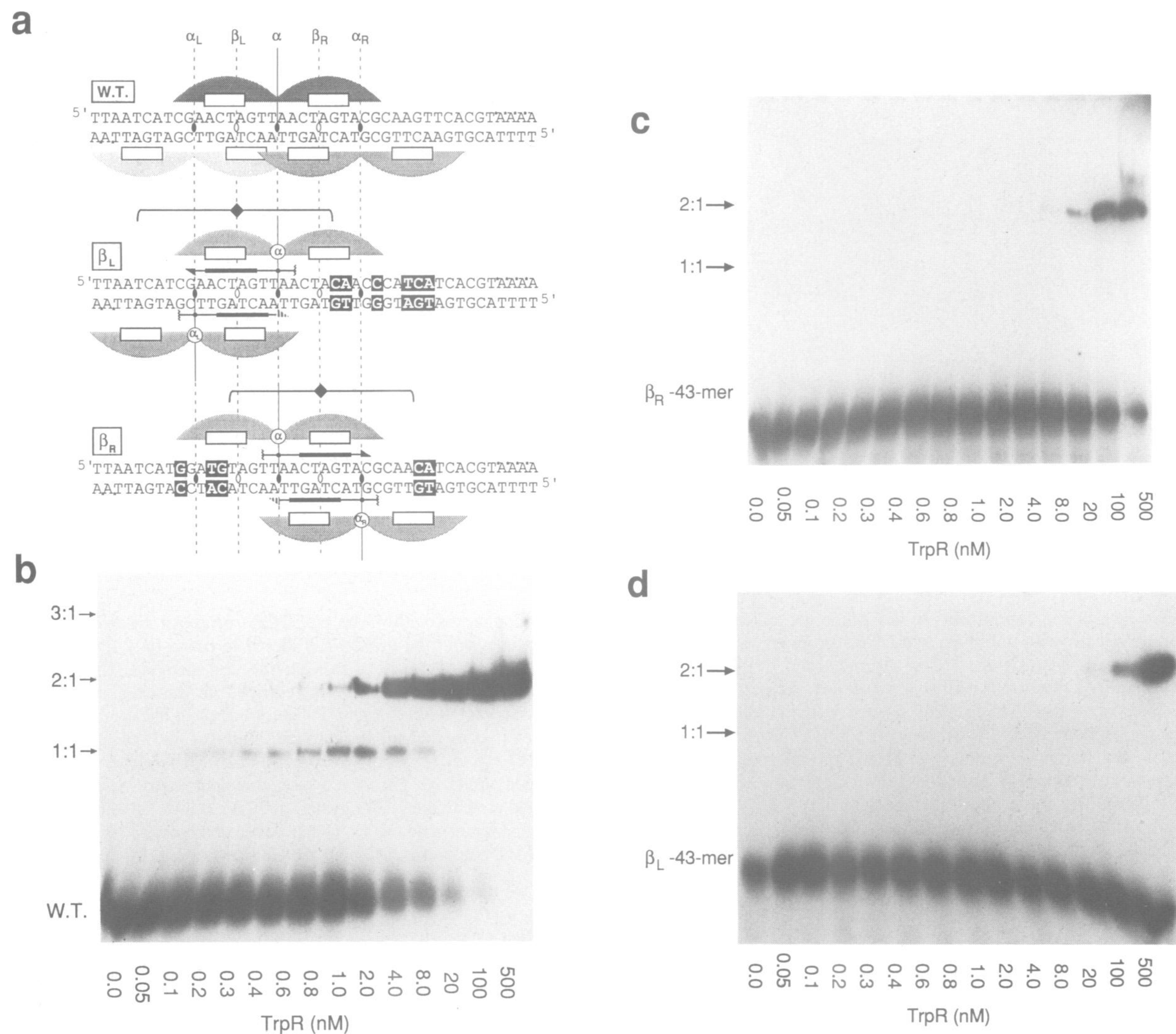


Fig. 5. Pattern of *trp* repressor binding to 43 bp *trpEDCBA* promoter fragments. (**a**) Wild type sequence (top) and two sequence variants modified to preserve only the left-most β-centered (β_L, middle) and only the right-most β-centered sequence (β_R, bottom). Gel retardation of wild-type fragment (**b**), β_R-centered target (**c**), β_L-centered target (**d**). Stars indicate sites of ^{32}P-labeled nucleotides introduced by filling in 3'-recessed ends with [α-^{32}P]dATP. DNA concentration was <0.1 nM in all reactions.

