Stryer, L. (1991) *J. Biol. Chem. 266*, 10711–10714.
Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212–4228.
Wensel, T., & Stryer, L. (1986) *Proteins: Struct., Funct., Genet. 1*, 90–99.
Wensel, T., & Stryer, L. (1990) *Biochemistry 29*, 2155–2161.
Yamazaki, A., Hayashi, F., Tatsumi, M., Bitensky, M. W., & George, J. S. (1990) *J. Biol. Chem. 265*, 11539–11548.
Yau, K.-W., & Baylor, D. (1989) *Annu. Rev. Neurosci. 12*, 289.
Zweig, M., Showalter, S., DuBois, G. C., Sisk, W. P., & Court, D. L. (1987) *Gene 55*, 47–53.

Interactions between the *trp* Repressor and Its Operator Sequence As Studied by Base Analogue Substitution†

Joan M. Mazzarelli,‡ Sharanabasava B. Rajur, Patricia L. Iadarola, and Larry W. McLaughlin*
Department of Chemistry, 140 Commonwealth Avenue, Boston College, Chestnut Hill, Massachusetts 01267
Received January 28, 1992; Revised Manuscript Received April 6, 1992

ABSTRACT: A series of modified *trp* operator sequences has been prepared by the incorporation of seven different base analogues. Four of the analogues allow the site-specific deletion of functional groups present on the dA-dT and dT-dA base pairs at positions −4/+4 and −5/+5 in the *trp* operator. The remaining three analogues permit the incorporation of structural analogues of the native dA-dT or dG-dC base pairs. The duplex operator sequences all exhibit T_m values well above ambient temperature (48–70 °C), and these values generally correlate very well with the number of interstrand hydrogen bonds present. The affinity between the *trp* repressor and 14 modified operator sequences was examined using a recently developed alkaline phosphatase protection assay. The results from the analogue sequences used in this study suggest that the structure of the dA-dT or dT-dA base pairs at positions −4/+4 and −5/+5, respectively, has relatively little effect upon the solution binding by the *trp* repressor, but the protein is very sensitive to the orientation of the amino and carbonyl functional groups at the −4/+4 positions, which are involved in the formation of an interbase hydrogen bond present in the major groove. (The term structure in this case refers to the hydrogen bonding structure of the base pairs. We recognize that the introduction of conservative functional group deletions or reversals may affect other structural criteria such as hydration.) The deletion of individual functional groups from the operator sequence suggests that the carbonyl at dT_{+4} is critical for formation of the high-affinity sequence-specific complex. Additionally, the thymine methyl group at dT_{+4} and the N7 nitrogen of dA_{+5} appear to be critical contacts necessary for high-affinity binding by the repressor. The thymine carbonyl and the adenine N7 nitrogen are each responsible for approximately −1.5 kcal/mol of apparent free energy of binding. The thymine methyl provides a somewhat smaller contribution of −0.7 kcal/mol. Deletion of either of the adenine amino groups at dA_{-4} or dA_{+5} results in a sequence that binds to the repressor with a higher affinity than observed with the native sequence; this can be explained in that the functional groups lost are not critical for binding, and the resulting increased flexibility of the operator, or the creation of a more hydrophobic surface at these sites, enhances van der Waals contacts between the protein and the nucleic acid.

The formation of high-affinity sequence-specific recognition complexes between proteins and nucleic acids can be explained most easily in terms of multidentate ionic interactions, hydrogen bonding interactions, and van der Waals contacts between the amino acid side chains of the protein and the base, carbohydrate, and phosphate residues of the nucleic acid (Seeman et al., 1976; von Hippel & Berg, 1986). While some protein–DNA contacts contribute to the overall binding affinity between the two macromolecules, it is the direct or indirect readout of the bases which allows the protein to discriminate correct from incorrect base sequences. With direct sequence readout, the protein appears to contact specific functional groups present on the base residues, and a target sequence will ideally contain a unique set of functional groups available for such interactions. Binding at the correct or incorrect sequence will depend both on the binding energy realized from the formation of specific interactions and on the potential binding energy lost as the result of steric interactions that preclude the formation of such bonds. Reports on the structure of a number of sequence-specific protein–DNA complexes, including bacterial phage repressor proteins (Anderson et al., 1987; Aggarwal et al., 1988; Jordan & Pabo, 1988; Wolberger et al., 1988) and the *Eco*RI restriction endonuclease (McClarin et al., 1986; Kim et al., 1990; Rosenberg, 1991), using X-ray diffraction methods have generally confirmed the existence of the direct readout mechanism. Indirect readout of base sequences has been suggested as an alternative mechanism by which proteins can form high-affinity complexes (Otwinowski et al., 1988). In this case, the functional groups present on the base residues will contribute to the overall geometry of the recognition site and may be important in allowing the intimate association of the protein with the nucleic acid, but most of the direct contacts occur between the amino acid side chains of the protein and the phosphodiester residues of the DNA.

† This work has been supported by a grant to L.W.M. from the National Science Foundation (DMB-8904306). L.W.M. is the recipient of an American Cancer Society Faculty Research Award (FRA-384).
* Author to whom correspondence should be addressed.
‡ Present address: Department of Molecular Biophysics and Biochemistry, School of Medicine, Yale University, New Haven, CT.

0006-2960/92/0431-5925$03.00/0

The reported structure of the *trp* repressor with its operator site (Otwinowski et al., 1988) best represents the indirect readout mechanism. Although the X-ray crystallographic structure of this repressor–operator complex has been reported, there has been some controversy about the nature of the operator site recognized by the repressor and whether the reported crystal structure represents a specific or a nonspecific protein–DNA complex (Brennan & Matthews, 1989; Staacke et al., 1990; Benson et al., 1992).

The *trp* repressor is composed of two identical subunits of 107 amino acids (Gunsalus & Yanofsky, 1980; Joachimiak et al., 1983) and binds with dyad symmetry (Otwinowski et al., 1988). The subunits are formed from α-helices, and in common with many other bacterial repressors [for reviews see Olendorf and Matthews (1983) and Pabo and Sauer, 1984], the *trp* repressor contains a helix–turn–helix structural motif which is primarily responsible for interacting with the operator site (Otwinowski et al., 1988). The amino acid tryptophan functions as a co-repressor (Joachimiak et al., 1983), which upon binding to the protein induces a conformational change and increases binding affinity for the operator by 10000-fold (Marmorstein et al., 1991). Bass et al. (1987) have used an in vivo assay to study the binding of the *trp* repressor to a series of operators containing symmetrical base changes. Their work suggests that the five nucleotides, 5′-ACTAG-3′ including base pairs −7/+7 through −3/+3 are most important for the formation of the high-affinity complex between the repressor and operator.

The crystal structure of the protein–DNA complex at 2.4 Å has been reported (Otwinowski et al., 1988) using the following self-complementary 19-mer:

```
     -9   -5   -1 1    5    9
5'-TGTACTAGTTAACTAGTAC
   CATGATCAATTGATCATGT-5'
    9    5    1-1   -5   -9
```

The crystal structure indicates the presence of only a single direct amino acid–nucleoside base contact: a bidentate interaction between Arg_{69} and G_{-9}. However, there are six major groove contacts to functional groups on the bases in each operator half-site mediated by three water molecules, and there are 12 direct contacts involving the unesterified oxygens of six phosphate residues. The complex may be stabilized additionally by extensive nonspecific van der Waals contacts at the protein–DNA interface that largely exclude solvent water molecules.

It has been suggested that the observed specificity of the *trp* repressor for its target operator is largely the result of the structure of the operator itself (indirect readout) rather than the result of interactions to the bases (direct readout) (Otwinowski et al., 1988; Luisi & Sigler, 1990). The DNA operator in the complex is bent, principally at the −5/−4 and +4/+5 steps, where there is large positive roll (~11°) that compresses the major groove and opens up the minor groove, and there is evidence to suggest that d(TA) steps, particularly in the sequence d(CTAG), tend to deviate from the classical B-form DNA structure with little if any induced strain (Hunter et al., 1989). The presence of such sequence-mediated structure modulation in B-form DNA in the area of contact between the *trp* repressor and its operator site has been suggested to be primarily responsible for the formation of the high-affinity recognition complex in the presence of the 10^6–10^7 other possible binding sites in the prokaryotic genome (Otwinowski et al., 1988; Luisi & Sigler, 1990).

The incorporation of modified nucleobases into short chemically synthesized oligodeoxynucleotides can be a useful approach for the study of recognition processes in that the functional groups of the base residues can be systematically altered [for a review see Aiken and Gumport (1991)]. The presence or absence of a given functional group at a specific location will vary in these sequences, and corresponding structure modulation of the DNA recognition site can be expected in some cases. Structure alteration will be most pronounced when functional groups involved in interstrand hydrogen bonding have been deleted, or in cases where additional functional groups have been incorporated into the sequence. This approach has been employed in studies with other repressor proteins (Goeddel et al., 1977, 1978; Caruthers, 1985) as well as restriction endonucleases (Dwyer-Hallquist et al., 1982; Ono et al., 1984; Yolov et al., 1985; Brennan et al., 1986a, Jiricny, et al., 1986; Seela & Driller, 1986; Seela & Kehne, 1987; McLaughlin et al., 1987; Nwosu et al., 1988; Mazzarelli et al., 1989; Lesser et al., 1990; Newman et al., 1990a,b; Aiken et al., 1991) modification enzymes (Brennan et al., 1986b), and RNA polymerase (Dubendorff et al., 1987).

We have prepared a series of 14 modified *trp* operator sequences in which we have altered the nature and position of functional groups present in the major groove at the −4/+4 and the −5/+5 base pairs. The structure of the dA-dT and dT-dA base pairs at these positions has been suggested to account for the high-affinity binding of the *trp* repressor to its operator site. We have examined the binding affinities of the *trp* repressor to these sequences using a recently developed kinetic assay employing alkaline phosphatase (Marmorstein et al., 1991).

