Bioorganic Chemistry **27,** 215–226 (1999)
Article ID bioo.1998.1129, available online at http://www.idealibrary.com on IDEAL®

# Probing the Interface of the *trp* Repressor–Operator Complex Using Operator Sequences Containing Isosteric Base-Pair Analogues

Stephen A. Smith and Larry W. McLaughlin[1]

*Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02167*

*Received July 13, 1998*

A series of base pair analogues, dI-dM and dD-dU, have been used in place of the native dA-dT and dG-dC residues to probe the interface between the *trp* repressor and its operator sequence. Both analogues maintain the interstrand hydrogen bond character that correlates with the corresponding native base pair, but alters the relative positions of the amino and carbonyl residues of the base pairs that extend into the major groove of B-form DNA. None of the sequences tested exhibit native-like binding with the *trp* repressor. Although functional group contacts between complementary hydrogen bond donors and acceptors could account for these differences, such a scenario is not consistent with the published crystal structures. Differences in binding by these analogue sequences is suggested to result from incremental differences in the van der Waals contacts at the interface between the protein and the nucleic acid. Although this surface would appear to be nonspecific in nature, previous studies have indicated that the state of hydration in the major groove for the free and complexed operators is in part conserved. These bound water molecules alter the nature of the van der Waals surface and could function as an extension of the van der Waals surface of the major groove. Either of the base pair analogues, dI-dM or dD-dU, results in either the loss of a conserved site of hydration, or the introduction of a new hydration site. Either effect would alter the van der Waals surface relative to the native operator and result in reduced affinity between the protein and nucleic acid. Additional analogues designed to alter the nature of the van der Waals surface by removing polar functional groups (adenine → purine) result in some cases in higher affinity binding than is observed with the native operator sequence. 

## INTRODUCTION

Sequence-specific binding between proteins and nucleic acids can generally be explained in terms of a series of specific hydrogen bonding, ionic and hydrophobic interactions that occur largely at the interface between the two molecules. Many proteins recognize their target sequences by what has been termed a “direct readout” mechanism in which the base residues are probed for complementary hydrogen-bonding sites by residues present on the protein. A number of crystal structures,

[1] To whom correspondence and reprint requests should be addressed. Fax: (617) 552-2705. E-mail: larry.mclaughlin@bc.edu.



0045-2068/99 $30.00

particularly of bacterial phage repressor proteins (*1–4*) have generally confirmed this mechanism for target identification. "Indirect readout" of target sequences has been suggested (*5*) as an alternate mechanism for sequence targeting and the formation of high affinity complexes. With indirect recognition, the functional groups of the base residues contribute to the overall geometry of the target sequence, and complementary van der Waals surfaces on both the protein and the nucleic acid permit intimate association of the two macromolecules, but few direct hydrogen bonding interactions are used to "read" the DNA sequence. The two reported structures of the *trp* repressor–operator complex represent examples of the indirect readout mechanism (*5,6*).

The interface between the *trp* repressor dimer and operator sequence as described by the crystal structure analysis (*5*) has but a single direct contact in each half-site between a base residue and an amino acid side chain; $Arg_{69}$ makes a bidentate hydrogen-bonding interaction with $G_{-9}$. But even this interaction may only contribute minimally to target recognition since the symmetrical mutations of $C_{+9}-G_{-9} \rightarrow G_{+9}-C_{-9}$ has relatively little effect upon repressor binding monitored *in vivo* (*7*). A number of direct contacts to the phosphodiester residues are also present in the complex (*5*) that afford complex stabilization (*8*). In addition to these direct interactions, a number of water-mediated contacts are present involving both the base residues (*5,9,10*) and the internucleotide phosphates (*5,8*). Base analogue studies have assisted in illustrating that such interactions contribute to high affinity binding in this complex (*8,10,11*). But in addition to both the direct and the water-mediated contacts that can be used to specify interactions with individual functional groups present in the operator sequence, there is a large nonspecific (at least in terms of functional group contacts) contact surface on the DNA that is rendered inaccessible to solvent upon binding by the protein (*5*). The 2,900 $A^2$ of contact surface area at the protein–nucleic acid interface arises in part from the flexibility of both the DNA and the protein that appears to permit some molding of one surface to the other. Although generally inaccessible to solvent, some functional groups at this interface remain hydrated by specifically bound water molecules (*12*)—some assist in the noted water mediated hydrogen bonding interactions, while others may simply extend or enhance the nature of the surface contacted. This large van der Waals contact area may be critical to sequence recognition since the contours of the major groove van der Waals surface would alter significantly with sequence mutations (even in the absence of functional group contacts), and such alterations to the interface could reduce affinity and thus sequence recognition.