complex (Figure 5d). The pattern of binding shown in Figure 5b and c can explain with an idea introduced by Haydock *et al.* (1983), namely, that the symmetrical dimeric *trp* repressor target is best considered as two half sites centered around an α-dyad. The ideal high-affinity target is composed of two half sites arranged as inverted repeats centered on α, but a sequence containing only a single half site will also support binding, albeit more weakly. These half sites are indicated by the horizontal arrows in Figures 2, 3 and 5. When the band shifts of the three sequences of Figure 5 are interpreted in terms of α-centered half sites, we can see that the wild-type sequence provides three targets: two strong sites (centered on α and α_R) each having two perfect or nearly perfect half sites and a third binding site (centered on α_L) that is noticeably weaker because it has only half-site sequence. The altered 43 bp promoter fragments containing intact 18 bp sequences around either β_R or β_L (Figure 5a) provide two overlapping binding sites each composed of only one intact (or nearly intact) half-site sequence. Thus, two repressors bind more weakly. If β-centered sequences were the preferred target, one would expect an isolated β-centered 18 bp operator to form a strong 1:1 complex.

Extreme cooperativity is seen when two repressors bind to a single β-centered target; indeed, the binding of one repressor is never observed. Figures 3b and 5a explain this stoichiometry in terms of two repressors binding the overlapping half sites of two α-centered targets. Based on the crystal structure of Otwinowski *et al.* (1988) modelling indicates the possibility of positive protein–protein interactions between the reading heads of tandemly bound repressors. This interaction may account for the noticeable cooperativity in tandem binding to the wild-type *trpEDCBA* promoter as well (Figure 5b). Perhaps, when *trp* repressor is bound in tandem to isolated β-centered sequences (either in short duplexes or in the modified *trpEDCBA* promoter), the weak non-specific interactions between the outer regions of the repressor and non-cognate flanking DNA elements permits readjustments in the region of protein overlap that enhance this cooperative interaction.

Filter binding of the 43 bp duplex is compatible with the band-shift results (Figure 7a). The wild-type promoter