Experimental Procedures

Materials

Thin layer chromatography (TLC) was performed on 5 × 10 cm silica gel $60F_{254}$ glass-backed plates (E. Merck, Darmstadt, Germany). For flash chromatography, silica gel 60 (particle size less than 0.063 mm; E. Merck, Darmstadt, Germany) was used under positive pressure from nitrogen gas. High-performance liquid chromatography (HPLC) was carried out on ODS-Hypersil (4.6 × 250 mm) or MOS-Hypersil (9.4 × 250 mm) (Shandon Southern, England), using a Beckman HPLC system. ^{31}P and ^{1}H NMR spectra were obtained with a Varian XL-300 multinuclear spectrometer at 121.421 and 300 MHz, respectively. Absorption spectra were recorded by a Perkin-Elmer Lambda 3B UV/vis spectrophotometer. Oligodeoxynucleotides were synthesized using nucleoside phosphoramidites on an Applied Biosystems 381A DNA synthesizer. The fully protected 2′-deoxynucleoside phosphoramidites containing arene- or isobutanamides were purchased from Cruachem through Fisher Chemical Co. The phosphoramidite derivatives of purine 2′-deoxyriboside (dP) (Clore et al., 1988), 5-methyl-2-pyrimidone 2′-deoxyriboside (d5) (Gildea & McLaughlin, 1989), and 2,6-diaminopyrine 2′-deoxyriboside (dD) (Gaffney et al., 1984) were prepared as described in the appropriate references. The phosphoramidites of 2′-deoxyuridine (dU), 2′-deoxyinosine (dI), and 2′-deoxy-5-methylcytosine (dM) were prepared by standard procedures in our laboratory, but they are also available from Applied Biosystems, Inc. (Foster City, CA). The phosphoramidite derivative of 7-deaza-2′-deoxyadenosine (dc^7A) was prepared by a procedure similar to that described earlier (Seela & Kehne, 1988) and is described in full in the present report (see below). The controlled pore glass support containing the 3′-terminal nucleoside was a product of Applied Biosystems, Inc. The *trp* repressor was isolated from an overproducing strain provided as a gift from Prof. P. Sigler (Yale University). Calf intestinal alkaline phosphatase was a product of Boeh-

ringer, Mannheim (Germany).

Methods

Synthesis of N^4-(Benzoylamino)-7-[5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-β-D-erythro-pentofuranosyl]pyrrolo[2,3-d]pyrimidine (**1**). *N*-Benzoyl-2'-deoxytubercidin was prepared from tubercidin as described by Robins et al. (1983). To 502 mg (1.42 mmol) of *N*-benzoyl-2'-deoxytubercidin which was evaporated from dry pyridine (3×) and then dissolved in 25 mL of dry pyridine was added *N,N*-diisopropylethylamine (0.5 mL, 5.4 mmol) followed by 4,4'-dimethoxytrityl chloride (723 mg, 2.14 mmol). The reaction mixture was stirred at room temperature under an argon atmosphere for 2 h. At this point, TLC analysis (CH_2Cl_2/CH_3OH 9/1) indicated a complete reaction and the presence of a single product (R_f = 0.52). The reaction was stopped with methanol and stirred an additional 5 min. The solution was evaporated to dryness, and the residue was dissolved in dichloromethane and washed successively with 5% $NaHCO_3$ and water. The organic layer was dried (Na_2SO_4), filtered, and evaporated to dryness. The product was purified by short column chromatography (dichloromethane with a gradient of methanol) to give **1**. Yield: 964 mg (69%) as yellow foam. R_f (CH_2Cl_2/CH_3OH 9/1): 0.52. UV (CH_3OH): λ_{max} = 222, 272, 280; λ_{min} = 258. ^1H NMR (DMSO-d_6) (TMS): δ 2.3 (1 H, m, $H_{2'}$), 2.6 (1 H, m, $H_{2''}$), 3.2 (2 H, m, $H_{5'}$, $H_{5''}$), 3.7 (6 H, s, OCH_3), 4.0 (1 H, m, $H_{4'}$), 4.4 (1 H, m, $H_{3'}$), 5.75 (1 H, d, 3'-OH), 6.7 (1 H, dd, $H_{1'}$, J = 6.8, 6.1 Hz), 6.8–7.8 (18 H, m, Ar-H), 7.1 (1 H, d, H_7), 8.1 (1 H, d, H_8), 8.6 (1 H, s, H_2), 11.16 (1 H, s, NH).

Synthesis of N^4-(Benzoylamino)-7-[3'-O-(N,N-diisopropylamino-β-cyanoethoxyphosphinyl)-5'-O-(4,4'-dimethoxytrityl)-2'-deoxy-β-D-erythro-pentofuranosyl]pyrrolo[2,3-d]pyrimidine (**2**). To a solution of **1** (600 mg, 0.914 mmol) in anhydrous dichloromethane (20 mL) at room temperature was added *N,N*-diisopropylethylamine (353.6 mg, 2.73 mmol), and the reaction mixture was stirred for 15 min under an argon atmosphere. To this solution was added 2-cyanoethyl *N,N*-diisopropylphosphoramidite (323.7 mg, 1.36 mmol) dropwise over a period of 10 min, and the reaction mixture was stirred an additional 2 h. The reaction was stopped by the addition of methanol (2 mL) and stirred for 5 min. The solvent was evaporated to yield a yellow oil and purified by a short column chromatography. The product eluted in dichloromethane/methanol/triethylamine 9.8/0.1/0.1. The fractions containing product were collected and evaporated to an oil that was dissolved in a small quantity of anhydrous dichloromethane and precipitated into cold petroleum ether (−70 °C) to give **2** as light yellow solid. Yield: 700 mg (89%). R_f (CH_2Cl_2/hexane/$(CH_3CH_2)_3N$, 45/45/10): 0.75, 0.80. ^{31}P NMR ($CDCl_3$) (H_3PO_4 ext.): 149.0, 149.3.

Oligonucleotide Synthesis. The oligonucleotides were synthesized on controlled pore glass supports using phosphoramidite chemistry (Matteucci & Caruthers, 1981) and an Applied Biosystems 381A DNA synthesizer. After deprotection by base (concd ammonia for 6 h at 50 °C), the sequences were isolated using a 9.4 × 250 mm column of ODS-Hypersil at a flow rate of 3.0 mL/min in 50 mM triethylammonium acetate (pH 7.0) with a gradient of acetonitrile (20–65% M over 40 min) (McLaughlin & Piel, 1984). After isolation, the 5'-terminal DMT group was removed in 80% aqueous acetic acid at 0 °C for 30 min. Most of the acid was then removed by evaporation from ethanol/water 1/1 (3×), and the fragments were desalted (Sephadex G-10) and lyophilized to dryness.

Nucleoside Analysis. Nucleoside composition was determined after snake venom phosphodiesterase/bacterial alkaline phosphatase hydrolysis. A 20-μL reaction mixture containing 0.5 A_{260} unit of oligomer in 100 mM Tris-HCl, pH 8.0, 20 mM $MgCl_2$, 3 units of snake venom phosphodiesterase, and 2 units of bacterial alkaline phosphatase was incubated at 37 °C overnight. A 5-μL aliquot was analyzed by HPLC using a 4.6 × 250 mm column of ODS-Hypersil in 20 mM potassium phosphatase, pH 5.5, and a gradient of 0–70% methanol (60 min). Under these conditions the following retention times were observed (260 nm): dC, 6.7; dU, 7.2; dM, 9.6; dI, 10.3; dG, 11.0; d5, 11.8 (at 310 nm); dT, 11.9; dD, 12.4; dP, 14.5; dA, 15.5; dc^7A, 17.1 min.

T_m Values. T_m values were obtained in 10 mM Tris-HCl (pH 7.4) and 250 mM sodium chloride at duplex concentrations in the low micromolar range. Absorbance values were measured with a Perkin-Elmer Lambda 3B UV/visible spectrophotometer equipped with a digital temperature control. The solution temperatures were measured directly with a thermister probe (OMEGA Engineering, Stanford, CT). Absorbance and temperature data were collected after analog to digital conversion (DT-2800; Data Translation, Marlboro, MA) using an IBM-XT computer and the ASYST (version 1.53) scientific software package (MacMillian Software, New York, NY). T_m values were determined from first- and second-order derivatives of the absorbance vs temperature plots.

Radioisotopic Labeling. Each 20-mer was 5'-end-labeled with [γ-^{32}P]ATP as follows: A 200-μL reaction mixture contained 0.5 A_{260} unit of 20-mer (~0.1 mM), 40 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 0.2 mM Na_2EDTA, 0.005–0.025 mM ATP, ~50 μCi of [γ-^{32}P]ATP, 5 μg/mL BSA, and 10 units of T_4 polynucleotide kinase was incubated overnight at 37 °C. The product was isolated by adsorption on a C_{18} Sep-Pak cartridge. The cartridge was washed with 4% aqueous methanol to remove the excess ATP and then with 20% aqueous methanol to elute the product. The product was lyophilized to dryness and dissolved in 10 mM Tris-HCl, pH 7.4, and 250 mM sodium chloride at a duplex concentration of approximately 0.5 mM. This solution was heated to 80 °C for 5 min, slowly cooled to 5 °C, and finally diluted with buffer to a final duplex concentration of 0.5–1.0 μM.

Alkaline Phosphatase Binding Analysis. The assay as described by Marmorstein et al. (1991) was employed. Some of the details of the procedures are noted below: Both the *trp* repressor and the calf intestine alkaline phosphatase were dialyzed against 10 mM Tris-HCl, pH 7.4, and 250 mM sodium chloride. Approximately 1 mg of repressor in 30 μL of storage buffer (0.5 mM sodium phosphate, 250 mM sodium chloride, pH 7.0) was diluted to a final volume of 500 μL and dialyzed at 4 °C. Three buffer changes were completed at 4-h intervals, and after 16 h the dialysis was stopped, the protein was transferred to a suitable container, and the final concentration was determined spectrophotometrically at 280 nm ($\epsilon = 1.48 \times 10^4$ M^{-1} cm^{-1}). A total of 100 units of alkaline phosphatase (1 unit/μL) was also dialyzed under the same conditions.