DNA sequences containing nucleoside analogues are valuable tools to probe the nature of the interface between the protein and the nucleic acid (*13,14*). A number of reports describe the use of such incrementally altered sequences to alter single functional groups and probe their contribution to complex stability. In the present study we have used two base pair analogues that are structural mimics of the native dA-dT and dG-dC base pairs, but alter the nature of the functional groups present in the major groove. Such analogues can both incrementally change the hydration state of the major groove, as well as alter the nature of the van der Waals surface. A series of analogue operator sequences have been prepared to probe the nature of the protein–nucleic acid interface in the *trp* repressor–operator complex.

## EXPERIMENTAL

Oligodeoxynucleotides were synthesized using nucleoside phosphoramidites on an Applied Biosystems 381A DNA synthesizer. The four fully protected common 2′-deoxynucleoside phosphoramidites containing aryl- or isobutyrylamides were purchased from Cruachem through Fisher Chemical Co. The phosphoramidite derivatives of 2′-deoxyinosine (dI), 5-methyl-2′-deoxycytidine (dM), and 2′-deoxyuridine (dU) were prepared by standard procedures, beginning with the native nucleoside, or purchased from Glen Research (Sterling, VA). 2,6-Diaminopurine-2′-deoxyriboside (dD) was prepared as described (*15*) and then converted to the desired phosphoramidite derivative. The *trp* repressor was isolated from an overproducing strain provided as a gift from Professor P. Sigler (Yale University). Calf intestinal alkaline phosphatase was a product of Boerhinger-Mannheim (Germany).

*Oligonucleotide synthesis.* The oligonucleotides were synthesized on controlled pore glass supports using phosphoramidite chemistry (*16*) and an Applied Biosystems 381A DNA synthesizer. After deprotection by base (conc. 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%, over 40 min). 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 methanol/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-$\mu$l reaction mixture containing 0.5 $A_{260}$ unit of oligomer in 100 mM Tris–HCl, pH 8.0, 20 mM $MgCl_2$, 2 units of snake venom phosphodiesterase and 1 unit of bacterial alkaline phosphatase was incubated at 37°C overnight. A 5-$\mu$l aliquot was analyzed by HPLC using a 4.6 × 250 mm column of ODS-Hypersil in 20 mM potassium phosphate, 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; dT, 11.9; dA, 15.5; dD, 21.5 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 of 1 $\mu$M. Absorbance and temperature values were measured with an AVIV 14DS UV/Visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 1°C steps (from 0 to 95°C) and when thermal equilibrium was reached, temperature and absorbance data were collected. $T_m$ values were determined both from first-order derivatives and by graphical analysis of the absorbance vs temperature plots.

*Radioisotopic labeling.* Each 20 mer was 5′-end labeled with [$\gamma$-$^{32}$P]ATP as follows: A 200-$\mu$l 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 $\mu$Ci[$\gamma$-$^{32}$P]ATP, 5 $\mu$g/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,

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 $\mu$M.