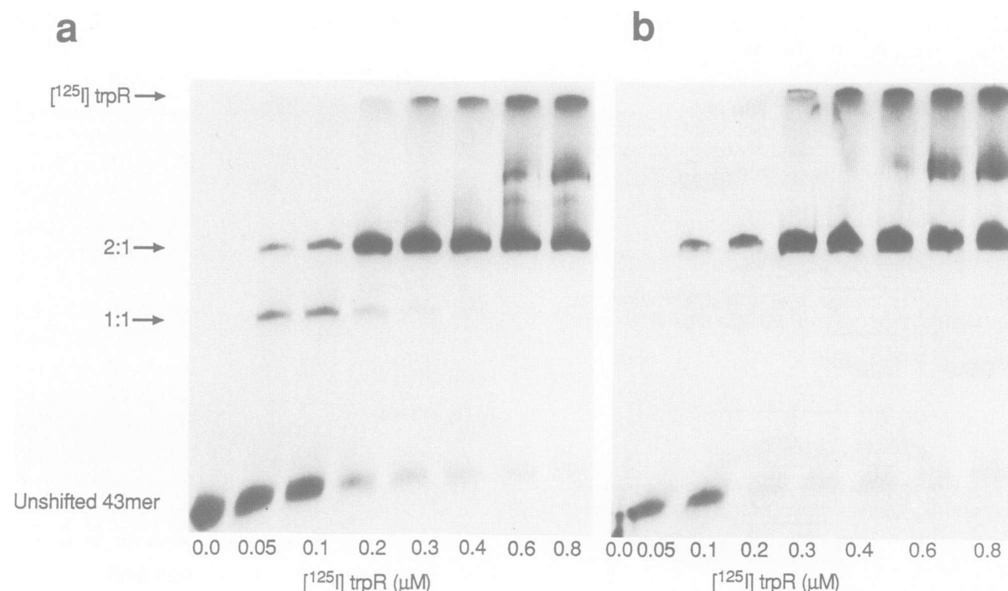


Fig. 6. Double-labeling experiments to determine the stoichiometry of *trp* repressor complexes with 43 bp *trpEDCBA* promoter fragments. (**a**) *trp* repressor complexes with wild-type *trpEDCBA* promoter and (**b**) *trp* repressor complexes with the *trpEDCBA* altered to retain only the β_R sequence. Gel electrophoresis analysis of [^{125}I]*trp* repressor and [^{32}P]DNA fragments was carried out as described in Materials and methods. The DNA concentration in all reaction mixtures was 0.1 μM. The ratio of protein:DNA in each band was quantified as described in Materials and methods. For example, in (a) the '2:1' complex formed in the lane containing 0.1 μM repressor has 1949 c.p.m. (^{32}P; uncorrected), 2787 c.p.m. (^{32}P; corrected for background and quenching), 2176 (^{125}I uncorrected), 22 835 c.p.m. (^{125}I; corrected for background, spillover and quenching). Specific activities for the starting materials was 22 201 c.p.m./pmol for [^{32}P]DNA and 90 247 c.p.m./pmol for [^{125}I]*trp* repressor. Therefore, the molar ratio of repressor dimers:operator duplex is 2.02 for this lane. All other bands were treated similarly. Thus the averaged stoichiometry designed as '2:1' is 2.00 ± 0.06 and for those designated as '1:1' it is 1.02 ± 0.04. In (b) the averaged stoichiometry designated as '2:1' is 2.02 ± 0.06. Analyses include all visible bands but do not include data from the right-most lane showing significant amounts of higher order complexes (i.e. >2:1). This is because 'leakage' of [^{32}P]DNA from the weak complexes contaminates the 2:1 and 1:1 complexes. A more detailed description of the methodology is described elsewhere (T.E.Haran *et al.*, in preparation).

sequence appears to bind two repressors strongly, and a third repressor is bound more weakly. The modified promoters show weaker double-site binding. As expected, the weakest binding site is the region that remains unaltered around β_L.

In order to confirm which sequence segments within the promoter are responsible for affinity in the filter-binding assay, we carried out competition experiments. The radioactive probe was the wild-type 43 bp *trp* promoter fragment and the competing targets were unlabeled 21 bp oligoduplexes produced by hairpins identical to those shown in Figure 3. As expected for an α-centered model and consistent with the band-shift and direct filter-binding experiments, the hairpin whose stem is the α-centered traditional operator sequence competes most effectively, indeed, as well as an unlabeled 43 bp probe. This suggests that the loop of this hairpin construct does not influence affinity significantly. Hairpins with β-centered sequences compete poorly (Figure 7b); the idealized version (Figure 3b) binds 17 times more weakly than the α-centered sequence (Figure 3a) and the construct containing β_L-centered sequence (Figure 5a) competes about 1100 times more weakly.

Discussion

The principal bias in designing a crystallographic study of a specific protein–DNA complex is the choice of the macromolecular partners. For the study of the *trp* repressor–operator complex, the choice of the target DNA sequence was based mainly on the sequence consensus exhibited in the three *E.coli trp* promoters/operators (and since supported by the sequence of the *mtr* operator; Heatwole and Somerville, 1991) and, secondarily, on the distribution of operator-constitutive mutations (Bennett and Yanofsky, 1978). This was confirmed by the systematic mutagenesis studies of Bass *et al.* (1987), and the biochemical analyses of Carey (1988, 1989); however, any choice carries the limitations inherent in substituting a short linear DNA segment for a target of the same sequence in the context of the naturally supercoiled bacterial genome. The refined crystal structure of the complex also agreed perfectly with mutational analysis of both the DNA (Bennett and Yanofsky, 1978; Bass *et al.*, 1987) and the protein (Kelley and Yanofsky, 1982; Bass *et al.*, 1988), although there is a minor disagreement in the structural interpretation of the genetic results (Bass *et al.*, 1988; Luisi and Sigler, 1990). Moreover, the crystal structure was compatible with chemical protection studies (Oppenheim *et al.*, 1980; Kumamoto *et al.*, 1987) and showed that the tandem binding of repressor to overlapping operator sites was stereochemically permitted. This overlapping binding pattern has since been shown to be also required in the repression of methionine-biosynthetic operons by the MetJ protein (Phillips, 1991). The chemical details of the protein–DNA interface seen in the crystal structure of the *trp* repressor–operator complex did not, however, agree with the prevailing paradigm. The choice of the operator sequence was so forcefully challenged (Staacke *et al.*, 1990a), that the relevance of the cystallographic result has been queried (Kissinger *et al.*, 1990; Brennan, 1991). Where the results