Alkaline Phosphatase Reaction in the Absence of Repressor. A 200-μL reaction solution at 22 °C containing 20 nM duplex operator in 10 mM Tris-HCl, pH 7.4, and 250 mM sodium chloride and containing 10 μg/mL BSA was initiated by the addition of 2.5 or 5.0 units of alkaline phosphatase. Aliquots of 20–30 μL were removed at 15-s intervals (usually for a 2-min total period), and the reaction was quenched by addition of the aliquot into 25 μL of 0.15 M potassium phosphate, pH 7.0, containing 100 A_{260} units/mL of bulk tRNA. The nucleic acids were precipitated by the addition

of 3 mL of 5% aqueous trichloroacetic acid (TCA) followed by incubation for 5 min at ambient temperature. The TCA solution was filtered through a nitrocellulose membrane, followed by filtration of an additional 3-mL "wash" solution of TCA. The membrane was dried, and the radioactivity was determined in a toluene-based scintillation fluid.

Alkaline Phosphatase Reaction in the Presence of Repressor. The alkaline phosphatase reaction in the presence of the *trp* repressor was performed by essentially the same procedure with the following exceptions: The 200-μL reaction solution contained 0.1, 1.0, or 10 μM repressor (dimer) and 0.5 mM L-tryptophan co-repressor, and this mixture was incubated 30 min at 22 °C prior to the addition of the phosphatase. Control reactions were also performed in absence of the co-repressor. The time invervals for removing aliquots of solution for precipitation and analysis varied depending upon the rate of reaction.

Data Analysis. The amount of radiolabeled duplex present at various time periods, determined from the precipitated radioactivity (as a ratio of the starting labeled duplex), was plotted logarithmically as a function of reaction time as described previously. In some reactions with the repressor, a biphasic reaction rate appeared to be present: an initial rapid dephosphorylation occurred, similar to the rate of the alkaline phosphatase reaction in the absence of the repressor, followed by a second slower rate. The initial rate accounted for approximately 10–15% of the labeled oligodeoxynucleotide and likely reflected the presence of non-duplex structures (e.g., hairpins). The presence of such structures is not generally considered in assays using self-complementary oligodeoxynucleotides, but they are likely to occur, particularly when the concentration of the oligodeoxynucleotide is relatively low. In cases where a biphasic reaction rate appeared to be present, the second portion of the rate curve was used to extrapolate the concentration of "active" duplex 20-mer present when the reaction was initiated. Rate constants were obtained from the plots of the alkaline phosphatase reaction with (k_i') and without (k_i) repressor (R) and an apparent dissociation constant (K_D) was obtained from the following relationship: $K_D = [R]/(k_i/k_i' - 1)$, as has been described previously (Marmorstein et al., 1991). The K_D values reported are the result of a minimum of four independent assays and the standard deviations reported reflect deviations in these independent assays.

RESULTS

Design of the Modified Operator Sequences. We have used a 20-base-pair sequence containing the 18 base pairs of the *trp* operator sequence used in crystallographic analysis (Otwinowski et al., 1988) of the repressor–operator complex:

```
          -9   -5   -1 +1   +5   +9
5'-CGTACTAGTTAACTAGTACG
  GCATGATCAATTGATCATGC-5'
    +9   +5   +1 -1   -5   -9
```

Although two distinctly different alignments of the repressor about the operator sequence have been proposed (Otwinowski et al., 1988; Staacke et al., 1990), other experiments, including genetic studies (Bass et al., 1987, 1988), hydroxyl radical footprinting (Carey, 1989), and in vitro binding assays (Marmorstein et al., 1991; Haran et al., 1992) suggest that the traditional operator sequence, and that employed in this study, is the preferred target of the *trp* repressor molecule. In vivo studies have implicated a portion of the 18-base-pair operator sequence, including the −7/+7 through −3/+3 base pairs as being critical for binding by the repressor (Bass et al., 1987, 1988). In the crystallographic analysis of the repressor–operator complex, the structure at the base pair steps −5/−4 and +4/+5 have been suggested to be crucial for sequence-specific binding of the repressor to the operator sequence (Otwinowski et al., 1988), and these two sites appeared a likely area of the recognition sequence to initiate studies involving base analogues. Seven base pairs have been substituted for the dA-dT base pair at the −4/+4 positions and the dT-dA base pair at the −5/+5 positions. In addition to base pairs containing altered functional groups, we have incorporated a corresponding dG-dC or dC-dG base pair at each of the appropriate sites. The base pairs with altered functional groups fall into two classes: deletion modifications and functional group reversals. In the first group, four base pair analogues have been employed to systematically delete each of the four critical functional groups located in the major groove of a dA-dT (or dT-dA) base pair (the adenine N7, the adenine NH_2, the thymine O4, and the thymine CH_3) (Figure 1). Each functional group deletion was accomplished by the incorporation of the appropriate analogue: 7-deaza-2′-deoxyadenosine (dc^7A), purine 2′-deoxyriboside (dP), 5-methyl-2-pyrimidinone 2′-deoxyriboside (d5), and 2′-deoxyuridine (dU).

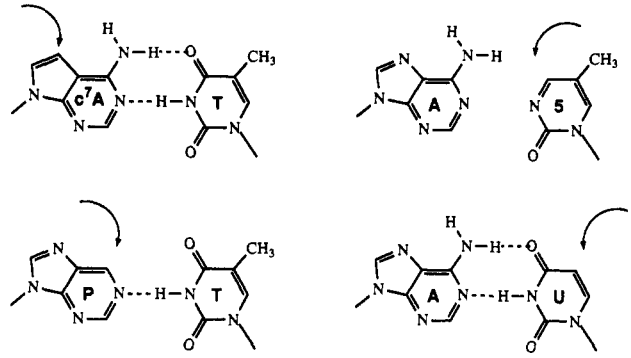

FIGURE 1: Deletion-modified adenine–thymine base pairs (arrows mark the site of the deleted functional group in each base pair).

In some cases (e.g., dP-dT and dA-d5) the deletion of a specific functional group involved in Watson–Crick base pairing results in the loss of one (or more) interstrand hydrogen bond. The resulting local or global structural effects resulting from the loss of such hydrogen bonds at specific sites may themselves contribute to an altered affinity between the protein and the nucleic acid and confuse the interpretation of the binding assays. In order to partially alleviate such concerns, we have employed two additional base pair analogues: hypoxanthine–5-methylcytosine (dI-dM) and 2,6-diaminopurine–uracil (dD-dU) can be viewed as dA-dT and dG-dC structural analogues (Figure 2). Both modified base pairs maintain the two or three interbase hydrogen bonds characteristic of dA-dT or dG-dC base pairs, respectively, but reverse the relative orientation of the amino and carbonyl functional groups involved in the interbase hydrogen bond present in the major groove.[1] The use of these analogues permits alteration of the functional group character present in the major groove with little accompanying structural effects.

Synthesis of the Modified Operator Sequences. In addition to the native sequence, 14 modified sequences were prepared in which the dA-dT and dT-dA base pairs at positions −4/+4

[1] The formal attachment of the hydrogen atom involved in the purine N1–pyrimidine N3 hydrogen bond is also reversed in these two base pairs. However, there is no evidence that the *trp* repressor opens up the helix to probe the Watson–Crick faces of the base pairs, so this reorientation of the formal bonding of the hydrogen atom at this site is unlikely to have a significant effect upon the structure of the DNA or the manner in which the protein interacts with the sequence.

Table I: Consequences of Base Analogue Substitutions at the −4/+4 Positions within the *trp* Operator

trp modified operator sequences (−5-4-3-2-1+1+2+3+4+5)	position	base pair	T_m (°C)	0.1 μM repressor[a] K_D (μM)	1.0 μM repressor[a] K_D (μM)	10 μM repressor[a] K_D (μM)
1 CGTACTAGTTAACTAGTACG	−4/+4	AT	65.0	0.003 ± 0.0012	0.024 ± 0.005	
2 CGTACTGGTTAACCAGTACG	−4/+4	GC	70.2		0.57 ± 0.04	
3 CGTACTDGTTAACUAGTACG	−4/+4	DU	67.4	0.011 ± 0.002	0.018 ± 0.007	
4 CGTACTIGTTAACMAGTACG	−4/+4	IM	63.0		0.33 ± 0.02	
5 CGTACTAGTTAACUAGTACG	−4/+4	AU	62.7	0.032 ± 0.004	0.046 ± 0.007	
6 CGTACTPGTTAACTAGTACG	−4/+4	PT	56.0	0.0013 ± 0.0004	0.039 ± 0.004	
7 CGTACTAGTTAAC5AGTACG	−4/+4	A5	52.2		np[b]	0.47
8 CGTACT(c⁷A)GTTAACTAGTACG	−4/+4	c⁷AT	64.5	0.008 ± 0.002	0.024 ± 0.002	
gal CGGTGTAAACGTTTACACCG	*gal* operator					>14

[a] Concentration of repressor used in the protection assay. The apparent dissociation constant (K_D) varies artificially with the concentration of the repressor used. It is best to compare K_D values obtained under identical repressor concentrations. [b] np = no protection was observed at this repressor concentration.