*Alkaline phosphatase binding analysis.* The assay as described by Marmorstein *et al.* (1991) (*18*) 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, 250 mM sodium chloride. Approximately 1 mg of repressor in 30 $\mu$l of storage buffer (0.5 mM sodium phosphate, 250 mM sodium chloride, pH 7.0) was diluted to a final volume of 500 $\mu$l 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 determined spectrophotometrically at 280 nm ($\varepsilon = 1.48 \times 10^4\ M^{-1}cm^{-1}$). 100 units of alkaline phosphatase (1 U/$\mu$l) were also dialyzed under the same conditions.

*Alkaline phosphatase reaction in the absence of repressor.* A 200-$\mu$l reaction solution at 22°C containing 20 nM of duplex operator in l0 mM Tris–HCl, pH 7.4, 250 mM sodium chloride, and containing 10 $\mu$g/ml BSA was initiated by the addition of 2.5 or 5.0 units of alkaline phosphatase. Aliquots of 20–30 $\mu$l were removed at l5-s intervals (usually for a 2-min total period) and the reaction was quenched by addition of the aliquot into 25 $\mu$l of 0.15 M potassium phosphate, pH 7.0, containing 100 $A_{260}$ units/ml 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 determined in 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 $\mu$l reaction solution contained 0.1, 1.0, or 10 $\mu$M repressor (dimer) and 0.5 mM L-tryptophan corepressor and this mixture incubated 30 min at 22°C prior to the addition of the phosphatase. Control reactions were also performed in absence of the corepressor. The time intervals 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 nonduplex structures (e.g., hairpins). The presence of such structures is not generally considered in assays using self-complementary oligodeoxynucleotides, but 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 relationship: $K_D = [R]/(k_i/k_i' - 1)$, as has been described previously. 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

In order to probe the nature of the interface between the *trp* repressor and its operator DNA sequence, two base pair analogues were chosen that could mimic the structure of either of the native dA-dT and dG-dC base pairs and also alter the nature of the functional group character, thus the van der Waals surface present in the major groove. The base pair analogue for dA-dT used in this study is that formed from 2′-deoxyinosine (dI) and 2′-deoxy-5-methylcytidine (dM). The dI-dM base pair has structural features that are similar to dA-dT (compare Figs. 1a and 1b). Two interstrand hydrogen bonds are present in both base pairs, and the pyrimidine methyl group is retained in the analogue base pair to mimic that of the native base pair. The major difference between the analogue and the native base pair is the relative orientation of the amino and carbonyl groups that form an interstrand hydrogen bond in the major groove; their orientation is reversed. This change in relative orientation of two functional groups alters the position of potential hydrogen bonding sites, the nature of the van der Waals surface, and/or hydration characteristics within the groove relative to those properties of the native dA-dT base pair. In fact, when viewed from the perspective of the major groove, the dI-dM base pair maintains a pattern of functional groups that is, with the exception of the pyrimidine methyl group, similar to that of a dG-dC base pair. In this respect, the analogue base pair is structurally similar to dA-dT but provides a van der Waals surface in the major groove that is