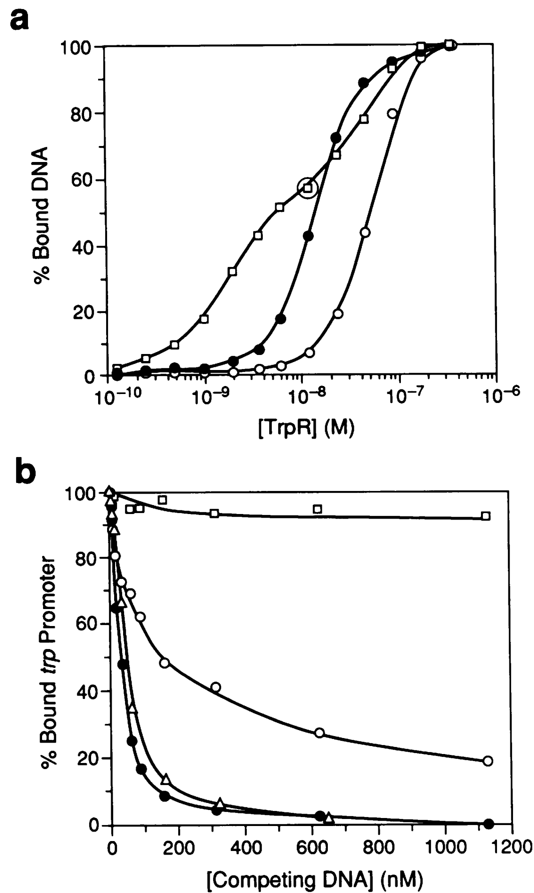


Fig. 7. Filter-binding analysis of *trp* repressor to 43 bp *trpEDCBA* promoter fragments. (**a**) Repressor-mediated binding of labeled fragments to the filter. At low repressor concentration when the DNA is largely free of bound protein, the *trp* repressor binds more strongly to the wild-type sequence (□) than to either of the fragments that were altered to preserve either the β_R-centered (●) or the β_L-centered sequence (○). At higher concentrations where repressor binds largely to DNA targets already bound by protein, the wild-type sequence exhibits multiple equilibria while the β_R- and β_L-centered sequences show the cooperativity discussed in the text. (**b**) Competition for *trp* repressor binding between the labeled 43 bp *trpEDCBA* promoter fragment and unlabeled short sequences in the stems of hairpin constructs similar to those used in Figure 3. The mixture being challenged with competing oligonucleotides is that used in the experiment designated by ⊚ in (a). The traditional α-centered operator sequence (●) competes as well as the unlabeled wild-type 43 bp probe (△) and significantly better than the symmetrized β-centered sequence (○). A symmetrized version of the β_L-centered sequence defined in Figure 5a competes poorly (□).

of detailed structural studies contradict the 'conventional wisdom' it is especially important to establish the validity and relevance of the structure; but, it is equally important to scrutinize the studies that challenge the structural work and to identify potential sources of error. For this reason, we have carried out experiments to determine the true target for the *trp* repressor.

We have shown that the mutational analysis of Staacke *et al.* (1990a) are consistent with previous mutagenic studies of the operator and the consensus of *E.coli* trpR-repressible promoters, that is, there is general agreement as to which base pairs contribute most to specificity. However, the alignment of *trp* repressor on the *trpEDCBA* promoter–operator proposed by the Köln group appears stereochemically implausible because it positions the experimentally defined recognition apparatus of the repressor so that it cannot contact most of the operator's identity elements.