Table II: Consequences of Base Analogue Substitutions at the −5/+5 Positions within the *trp* Operator

trp modified operator sequences (−5-4-3-2-1+1+2+3+4+5)	position	base pair	T_m (°C)	0.1 μM repressor[a] K_D (μM)	1.0 μM repressor[a] K_D (μM)	10 μM repressor[a] K_D (μM)
1 CGTACTAGTTAACTAGTACG	−5/+5	TA	65.0	0.003 ± 0.0012	0.024 ± 0.005	
2 CGTACCAGTTAACTGGTACG	−5/+5	CG	70.5		0.067 ± 0.010	
3 CGTACUAGTTAACTDGTACG	−5/+5	UD	64.2	0.005 ± 0.0008	0.048 ± 0.001	
4 CGTACMAGTTAACTIGTACG	−5/+5	MI	64.3		0.13 ± 0.01	
5 CGTACUAGTTAACUAGTACG	−5/+5	UA	64.4	0.002 ± 0.0007	0.051 ± 0.015	
6 CGTACTAGTTAACTPGTACG	−5/+5	TP	56.6	0.00038 ± 0.00001	0.017 ± 0.001	
7 CGTAC5AGTTAACTAGTACG	−5/+5	5A	47.6		np[b]	np[b]
8 CGTACTAGTTAACT(c⁷A)GTACG	−5/+5	Tc⁷A	64.8		0.65 ± 0.10	
gal CGGTGTAAACGTTTACACCG	*gal* operator					>14

[a] Concentration of repressor used in the protection assay. The apparent dissociation constant (K_D) varies artificially with the concentration of the repressor used. It is best to compare K_D values obtained under identical repressor concentrations. [b] np = no protection observed at this repressor concentration.

and −5/+5 were replaced by dG-dC (dC-dG) as well as each of the modified base pairs illustrated in Figures 1 and 2. Seven 2′-deoxynucleoside phosphoramidites were required for these syntheses. Six of the phosphoramidite derivatives were prepared as we or others have previously described (Gaffney et al., 1984; Clore et al., 1988; Gildea & McLaughlin, 1989). The 7-deaza-2′-deoxyadenosine derivative was prepared by a modification of a previously described (Seela & Kehne, 1988) procedure (see Methods). The native and modified sequences were synthesized using solid-phase based procedures that employed a controlled pore glass bead support (Matteucci & Caruthers, 1981).

After assembly, deprotection, and purification, the sequences were analyzed by HPLC and polyacrylamide gel electrophoresis. A typical HPLC analysis is illustrated in Figure 3a. On the basis of these analyses, the oligodeoxynucleotides were estimated to be >95% pure. Nucleoside analysis (see Figure 3b) confirmed the presence of one equivalent of the modified residue(s).

T_m Values of the Modified Operator Sequences. The thermal stabilities of the self-complementary operator duplexes were determined under the pH and salt conditions used in the binding assays (10 mM Tris-HCl, pH 7.4, 250 mM NaCl). Under these conditions, the T_m value for the native *trp* operator sequence at ~5 μM duplex concentration was 65.0 °C. The incorporation of modified base pairs into two symmetric positions (−4/+4 or −5/+5) had varying effects upon stability. Base pairs that maintained the same overall number of interstrand hydrogen bonds in the 20-mer sequence exhibited T_m values that were within approximately 2 °C of that obtained for the native sequence (compare entries 4, 5, and 8 in Tables I and II with the native sequence, entry 1). Base pairs that resulted in an increase in the number of interstrand hydrogen bonds generally exhibited slightly higher T_m values (compare entries 2 and 3 in Tables I and II with the native sequence, entry 1). Four modified sequences, those containing dP-dT and dA-d5 base pairs result in a net loss of interstrand hydrogen bonds relative to the native sequence. Substitution of purine for adenine eliminates one hydrogen bond (at each of the symmetric sites) present in the major groove of the duplex DNA (see Figure 1). However, previous NMR experiments (Ikuta et al., 1987; Clore et al., 1988) suggest that a single Watson–Crick hydrogen bond remains between the purine N1 and the thymine NH3 (as shown in Figure 1). The substitution of two dP-dT base pairs into the operator sequence results in a net loss of two interstrand hydrogen bonds from the duplex operators, and the T_m values for these derivatives

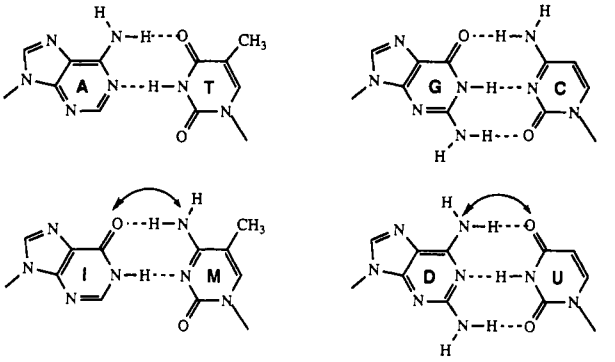

FIGURE 2: Native dA-dT and dG-dC base pairs (top) and the corresponding base pair analogues containing functional group reversals (bottom, arrows indicated the relative positions of the "reversed" functional groups).

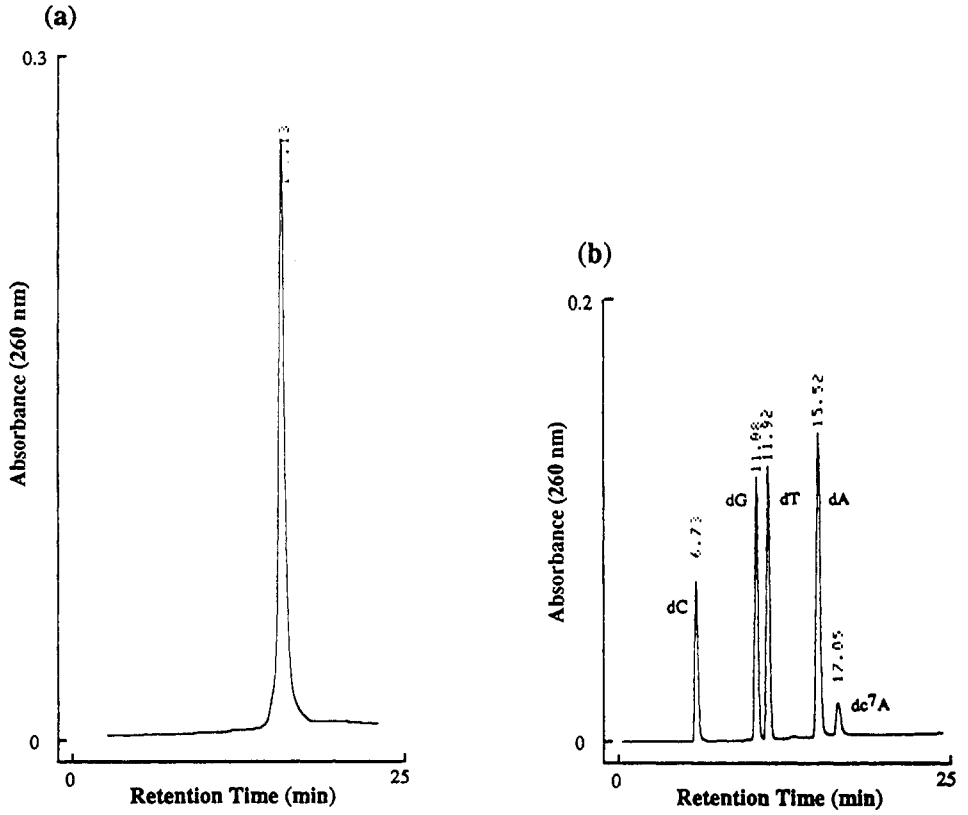

FIGURE 3: (a) HPLC analysis of the 20-mer containing the *trp* operator sequence and a dc^{7}A-dT analogue at the −4 position. Column: 4.6 × 250 mm ODS-Hypersil. Buffer: 50 mM KH_2PO_4, pH 5.5. Gradient: 0–75% methanol over 60 min. (b) HPLC analysis of the nuclease digest of the 20-mer described above in (a). HPLC conditions were as described in (a).

are approximately 8–9 °C lower than that of the native fragment (see entry 6 in Tables I and II). By comparison, substitution of the 2-pyrimidinone derivative, d5, for dT also results in the loss of the interstrand hydrogen bond present in the major groove. However, in this case a tautomeric shift in the character of the N3 nitrogen eliminates the hydrogen formally attached to the thymine N3. This tautomeric shift results in the loss of both Watson–Crick hydrogen bonds present in a native dA-dT base pair (see Figure 1). A net loss of four hydrogen bonds (two at each symmetric site) within the operator sequence results in approximately a 13 °C (at −4/+4) and a 17 °C (at −5/+5) decrease in T_m values (compare entry 7 of Tables I and II with the native sequence, entry 1).

Binding of the Native Operator Sequence to the trp Repressor. We have used the recently developed alkaline phosphatase protection assay (Marmorstein et al., 1991) for the determination of equilibrium dissociation constants between the synthesized operator sequences and the *trp* repressor. In this assay, binding by the repressor to radiolabeled sequences of 20 nucleotide residues protects the sequences from dephosphorylation by calf intestinal alkaline phosphatase. The rate of dephosphorylation in this case is dependent upon both the dissociation constant (K_D), characterizing the binding between the repressor[2] and operator, and the concentration of the repressor (R) available in solution. Plots of the natural logarithm of the fraction of phosphorylated operator versus time should generate a linear function with the slope of the line equal to the apparent rate constant. With a given dissociation constant characterizing the binding between the repressor and operator, the extent of protection from alkaline phosphatase hydrolysis will vary with the concentration of repressor available. As illustrated in Figure 4, significantly more protection was observed for the dD-dU-containing sequence with increasing concentrations of repressor.