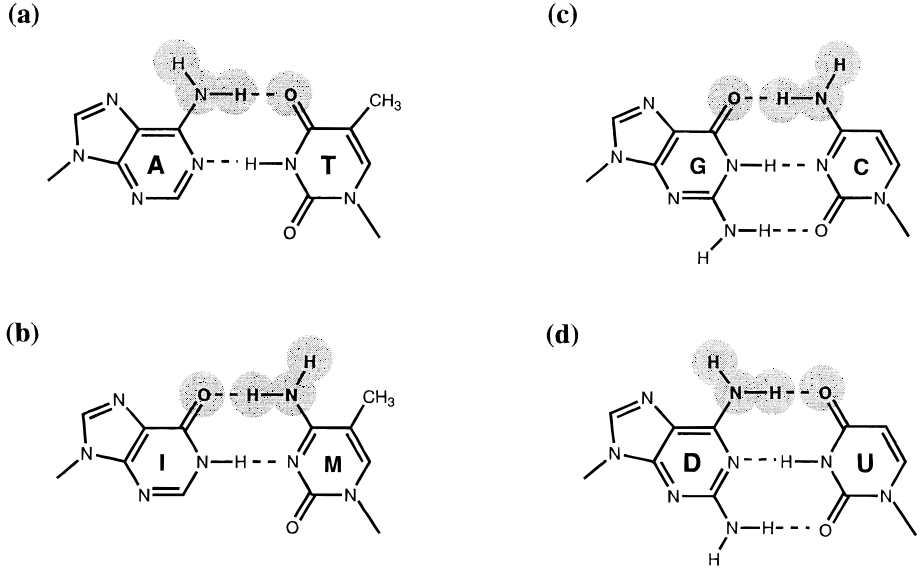


**FIG. 1.** Structures of the (a) dA-dT, (b) dI-dM, (c) dG-dC, and (d) dD-dU base pairs.

also similar to that of dG-dC. In a similar manner, the base pair formed from 2,6-diaminopurine-2′-deoxyriboside (dD) and 2′-deoxyuridine (dU) (Fig. 1d) provides a structural mimic of the native dG-dC base pair (Fig. 1c), but when viewed from the major groove has many of the functional group characteristics of a dA-dT base pair.

A second series of analogue operator sequences were prepared using the purine base, an analogue of adenine from which the exocyclic amino group has been deleted. One of the adenine amino groups has been implicated in a contact with the repressor; a water-mediated interaction involving $A_{-7}$. Additionally, all of the adenine exocyclic amino groups are hydrated in both the free operator sequence and in the bound repressor–operator complex (*12*). Deletion of this amino group at selected sites could incrementally alter the nature of the major groove surface by eliminating both the amino group and altering the van der Waals surface and/or hydration effects at a given site.

*Sequence preparation and characterization.* DNA sequences twenty residues in length were synthesized using the appropriate nucleoside phosphoramidites and purified by HPLC. The presence of the two modified nucleosides of each analogue base pair could be confirmed by digestion of a small quantity of oligonucleotide and analysis of the resulting mixture by HPLC. Figure 2 illustrates two examples, one for sequence containing a dI-dM base pair (Fig. 2a), and a second analysis for a sequence containing a dD-dU base pair (Fig. 2b). Initial characterization of the 20-mer operator sequences was by simple measurements of $T_m$ values. These values were obtained in the buffer/salt conditions used for analysis of repressor binding. Substitution of the base pair analogue dI-dM for dA-dT at two sites in the 20-mer sequence (one in each half-site) had very little effect upon the $T_m$ values for the sequences (Table 1). This is consistent with the fact that both the analogue and the native base pairs can form two interstrand hydrogen bonds and additionally suggests that base stacking effects for the dI vs dA and for dM vs dT are not dramatically different. By comparison, the substitution of dD-dU for dG-dC in each case resulted in roughly a 7°C decrease in $T_m$. Although the dD-dU base pairs can form three interstrand hydrogen bonds, as do dG-dC base pairs, it is well-documented that DNA duplexes containing dD-dU base pairs have reduced $T_m$ values, relative to those containing dG-dC base pairs (*15*), and this phenomenon can be understood in terms of the orientation of the dipoles generated upon base pairing (*17*). By comparison, DNA duplexes containing purine for adenine substitutions generate dP-dT base pairs, but owing to the loss of the exocyclic amino group only a single interstrand hydrogen bond is present. A corresponding reduction in $T_m$ values would be expected, and the observations for the purine-containing sequences are consistent with this expectation (Table 2).

*Protein binding studies.* We have employed a phosphatase protection assay described in other studies (*8,10,11,18*) to measure relative binding constants for the native and base analogue-substituted operator sequences. This assay relies upon binding by the *trp* repressor to the 20-mer operator sequence for protection of a terminal (radiolabeled) phosphate from cleavage by alkaline phosphatase. By comparing the rate of phosphatase activity in the absence and in the presence of the repressor, an apparent dissociation constant can be obtained based upon substrate depletion kinetics. The results from this assay are quite sensitive to the conditions of the assay, particularly