The challenge to the conventional choice of the *trp* operator sequence lies in a gel retardation analysis carried out with symmetrical duplex targets between 14 and 18 bp long. By using radiolabeled protein, we have shown that the shift of the 14 bp β-centered target proposed by Staacke *et al.* (1990) involves a complex containing two dimeric repressors where their model predicts only one repressor molecule straddling the β-dyad. By contrast, short α-centered duplexes are bound by one repressor as predicted by the traditional model. Recent gel retardation studies by Carey *et al.* (1991) also show that the *trp* repressor binds an α-centered duplex (20 bp) fifty times more strongly than a β-center duplex (18 bp). It remains to be shown what component of this preference is due to the sequence *per se* and what is due to the difference in length (20 bp versus 18 bp). We have also repeated the experiments of Staacke *et al.* (1990) using their experimental conditions and DNA constructs, and have shown that the 1:1 complexes formed with short two-stranded α-centered targets are kinetically unstable in the gel during the course of the electrophoresis, whereas the 2:1 complexes formed on the β-centered target are not. This leads to the immediate impression that α-centered duplexes are a poorer target. We have shown this to be a misleading artifact of the analytical method by using a construction in which the target duplex is stabilized in the stem of a hairpin. Once the kinetic instability is removed, a clear preference (at least 17-fold) is shown for α-centered targets.

The conclusions drawn from experiments with short hairpin-stabilized sequences are consistent with gel retardation analyses and filter-binding studies using wild-type and modified 43 bp *trpEDCBA* promoter segments. Thus, the traditional operator sequence used in the crystal structure of the *trp* repressor–operator complex, is the preferred target for *trp* repressor.

Materials and methods

Purification and labeling of trp repressor and DNA fragments

trp repressor was overexpressed in and purified from an overproducing strain (pET13atrpr3/BL21DE3) kindly provided by Dr C.Lawson (Brookhaven National Laboratory). The purification protocol was the same as described Joachimiak *et al.* (1987). Repressor preparations were shown to be electrophoretically homogeneous and fully active in DNA binding assays.

Pure *trp* repressor (6 nmol) was labeled with $\sim {}^{125}$I[NaI] (~6000 Ci/mmol) and Iodogen (Pierce) according to the manufacturer's instructions. Free iodine was removed by gel filtration and dialysis and the radiolabeled protein was shown by isotopic dilution to be as active in a specific DNA binding as unlabeled repressor (T.E.Haran *et al.*, in preparation). Two percent of the repressor's subunits were labeled and virtually all of the label was incorporated in Tyr7 as determined by N-terminal sequence analysis. The specific radioactivity of the labeled protein was established by liquid scintillation counting and UV spectroscopy.

All deoxynucleotides were synthesized on an automated DNA synthesizer (Applied Biosystems). The oligonucleotides were purified by electrophoresis on polyacrylamide gels containing 7 M urea. DNA fragments were eluted from the gel and ethanol precipitated.

Oligonucleotides were labeled with ^{32}P in two ways: (i) the 5′ terminus of the hairpin duplexes (Figure 3) and short double-stranded duplexes (Figure 4) was labeled with [γ-^{32}P]ATP (~5000 Ci/mmol) and T4 polynucleotide kinase, and purified by polyacrylamide gel electrophoresis in the presence of 7 M urea. (ii) 3′-recessed ends of *trpEDCBA* promoter fragments (Figure 5) were filled in with [α-^{32}P]dATP (~6000 Ci/mmol), dTTP and the Klenow fragment of DNA polymerase I. DNA duplexes were subsequently purified on native polyacrylamide gels.

Gel retardation assay

The gels used in Figures 3 and 5 were native 16% polyacrylamide gels (acrylamide/bisacrylamide ratio = 75:1) cast and run in 10 mM phosphate buffer (pH 6.5) and 0.25 mM L-Trp. The gels were prerun at 12.5 V/cm for 1.5 h, and run at 25 V/cm for 3 h, with recirculation of the buffer. The gel temperature was kept at 18°C by using constant temperature gel apparatus (Hoefer) and a circulating refrigerated water bath (VWR). Ten microliter samples were incubated for 30 min at room temperature. The buffer conditions were 0.2 nM DNA, 10 mM phosphate buffer (pH 6.5), 0.5 mM L-Trp, 250 mM NaCl and 0.1% NP40. Samples were loaded while the gels were running.