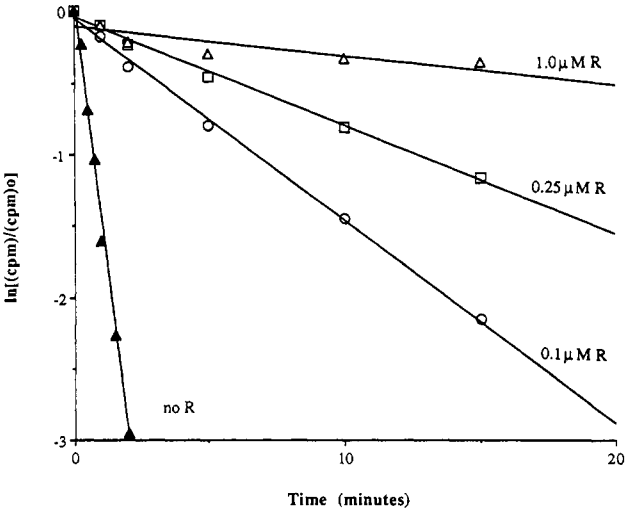

FIGURE 4: Plots illustrating the protection of the 20-mer, containing the *trp* operator sequence (~20 nM) and a dD-dU analogue at the −4/+4 positions, from alkaline phosphatase hydrolysis at varying concentrations of *trp* repressor (R).

In some cases, in the presence of the *trp* repressor, we observed a biphasic character for these rate plots. An initial rate, almost identical with the rate for alkaline phosphatase hydrolysis in the absence of the repressor, was followed by a slower rate that appeared to reflect protection by the repressor. The first portion of such plots typically accounted for only

[2] Although not explicitly stated in each case, all assays were performed in the presence of 0.5 mM of co-repressor, tryptophan (see Experimental Procedures). The descriptions of repressor–operator binding actually reflect the binding of the repressor–corepressor complex to the operator sequence.

about 10–15% of the hydrolyzed material. The T_m values for all the operator sequences used in this study, and measured in the low micromolar concentration range, suggest the presence of well-characterized duplexes. However, at the low duplex concentrations used for the protection assay (~20 nM) some formation of hairpin loop or other unusual structures in addition to the desired 20-mer duplex can be expected. In cases where biphasic rate plots were obtained, we have assumed that the initial fast rate of hydrolysis, unaffected by the presence of the repressor, represents hairpin or other nonduplex structures. We have used the second portion of the curve to extrapolate the concentration of the "active" 20-mer duplex at time zero.

By comparing the rate constants for the dephosphorylation of the radiolabeled operator sequence in presence (k_i') and absence (k_i) of the repressor (R), the apparent dissociation constant (K_D) can be derived from the relationship $K_d = [R]/(k_i/k_i' - 1)$ (Marmorstein et al., 1991). However, since the concentration of repressor enters into this calculation for the value of K_D, it is best to compare the relative dissociation constants for a series of sequences under identical conditions, and at a repressor concentration that is not excessively high.

In our hands, a repressor concentration of 0.1 μM was sufficient to exhibit significant protection of the operator sequence. Under these conditions, a dissociation constant of 3 nM was obtained characterizing the binding between the native operator and the *trp* repressor. This value is essentially identical with the value obtained by Marmorstein et al. in the recent work describing this assay. Additionally, this value compares favorably with the K_D values of 2 nM obtained by Klig et al. (1987) and 5.9 nM reported by Marmorstein and Sigler (1989) using filter binding assays. The measured K_D is somewhat higher than the value of 0.5 nM measured by Carey (1988) using a gel shift assay at pH 6.0.

We have also used a second unrelated operator sequence (the *gal* operator) to estimate nonspecific binding. With the *gal* operator sequence, we were unable to observe any protection (binding) by the *trp* repressor at 0.1 or 1.0 μM concentrations of repressor. At 10 μM *trp* repressor, we were able to obtain a value, although difficult to reproduce with precision, indicating that the K_D value for binding of the *trp* repressor to the *gal* operator was not less than 14 μM.

Binding of Modified Operator Sequences to the trp Repressor. Initial binding experiments indicated that some of the modified operator sequences were only poorly protected at a repressor concentration of 0.1 μM. In general, it was most efficient to analyze all operator sequences at the higher repressor concentration of 1.0 μM. Although using a higher repressor (R) concentration artificially raised the measured K_D to 24 nM for the native system, we could then compare relative binding affinities for all the sequences prepared. Operator sequences that appeared to bind as well as did the native sequence were then reassayed at a lower concentration of repressor (0.1 μM). Sequences that did not appear to be protected at 1.0 μM repressor were assayed again in the presence of 10 μM repressor. Using these procedures, we could compare the relative binding affinities for all sequences at a single repressor concentration, and yet provide some discrimination between sequences, that appeared (at 1.0 μM repressor) to have relatively low dissociation constants similar to that of the native sequence or, conversely, very high dissociation constants.

Consequences of Modifications at the −4/+4 Base Pair. Analysis of the crystal structure of the *trp* repressor–operator complex (Otwinowski et al., 1988) did not suggest the presence of any amino acid–base contacts at the −4/+4 base pair. Structural anomalies of the DNA sequence within the repressor–operator complex, particularly the high degree of roll associated with both the −4/+4 and −5/+5 base pairs, was suggested to contribute to the overall binding affinity. Replacement of the dA-dT base pair at −4/+4 with a dG-dC base pair resulted in a K_D value of 0.57 μM, some 24-fold higher than that observed for the native sequence (compare entries 1 and 2 at 1.0 μM repressor, Table I). Although the dD-dU base pair can be viewed as a structural analogue of dG-dC (see Figure 2), the placement of a dD-dU at the −4/+4 position resulted in a K_D value very similar to that of the native sequence (entry 3, Table I). By comparison, the operator sequence containing a dA-dT structural analogue (dI-dM, entry 4, Table I) resulted in an apparent dissociation constant that was 14-fold higher than that characterizing the unmodified operator.

In addition to the structural analogues, the four "deletion-modified" base pairs (see Figure 1) were each placed into the −4/+4 site. Deletion of the thymine methyl group, the adenine amino group, or the adenine N7 nitrogen (entries 5, 6, and 8, Table I) resulted in sequences characterized by K_D values differing from that of the unmodified sequence by less than 2-fold. By comparison, deletion of the thymine carbonyl (entry 7, Table I) produced a sequence that did not exhibit significant protection in the presence of 1.0 μM repressor.

Sequences 3, 5, 6, and 8 (Table I) with K_D values similar to that obtained for the unmodified operator (at a concentration of 1.0 μM repressor) were reanalyzed in the presence of a reduced concentration of the repressor (0.1 μM, see Table I). The operators containing either a dD-dU or a dc^7 A-dT base pair exhibited binding characteristics that were very similar to those of the native sequence (11 nM and 8 nM, respectively, vs 3 nM). However, the dA-dU-containing sequence resulted in a K_D value approximately 10-fold higher (32 nM) than that observed with the unmodified operator (3 nM). Surprisingly, the dP-dT containing sequence (entry 6) exhibited a K_D value that was slightly, but reproducibly, lower (1.3 nM) than that of the native sequence.

The only sequence that did not exhibit any protection by a 1.0 μM concentration of the repressor (entry 7, Table I) was reanalyzed in the presence of 10 μM repressor. In this assay, some protection was observed, and a K_D value of 0.47 μM was obtained.

Consequences of Modifications at the −5/+5 Base Pair. This base pair, in conjunction with the −4/+4 base pair, also exhibits a high degree of roll, on the basis of analysis of the crystal structure. Together, the −4/+4 and −5/+5 base pairs are responsible for the formation of two symmetrically placed kinks within the DNA structure of the repressor–operator complex. Additionally, analysis of the crystal structure suggests the presence of a water-mediated contact that bridges the adenine N7 nitrogen at A_{+5}. This water molecule is also bound the O6 of dG_{+6} and the amide nitrogen of Ala_{80}. In contrast to the results obtained for substitutions at the −4/+4 position, replacement of the native dT-dA base pair by either a dC-dG base pair or either of the structural analogues dU-dD or dM-dI resulted in only marginal changes in binding affinity when measured in the presence of 1.0 μM repressor (compare entries 2, 3, and 4 with entry 1, Table II). The presence of a dC-dG base pair resulted in a 3-fold increase in K_D (0.067 μM) relative to the unmodified operator, and the greatest change in affinity for these derivatives occurred with the dT-dA structural analogue dM-dI, where a 5-fold increase in K_D was observed (129 nM).

Deletion of the individual functional groups present in the major groove of the dT-dA base pair had varying effects upon repressor affinity. Excision of either the thymine methyl group or the adenine amino group (entries 5 and 6, Table II) resulted in little change in the derived K_D value. However, deletion of the thymine carbonyl (entry 7, Table II) generated a sequence that could not be protected by the repressor, even at a repressor concentration of 10 μM. Additionally, substitution of 7-deazaadenine at the −5/+5 position resulted in a K_D of 0.65 μM, 27-fold higher than that measured for the native sequence.

Three modified operators were further analyzed in the presence of a lower concentration of repressor (0.1 μM, Table II). The dC-dG structural analogue dU-dD, containing three interbase hydrogen bonds, exhibited a binding affinity for the repressor that was virtually identical with that of the native operator (5 nM). Additionally the deletion of the methyl group from dT_{-5} had no effect upon the derived K_D value (2 nM). Surprisingly, deletion of the adenine amino group (replacement by dT-dP, entry 6, Table II) resulted in an approximately 8-fold decrease in the K_D value, suggesting the formation of a complex that is more stable (K_D = 0.38 nM) than that obtained with the unmodified operator.

The only sequence that did not exhibit any protection by the repressor was the 20-mer containing the d5-dA base pair at position −5/+5 (entry 8, Table II). Even with a repressor concentration of 10 μM, the rate of alkaline phosphatase hydrolysis of the terminal phosphomonoester was unchanged from that observed in the absence of the repressor. This result must be compared with that obtained for a non-*trp* operator sequence (the *gal* operator, entry 9, Table II). In the presence of 10 μM *trp* repressor, the *gal* operator sequence exhibited some protection from alkaline phosphatase activity such that a K_D value of approximately 14 μM could be derived.

Discussion

Two types of base analogues have been employed to probe both structural (interbase hydrogen bonding) relationships and specific functional group contacts involved in the formation of the high-affinity sequence-specific *trp* repressor–operator complex. Deletion-modified analogues have been used to excise a specific functional group at a preselected site within the DNA sequences of interest. Structural analogues of adenine–thymine and guanine–cytosine base pairs (e.g., dI-dM and dD-dU) have been used in order to maintain the fundamental interbase hydrogen bonded structure of the respective natural base pair and yet alter the functional group character present in the major groove of the DNA duplex. Both the deletion analogues and the structural analogues described here have been used in other recent studies to probe DNA curvature (Diekmann et al., 1987, 1992; Diekmann & McLaughlin, 1988) as well as sequence-specific recognition by phase repressor proteins (Goeddel et al., 1977, 1978; Caruthers, 1985), restriction endonucleases (Dwyer-Hallquist et al., 1982; Ono et al., 1984; Yolov et al., 1985, Brennan et al., 1986a; Jiricny et al., 1986; Seela & Driller, 1986; Seela & Kehne, 1987; McLaughlin et al., 1987; Nwosu et al., 1988; Mazzarelli et al., 1989; Lesser et al., 1990; Newman et al., 1990a,b; Aiken et al., 1991), modification enzymes (Brennan et al., 1986b), and RNA polymerase (Dubendorff et al., 1987).

T_m Values. The relative thermal stabilities obtained for the various 20-mer duplexes were generally within expected values. Replacement of the native dA-dT (or dT-dA) base pairs by dG-dC (or dC-dG) increased the overall number of interstrand hydrogen bonds by two, and this resulted in a corresponding (~5 °C) increase in T_m. The 2,6-diaminopurine–uracil base pair has been shown by crystallographic analysis to form three hydrogen bonds (Coll et al., 1986), analogous to a dG-dC base pair. Although the substitution of dD-dU for dA-dT results in the presence of two additional interstrand hydrogen bonds, the T_m values for these sequences were not significantly increased relative to the native 20-mer. This is a common phenomenon for dD-dU (or dD-dT) base pairs (Gaffney et al., 1984), and their stability relative to dA-dT and dG-dC base pairs has been explained in the elegant work by Jorgensen (1990). Base pair analogues that cause an overall loss of interstrand hydrogen bonds resulted in duplexes that exhibited reduced T_m values. Replacement of dA-dT by dP-dT at two symmetric sites removes two interstrand hydrogen bonds from the 20-mer duplex, and a corresponding 8–9 °C decrease in T_m was observed. Replacement of dA-dT by dA-d5 results in the loss of four hydrogen bonds, and a 13–17 °C decrease in T_m was observed. These results are also generally consistent with other studies that we have reported employing such base pair analogues [see for example, McLaughlin et al. (1987)]. We have previously observed that a self-complementary decamer containing two dA-d5 base pairs did not form a duplex at temperatures above 5 °C (Gildea & McLaughlin, 1989), while a similar dodecamer produced anomalous absorbance vs temperature curves. Both of these earlier results suggest that the presence of dA-d5 base pairs results in a significant loss of duplex stability. However, other reports describing the incorporation of dA-d5 or dP-dT base pairs into self-complementary duplexes indicates that the T_m values for the modified sequences were largely unchanged from those of the native (Connolly & Newman, 1989; Newman et al., 1990a). We are unable at present to resolve this discrepancy. However, in the work presented here, the stability of the 20-mer helices was generally directly dependent upon the number of interstrand hydrogen bonds present in each duplex.

Binding of the Native Operator Sequence to the trp Repressor. The 3 nM K_D value obtained using the alkaline phosphatase protection assay is essentially the same as that obtained previously using this assay (Marmorstein et al., 1991), it is corresponds well with the values reported for filter binding assays (Klig et al., 1987; Marmorstein & Sigler, 1989). The apparent dissociation constant determined in this work is about 6-fold higher than the 0.5 nM value reported by Carey from a gel shift assay (Carey, 1988). However, the gel shift assays were performed under slightly acidic conditions at a pH of 6.0. The reduced pH value in that study may enhance binding by providing additional electrostatic contacts as the result of protonation as noted previously (Marmorstein et al., 1991).

The results presented here cannot be used to discriminate between the two suggested orientations of the *trp* repressor bound to the operator sequence. We have used the traditional *trp* operator sequence, essentially identical to that used in the crystal analysis and observed K_D values that are very similar to those generated using longer sequences (Klig et al., 1987; Carey, 1988; Marmorstein & Sigler, 1989); these latter studies would permit either orientation. Recent studies (Haran et al., 1992) additionally suggest that the traditional operator sequence, and that used here, is the preferred binding site.

Binding of the Modified Operator Sequences to the trp Repressor. The apparent K_D values derived in the alkaline phosphatase assay must be compared under indentical concentrations of repressor. We have observed that high repressor concentrations tend to obscure differences in K_D values because the rates of alkaline phosphatase hydrolysis become very similar. Therefore, sequences that exhibit apparent K_D values similar to that of the native have been assayed a second time

at reduced repressor concentration (0.1 μM). Under these conditions the sequences are not protected as well by the repressor (see Figure 4), and such differences in protection permit the observation of rate differences that reflect varying apparent K_D values. For example, at 1.0 μM repressor, deletion of the thymine methyl group (sequence 5, Table I) has an apparent K_D value only 2-fold higher than the native sequence. This slight difference can be enhanced by assaying the sequences at lower repressor concentration where a 10-fold difference in apparent K_D is evident. The relative variations in the apparent K_D values for the weakly binding sequences can be determined from the values obtained at 1.0 μM repressor, while relative variations in apparent K_D values for the tighter binding sequences can be determined at 0.1 μM repressor.

Analogues at the −4/+4 Positions. The binding between the repressor and operator is sensitive to some of the base analogues introduced at these positions, although the crystal structure analysis did not report the presence of any direct contacts to the base residues at the −4/+4 positions. Replacement of the native dA-dT base pair by a dG-dC base pair results in a 24-fold increase in the K_D value (Table I) suggesting that the presence of a dA-dT base pair at these two symmetrical sites is critical for effective recognition of the sequence. This is in agreement with the genetic studies of Bass et al., who report that the incorporation of a dG-dC base pair at the −4/+4 base pair has a severe effect upon in vivo repression. Both the present in vitro binding results and the previous in vivo results could be explained in terms of a structural alteration of the 5′-CTAG-3′ portion of the sequence in such a manner that the dG-dC base pair is unable to adopt the high degree of base pair roll present at this site in the crystal structure. The presence of the additional 2-amino group in the minor groove with a dG-dC substitution, and its involvement in a third hydrogen bond for the base pair, may inhibit formation of the structure observed in the crystallographic complex.

The dD-dU base pair is a structural analogue of dG-dC; it maintains three interbase hydrogen bonds and has a 2-amino group located in the minor groove. However, substitution of this analogue at the −4/+4 position results in a K_D value that is very similar to that of the unmodified operator. By comparison, the dI-dM base pair is a structural analogue of dA-dT; it contains two interstrand hydrogen bonds, it maintains the methyl group at the pyrimidine 5-position, and it lacks an amino group at the 2-position of the purine. However, this substitution results in a 14-fold increase in the dissociation constant. If the structure of the 5′-CTAG-3′ sequence, and in particular the structure associated with dA-dT at −4/+4, is critical for binding by the repressor, then it seems likely that repressor would exhibit similar affinities for the native operator sequence and that sequence containing the structural analogue dI-dM. Conversely, sequences containing dG-dC or its structural analogue dD-dU, both containing an additional functional group and a third interbase hydrogen bond, would exhibit reduced affinities. However, this conclusion is not consistent with the data.

A second view of the importance of the dA-dT base pair at the −4/+4 position would relate the relative positions of the functional groups in the major groove to the observed binding affinities. The functional group character of the major groove should be in intimate contact with the repressor. With this approach, the dD-dU base pair can be viewed as a dA-dT analogue; both base pairs contain a purine exocyclic amino group (in the major groove) with a pyrimidine carbonyl. Although the number of hydrogen bonds in the two base pairs differ, and the pyrimidine methyl group is absent from the dD-dU base pair, the relative orientations of the amino and carbonyl functional groups (forming the interbase hydrogen bond present in the major groove) are identical. Similarly, the relative orientations of these functional groups in dG-dC and dI-dM are identical but reversed from that present in dA-dT or dD-dU. The similarity in K_D values for the native sequence and the dD-dU analogue, as well as the increase in K_D values for both dG-dC and dI-dM containing sequences, suggests that the repressor is very sensitive to the orientation of this amino and carbonyl group present in the major groove. A similar interpretation can be extrapolated from the results of Bass et al. (1987). In this case the substitution of dG-dC or dT-dA for the dA-dT base pair at the −4/+4 positions resulted in severe effects, while substitution of dC-dG causes only mild effects upon in vivo repression. Although the geometry is not exact, the relative orientation of the amino and carbonyl groups in dA-dT and dC-dG is similar.

Site-specific deletion of individual functional groups suggested differences in the relative importance of specific groups for formation of the high-affinity complex. The results described above indicate that the orientation of the amino and carbonyl functional groups in the major groove at the −4/+4 sites is critical for the formation of a high-affinity complex. However, deletion of the adenine amino group from this hydrogen bonding pair did not reduce the affinity of the repressor for the sequence; in fact, binding affinity may be slightly enhanced in this case (see entry 6, Table I). Conversely, the deletion of the thymine carbonyl from this hydrogen bonding pair had a severe effect upon the K_D value (entry 7, Table I). The results of both the functional group deletions and functional group reversals at these sites suggest that the thymine carbonyl (at T_{+4}) is required for high-affinity binding by the repressor. We should note that the inability to derive a K_D value for the dA-d5 containing sequence except at high concentrations of repressor may suggest additional structural modulation at this site because of the lack of interstrand Watson–Crick hydrogen bonds. The suggestion that the thymine carbonyl is critically important for the formation of the *trp* repressor–operator complex was also suggested on the basis of the results of in vivo studies (Bass et al., 1987), but this conclusion does not appear to correlate well with the observations in the reported crystal structure (Otwinowski et al., 1988).