The 5′ terminus of the DNA in Figure 4 was labeled with [α-^{32}P]dATP and polynucleotide kinase as described above, and subsequently mixed with a large excess of unlabeled DNA of the same sequence. The final concentration of DNA was 10 μM in (a) and 100 μM in (b). The experimental conditions for gel electrophoresis shown in Figure 4 were identical to those of Staacke *et al.* (1990a).

Double labeling experiments

[^{125}I]*trp* repressor was labeled and purified as described above. [^{32}P]DNA was labeled and purified as described earlier and subsequently mixed with a 150-fold molar excess of DNA of known concentration. In this way, one ensures that the specific radioactivity and concentration of the DNA samples is accurately known. Gel retardation analysis of [^{125}I]*trp* repressor and [^{32}P]DNA was carried out as described above. After electrophoresis the gels were placed on to nitrocellulose filters and vacuum dried. The bands were localized by autoradiography, excised together with the nitrocellulose backing and counted by liquid scintillation. The apparent specific activities of both ^{32}P and ^{125}I are diminished significantly in bands excised from electrophoresis experiments presumably because of unavoidable loss of material and fluorescence quenching by components in the gel. Therefore, samples of [^{32}P]DNA and [^{125}I]*trp* repressor; ranging respectively in concentration from 0.2 to 8 pmol and 0.5 to 8 pmol were electrophoresed in separate lanes, autoradiographed, excised, dried and counted as described above. Identical aliquots were counted directly and compared with their electrophoresed counterparts. [^{32}P]DNA and [^{125}I]*trp* repressor, retrieved from gels were corrected for 'quenching' by the factors 1.43 ± 0.05 and 11.0 ± 0.9 respectively. The effect is linear for both isotopes over the concentrations used in these experiments and the appropriate corrections were applied to all 'stoichiometry' experiments. Energy discrimination for ^{125}I and ^{32}P was optimized to minimize cross contamination of counting. The window for ^{125}I counting was selected between 0 and 36 keV, and that of ^{32}P between 50 and 1700 keV to minimize spillover. The spillover of ^{32}P to the ^{125}I window was 4.16 ± 0.06%. There was essentially no spillover of ^{125}I into the ^{32}P window.

Filter binding assay

Nitrocellulose filter-binding assay was performed as described by Hurlburt and Yanofsky (1990) using the same 43 bp *trpEDCBA* promoter fragment. The radioactive probe was purified and labeled as described above but with [α-^{35}S]dATP (~600 Ci/mmol). In competition experiments, the mixture of *trp* repressor and [^{35}S]*trpEDCBA* promoter fragment was challenged with increasing concentrations of competing non-radioactive oligonucleotides. These include the 43 bp probe and the short hairpin constructs described in Figure 3.

Acknowledgements

We thank Drs Donald Crothers and Thomas Steitz of Yale, and Charles Yanofsky of Stanford for helpful discussions; Mr Kyu Rhee for help with the filter-binding experiments; Mr Mark Saba for expert help with the figures; and the NIH for financial support (grant GM15225).

References

Arrowsmith,C.H., Pachter,R., Altman,R.B., Iyer,S.B. and Jardetzky,O. (1990) *Biochemistry*, **29**, 6332–6341.

Arrowsmith,C.H., Czaplicki,J., Iyer,S.B. and Jardetzky,O. (1991) *J. Am. Chem. Soc.*, **113**, 4020–4022.

Bass,S., Sugiono,P., Arvidson,D.N., Gunsalus,R.P. and Youderian,P. (1987) *Genes Dev.*, **1**, 565–572.

Bass,S., Sorrells,V. and Youderian,P. (1988) *Science*, **242**, 240–245.

Bennett,G.N. and Yanofsky,C. (1978) *J. Mol. Biol.*, **121**, 179–192.