The deletion of the adenine N7 at dA_{-4} does not alter the affinity of the repressor for the operator sequence. Deletion of the thymine methyl group appeared to have little effect upon K_D when assayed at 1.0 μM repressor, but upon reducing the repressor concentration, a K_D of 32 nM was obtained. This value is some 10-fold higher than that observed for the unmodified sequence under the same conditions. The role of the thymine methyl groups could not be addressed with previous in vivo studies, but both observations (related to the N7 of dA_{-4} and the thymine methyl group) correlate well with observations reported from the crystal structure. Otwinowski et al. (1988) report that a hydrophobic contact is present near dT_{+4}, but no contact was observed near the N7 of dA_{-4}. Although the hydrophobic contact observed in the crystal structure involves the α-carbon of Gly_{78} and the C_6-H of dT_{+4}, the adjacent methyl group of the thymine C_5 is likely close enough to enhance such interactions and contribute to overall binding affinity. The importance of thymine methyl groups for the formation of high-affinity complexes between repressor proteins and operator sequences is well documented in the case

of the *lac* repressor (Goeddel et al., 1977, 1978).

Analogues at the −5/+5 Positions. In contrast to the results at the −4/+4 positions, binding by the repressor appears to be much less sensitive to the replacement of the dT-dA base pair at the −5/+5 positions with either of the structural analogues dU-dD or dM-dI, or in fact by replacement with a native dC-dG base pair. Of this series of analogues, the dM-dI substitution resulted in the highest dissociation constant, but the value of 129 nM was only 5-fold higher than the K_D value obtained for the native sequence (at 1.0 μM repressor). These results suggest that binding by the repressor is much less sensitive to the structure of the base pair at this position, and in contrast to the results at the −4/+4 positions, there is little recognition of the orientation of the amino and carbonyl functional groups that form the hydrogen bond present at these sites in the major groove.

Analysis of the crystal structure predicts that a water-mediated functional group contact is present at this base pair and involves the N7 nitrogen of the dA_{+5} residue (Otwinowski et al., 1988). Of all the modified sequences examined in this study, the deletion of this nitrogen by substitution of dT-dc^{7}A for dT-dA at the −5/+5 positions resulted in the most dramatic effects in binding, with a 27-fold increase in the K_D value. Although this study cannot confirm the existence of a water-mediated contact, it is clear that the N7 nitrogen is a critical functional group necessary for the formation of the high-affinity sequence-specific complex. The necessity for this purine nitrogen within the operator sequence was suggested on the basis of the in vivo studies of Bass et al. (1987). They have reported that substitution of dC-dG for dT-dA at the −5/+5 positions has a negligible effect upon in vivo repression, while replacement of this native base pair (dT-dA) by dG-dC or dA-dT results in severe effects. Both the native sequence and the dC-dG substitution place a purine N7 nitrogen in roughly the same location at the −5/+5 sites, while either of the other possible native base pair substitutions moves this functional group to the opposite side of the major groove.

In contrast to the results obtained for the −4/+4 positions, deletion of the thymine methyl group has little effect upon repressor binding, and this result argues against a significant hydrophobic contact at this site. Surprisingly, deletion of the adenine amino group produced an operator with a K_D value approximately 8-fold better than that of the native sequences. This observation is similar to, but more dramatic than, that obtained at the −4/+4 positions (see above). The results obtained for both purine substitutions at dA_{-4} and dA_{+5} may reflect the presence of a large solvent-excluded interface between the protein and the nucleic acid as suggested by the crystal structure. Removal of the hydrophilic amino group, which does not appear to be a critical functional group contact at either position, and its replacement by the relatively hydrophobic C-H residue may enhance van der Waals contacts at these sites and ultimately result in a *more* stable protein–nucleic acid complex as is implicated by the lower K_D values.

The operator sequence containing a d5-dA base pair at the −5/+5 positions could not be protected from alkaline phosphatase hydrolysis even in the presence of 10 μM *trp* repressor. Although this could be interpreted to suggest that the carbonyl at this site is a critical contact, the lack of sensitivity to other substitutions that alter the position of this carbonyl (see entries 2, 3, and 4 in Table II) does not support such an interpretation. The use of the d5 analogue suffers from the problem that there are no Watson–Crick hydrogen bonds that can be formed with a complementary adenine residue (see Figure 1). Although in some cases base stacking interactions on either side of these residues may assist in positioning this base pair (dA-d5) in a geometric arrangement that approximates a dA-dT base pair, there is also the possibility for global helix disruption. The operator sequence containing the d5-dA base pair at positions −5/+5 has the lower T_m value of all sequences analyzed, and at very low concentrations (~20 nM) it may undergo denaturation or may disproportionate such that little or no "active" helix results. Such effects may also be present in sequence 7 (Table I), but in this case we were able to observe some protection (i.e., binding) by the repressor (K_D = 0.47 μM). The inability to derive a K_D value for the $d5_{-5}$-dA_{+5} sequence at a high repressor concentration, as we were able to do for even the non-tryptophan sequence (the *gal* operator), suggests that this sequence does not result in significant amounts of a well-characterized duplex at the concentrations used in the assay (20 nM).

Energetics of Binding. Relative differences in binding energies were obtained from the measured apparent dissociation constants for the native ($K_{D^{nat}}$) and modified ($K_{D^{mod}}$) sequences and the apparent free energy of binding for a specific contact (ΔG_{app}) could be estimated [$\Delta G_{app} = RT \ln (K_{D^{nat}}/K_{D^{mod}})$] [for a review, see Fersht (1987)]. Relative values were compared under identical conditions using the K_D values obtained at either 0.1 or 1.0 μM (and in the case of dA-d5 at −4/+4, at 10 μM) repressor. Differences in relative binding energies were divided in half in order to reflect the apparent energy of a single interaction in each operator half-site. The strengths of uncharged hydrogen bonds in enzyme substrate complexes have been estimated to vary from 0.5 to 4.0 kcal/mol (Fersht et al., 1985; Wells & Fersht, 1986; Bartlett & Marlowe, 1987). Analysis of uncharged hydrogen bond interactions in the tyrosyl-tRNA synthetase–substrate complex results in values ranging from 0.5 to 3 kcal/mol. Deletion of the hydroxyl groups from the sugar in the glycogen phosphorylase–glucose complex was analyzed in a similar fashion with apparent binding energies for uncharged hydrogen bonds of approximately 1.5 kcal/mol (Street et al., 1986). Similar studies with restriction endonucleases have generated values in the 1–2 kcal/mol range (Newman et al., 1990b; Lesser et al., 1990; Aiken et al., 1991). In the present study, deletion of both thymine carbonyl groups from the two dT_{+4} sites results in a change in apparent overall binding energy of approximately 3 kcal/mol or 1.5 kcal/mol per half-site at 22 °C. A similar value of 1.6 kcal/mol (22 °C) per half-site was obtained for the change in apparent binding energy resulting from the loss of the two N7 nitrogens at dA_{+5}. These values are consistent with those obtained from the tyrosyl-tRNA synthetase–substrate complex after deletion of uncharged hydrogen bonds or with the values obtained for the *Eco*RI (Lesser et al., 1990), *Eco*RV (Newman et al., 1990), and *Rsr*I (Aiken et al., 1991) restriction endonucleases. These relatively low energy values may also reflect the ability of the protein to compensate for the lack of specific hydrogen bonding partners. Without compensatory interactions, the loss of a single uncharged hydrogen bond may account for a free energy of binding as high as 4 kcal/mol (Bartlett & Marlowe, 1987).

The hydrophobic contribution to sequence-specific binding is more difficult to quantitate. Although hydrophobic interactions and their contribution as a driving force for the formation of protein–DNA complexes has been studied in some detail (Jeung-Hoi et al., 1989), it is still difficult to estimate the individual contribution made by a specific residue, perhaps with the exception of the thymine methyl group. The free energy of binding for a methyl group in protein or enzyme complexes has been determined in several cases and ranges

from 1.5 kcal/mol for the *lac* repressor–operator complex (Goeddel et al., 1977) to 3.3 kcal/mol for several aminoacyl-tRNA synthetases (Owens et al., 1970; Fersht & Dingwall, 1979; Fersht et al., 1980). In the present case, deletion of the two dT_{+4} methyl groups resulted in a complex that had approximately 1.4 kcal/mol (~0.7 kcal/mol per methyl group) less free energy of binding than the native sequence. Although this value is somewhat below the value of 1.5 kcal/mol reported by Goeddel et al. (1977) for methyl groups in the *lac* repressor, it still represents a significant difference in apparent free energy of binding. The lower value may reflect the protein's ability to compensate for the loss of the methyl groups by maintaining significant van der Waals contacts to the C_5-H at the dT_{+4} site.

The two purine-containing complexes (dP_{-4} and dP_{+5}) exhibited apparent free energies of binding that were approximately 0.5 and 0.8 kcal/mol more stable, respectively, than the native complex. This increase in stability may result from some reorientation of the protein in the dP-containing sequences vs the native and may reflect the protein's ability to take advantage of additional van der Waals interactions, not available in the native sequence, in order to enhance overall binding effects. Additionally, it is possible that the loss of two hydrogen bonds in the dP-dT-containing operators enhances the flexibility of the sequences and permits distortion of the operator with little energy input in such a fashion as to permit enhanced affinity. Deletion of the adenine amino group at dA_{-4} may also facilitate a stronger hydrogen bonding interaction to the carbonyl of dT_{+4}. However, a similar interaction at dT_{-5} seems at present unlikely. The adenine amino group itself does not appear to be a hydrogen bond donor in the complex so that its removal does not result in the loss of significant binding energy. The results of the dP-containing sequences suggest that an increase in the hydrophobic character or flexibility at specific sites within the sequence, relative to that of the native, can contribute significantly to the overall binding affinity.

Conclusions. The use of base analogues can be a valuable addition to other methods for examining solution binding by sequence-specific binding proteins. The results from the analogue sequences used in this study suggest that the structure of the dA-dT or dT-dA base pairs at positions −4/+4 and −5/+5, respectively, has relatively little effect upon the solution binding by the *trp* repressor, but the protein is very sensitive to the orientation of the amino and carbonyl functional groups at the −4/+4 positions, which are involved in the formation of an interbase hydrogen bond present in the major groove. The deletion of individual functional groups from the operator sequence suggests that the carbonyl at dT_{+4} is critical for formation of the high-affinity sequence-specific complex. Although this observation is not entirely consistent with the reported crystal structure (Otwinowski et al., 1989), it is supported by previous in vivo genetic assays (Bass et al., 1987). Additionally, the thymine methyl group at dT_{+4} and the N7 nitrogen of dA_{+5} appear to be critical contacts, and these results are generally supported by the reported crystal structure, and in part by the genetic studies. Two of the deletion modifications result in sequences that bind to the repressor with a higher affinities than that observed with the native sequence; this can be explained in that the functional groups lost are not critical for binding, and the creation of a more hydrophobic surface at these sites enhances van der Waals contacts between the protein and the nucleic acid.

Acknowledgments

We thank one of the referees for suggesting that the reduced apparent K_D values for the dP-dT-containing sequences (loss of the adenine amino group) might result from enhanced hydrogen bonding to the thymine carbonyl.

References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., & Harrison, S. C. (1988) *Science 242*, 899–907.
Aiken, C. R., & Gumport, R. I. (1991) *Methods Enzymol. 208*, 433–457.
Aiken, C. R., McLaughlin, L. W., & Gumport, R. I. (1991) *J. Biol. Chem. 266*, 19070–19078.
Anderson, J. E., Ptashne, M., & Harrison, S. C. (1987) *Nature 326*, 846–852.
Bartlett, P. A., & Marlowe, C. K. (1987) *Science 235*, 569–571.
Bass, S., Corrells, V., & Youderian, P. (1988) *Science 242*, 240–245.
Bass, S., Sugiono, P., Arvidson, P. W., Gunsalus, R. P., & Youuderian, P. (1987) *Genes Dev. 1*, 1565–572.
Benson, N., Adams, C., & Youderian, P. (1992) *Genetics 130*, 17–26.
Brennan, C. A., Van Cleve, M. D., & Gumport, R. I. (1986a) *J. Biol. Chem. 261*, 7270–7278.
Brennan, C. A., Van Cleve, M. D., & Gumport, R. I. (1986b) *J. Biol. Chem. 261*, 7278–7286.
Brennan, R. G., & Matthews, B. W. (1989) *J. Biol. Chem. 264*, 1903–1906.
Carey, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 975–979.
Carey, J. (1989) *J. Biol. Chem. 264*, 1941–1945.
Caruthers, M. H. (1980) *Acc. Chem. Res. 13*, 155–165.
Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Happ, C. S., Happ, E., & Gronenborn, A. M. (1988) *Biochemistry 27*, 4185–4197.
Coll, M., Wang, A. H. J., van der Marel, G. A., van Boom, J. H., & Rich, A. (1986) *J. Biomol. Struct. Dyn. 4*, 157–172.
Connolly, B. A., & Newman, P. C. (1989) *Nucleic Acids Res. 17*, 4957–4975.
Diekmann, S., & McLaughlin, L. W. (1988) *J. Mol. Biol. 202*, 823–834.
Diekmann, S., von Kitzing, E., McLaughlin, L. W., Ott, J., & Eckstein, F. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8257–8262.
Diekmann, S., Mazzarelli, J. M., McLaughlin, L. W., von Kitzing, E., & Travers, A. (1992) *J. Mol. Biol.* (in press).
Dwyer-Hallquist, P., Kezdy, F. J., & Agarwal, K. L. (1982) *Biochemistry 21*, 4693–4700.
Fersht, A. R. (1987) *Trends Biochem. Sci. 12*, 301–305.
Fersht, A. R., & Dingwall, C. (1979) *Biochemistry 18*, 1238–1245.
Fersht, A. R., Shindler, J. S., & Tsui, W.-C. (1980) *Biochemistry 19*, 5520–5524.
Fersht, A. R., Shi, J. P., Knill-Hones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., & Winter, G. (1985) *Nature 314*, 235–238.
Fliess, A., Wolfes, H., Rosenthal, A., Schwellnus, K., Blocker, H., Frank, R., & Pingoud, A. (1986) *Nucleic Acids Res. 14*, 3463–3474.
Gaffney, B. L., & Jones, B. A. (1984) *Tetrahedron 40*, 3–13.
Gildea, B., & McLaughlin, L. W. (1989) *Nucleic Acids Res. 17*, 2261–2281.
Goeddel, D. V., Yansura, D. G., & Caruthers, M. H. (1977) *Nucleic Acids Res. 4*, 3039–3055.
Goeddel, D. V., Yansura, D. G., & Caruthers, M. H. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 3578–3582.

Gunsalus, R. P., & Yanosfky, C. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 7117–7121.
Haran, T. E., Joachimiak, A., & Sigler, P. B. (1992) *EMBO J.* (in press).
Hunter, W. N., D'Estaintot, B. L., & Kennard, O. (1989) *Biochemistry 28*, 2444–2451.
Ikuta, S., Eritja, R., Kaplan, B. E., & Itakura, K. (1987) *Biochemistry 26*, 5646–5650.
Jeung-Hoi, H., Spolar, R. S., & Record, M. T., Jr. (1989) *J. Mol. Biol. 209*, 801–806.
Jiricny, J., Wood, S. G., Martin, D., & Ubasawa, A. (1986) *Nucleic Acids Res. 14*, 6579–6590.
Joachimiak, A., Kelly, R. L., Gunsalus, R. P., Yanofsky, C., & Sigler, P. B. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 668–672.
Jordan, S. R., & Pabo, C. O. (1988) *Science 242*, 893–899.
Jorgensen, W. L., & Pranata, J. (1990) *J. Am. Chem. Soc. 112*, 2008–2009.
Kim, Y., Grabel, J. C., Love, R., Greene, P. J., & Rosenberg, J. M. (1990) *Science 249*, 1307–1309.
Klig, L. S., Crawford, I. P., & Yanofsky, C. (1987) *Nucleic Acids Res. 15*, 5339–5351.
Lesser, D. R., Kurpiewski, M. R., & Jen-Jacobson, L. (1990) *Science 250*, 776–786.
Luisi, B. F., & Sigler, P. B. (1990) *Biochim. Biophys. Acta 1048*, 113–126.
Marmorstein, R. Q., & Sigler, P. B. (1989) *J. Biol. Chem. 264*, 9149–9154.
Marmorstein, R. Q., Sprinzl, M., & Sigler, P. B. (1991) *Biochemistry 30*, 1141–1148.
Matteucci, M., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.
Mazzarelli, J., Scholtissek, S., & McLaughlin, L. W. (1989) *Biochemistry 28*, 4616–4622.
McClairn, J. A., Fredrick, C. A., Wang, B., Greene, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526–1541.
McLaughlin, L. W., & Piel, N. (1984) in *Oligonucleotide Synthesis: A Practical Approach* (Gait, M. J., Ed.) pp 199–218, IRL Press, Oxford.
McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N., & Scholtissek, S. (1987) *Biochemistry 26*, 7238–7245.
Newman, P. C., Nwosu, V. U., Williams, D. M., Cosstick, R., Seela, F., & Connolly, B. A. (1990a) *Biochemistry 29*, 9891–9901.
Newman, P. C., Williams, D. M., Cosstick, R., Seela, F., & Connolly, B. A. (1990b) *Biochemistry 29*, 9902–9910.
Nwosu, V. U., Connolly, B. A., Halford, S. E. & Garnett, J. (1988) *Nucleic Acids Res. 16*, 3705–3720.
Olendorf, D. H., & Matthews, B. W. (1983) *Annu. Rev. Biophys. Bioeng. 12*, 259–284.
Ono, A., Masayuki, S. Ohtani, Y., & Ueda, T. (1984) *Nucleic Acids Res. 12*, 8939–8949.
Otwinowski, Z., Scheivitz, R. W., Zhang, R. G., Lauson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321–329.
Owens, S. L., & Bell, F. E. (1970) *J. Biol. Chem. 245*, 5515–5523.
Pabo, C. O., & Sauer, R. T. (1984) *Annu. Rev. Biochem. 53*, 293–321.
Robbins, M. J., Wilson, J. S., & Hansske, F. (1983) *J. Am. Chem. Soc. 105*, 4059–4065.
Rosenberg, J. M. (1991) *Curr. Opin, Struct. Biol. 1*, 104–113.
Seela, F., & Driller, H. (1986) *Nucleic Acids Res. 14*, 2319–2332.
Seela, F., & Kehne, A. (1987) *Biochemistry 26*, 2232–2238.
Seeman, N., Rosenberg, J. M., & Rich, A. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 804–808.
Staacke, D., Walter, B., Kisters-Woike, B. V., Wilcken-Bergmann, B., & Muller-Hill, B. (1990) *EMBO J. 9*, 1963-1967; 3023 (correction).
Street, I. P., Amrstrong, C. R., & Withers, S. G. (1986) *Biochemistry 25*, 6021–6027.
von Hippel, P. H., & Berg, O. G. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1608–1612.
Wells, T. N. C., & Fersht, A. R. (1986) *Biochemistry 25*, 1881–1886.
Wolberger, C., Dong, Y., Ptashne, M., & Harrison, S. C. (1988) *Nature 335*, 789–795.
Yolov, A. A., Vinogradova, M. N., Gromova, E. S., Rosenthal, A., Cech, D., Veiko, V. P., Metelev, V. G., Kosykyh, V. G., Buryanov, Ya. I., Bayev, A. A., & Shaborova, Z. A. (1985) *Nucleic Acids Res. 13*, 8983–8998.