Brennan,R.G. (1991) *Curr. Opin. Struct. Biol.*, **1**, 80–88.

Brennan,R.G. and Matthews,B.W. (1989) *J. Biol. Chem.*, **264**, 1903–1906.

Carey,J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 975–979.

Carey,J. (1989) *J. Biol. Chem.*, **264**, 1941–1945.

Carey,J., Lewis,D.E.A., Lavoie,T.A. and Yang,J.J. (1991) *J. Biol. Chem.*, **266**, 24509–24513.

Chu,B.C. and Orgel,L.E. (1991) *Nucleic Acids Res.*, **19**, 6958.

Crothers,D.M. (1987) *Nature*, **327**, 464–465.

Garner,M.M. and Revzin,A. (1986) *Trends Biochem. Sci.*, **11**, 395–396.

Haran,T.E., Joachimiak,A. and Sigler,P.B. (1991) *J. Biomol. Struct. Dyn.*, **8**, a276–a277.

Haydock,P.V., Bogosian,G., Brechling,K. and Somerville,R.D. (1983) *J. Mol. Biol.*, **170**, 1019–1030.

Heatwole,V.N. and Somerville,R.L. (1991) *J. Bacteriol.*, **173**, 3601–3604.

Hurlburt,B.K. and Yanofsky,C. (1990) *J. Biol. Chem.*, **265**, 7853–7858.

Joachimiak,A., Kelley,R.L., Gunsalus,R.P., Yanofsky,C. and Sigler,P.B. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 668–672.

Joachimiak,A., Marmorstein,R.Q., Schevitz,R.W., Mandecki,W., Fox,J.L. and Sigler,P.B. (1987) *J. Biol. Chem.*, **262**, 4917–4921.

Kelley,R.L. and Yanofsky,C. (1982) *Proc. Natl. Acad. Sci. USA*, **82**, 483–487.

Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.

Klig,L.S., Crawford,I.P. and Yanofsky,C. (1987) *Nucleic Acids Res.*, **15**, 5339–5351.

Kumamoto,A.A., Miller,W.G. and Gunsalus,P. (1987) *Genes Dev.*, **1**, 556–564.

Lawson,C.L., Zhang,R.-G., Schevitz,R.W., Otwinowski,Z., Joachimiak,A. and Sigler,P.B. (1988) *Proteins*, **3**, 18–31.

Luisi,B.F. and Sigler,P.B. (1990) *Biochem. Biophys. Acta*, **1048**, 113–126.

Marmorstein,R.Q. and Sigler,P.B. (1989) *J. Biol. Chem.*, **264**, 9149–9154.

Oppenhein,D.S., Bennett,G.N. and Yanofsky,C. (1980) *J. Mol. Biol.*, **144**, 132–142.

Otwinowski,Z., Schevitz,R.W., Zhang,R.-G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) *Nature*, **335**, 321–329.

Phillips,S.E.V. (1991) *Curr. Opin. Struct. Biol.*, **1**, 89–98.

Ptashne,M.A. (1986) *Genetic Switch.* Cell Press, Cambridge, MA.

Schevitz,R.W., Otwinowski,Z., Joachimiak,A., Lawson,C.L. and Sigler,P.B. (1985) *Nature*, **317**, 782–786.

Staacke,D., Walter,B., Kisters-Woike,B., Wilcken-Bergmann,V.v. and Müller-Hill,B. (1990a) *EMBO J.*, **9**, 1963–1967.

Staacke,D., Walter,B., Kisters-Woike,B., Wilcken-Bergmann,V.v. and Müller-Hill,B. (1990b) *EMBO J.*, **9**, 3023 (corrigendum).

von Hippel,P.H. and Berg,O.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1608–1612.

Youderian,P., Vershon,A., Bouvier,S., Sauer,R.T. and Susskind,M.M. (1983) *Cell*, **35**, 777–783.

Zhang,R.-G., Joachimiak,A., Lawson,C.L., Schevitz,R.W., Otwinowski,Z. and Sigler,P.B. (1987) *Nature*, **327**, 591–597.

Received on February 25, 1992; revised on May 4, 1992