(1987) *J. Am. Chem. Soc. 109*, 3461–3464.
Haseltine, J. N., & Danishefsky, S. J. (1989) *J. Am. Chem. Soc. 111*, 7638–7640.
Lee, M. D., Dunne, T. S., Change, C. C., Ellestad, G. A., Siegel, M. M., Morton, G. O., McGahren, W. J., & Borders, D. B. (1987) *J. Am. Chem. Soc. 109*, 3466–3468.
Lockhart, T. P., Comita, P. B., & Bergman, R. G. (1981) *J. Am. Chem. Soc. 103*, 4082–4090.
Long, B. H., Golik, J., Forenza, S., Ward, B., Rehfuss, R., Dabrowiak, J. C., Catino, J. J., Musial, S. T., Brookshire, K. W., & Doyle, T. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2–6.
Magnus, P., Lewis, R. T., & Huffman, J. C. (1988) *J. Am. Chem. Soc. 110*, 6921–6923.
Magnus, P., Fortt, S., Pitterna, T., & Snyder, J. P. (1990) *J. Am. Chem. Soc. 112*, 4986–4987.
Mantlo, N. B., & Danishefsky, S. J. (1989) *J. Org. Chem. 54*, 2781–2783.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Nagata, R., Yamanaka, H., Okazaki, E., & Saito, I. (1989) *Tetrahedron Lett. 30*, 4995–4998.
Nicolaou, K. C., Ogawa, Y., Zuccarello, G., & Kataoka, H. (1988a) *J. Am. Chem. Soc. 110*, 7247–7248.
Nicolaou, K. C., Zuccarello, G., Ogawa, Y., Schweiger, E. J., & Kumazawa, T. (1988b) *J. Am. Chem. Soc. 110*, 4866–4868.
Shiraki, T., & Sugiura, Y. (1990) *Biochemistry 29*, 9795–9798.
Snyder, J. P. (1989) *J. Am. Chem. Soc. 111*, 7630–7632.
Snyder, J. P. (1990) *J. Am. Chem. Soc. 112*, 5367–5369.
Sugiura, Y., Uesawa, Y., Takahashi, Y., Kuwahara, J., Golik, J., & Doyl, T. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7672–7676.
Sugiura, Y., Shiraki, T., Konishi, M., & Oki, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3831–3835.
Tomioka, K., Fujita, H., & Koga, K. (1989) *Tetrahedron Lett. 30*, 851–854.
Uesawa, Y., Kuwahara, J., & Sugiura, Y. (1989) *Biochem. Biophys. Res. Commun. 164*, 903–911.
Zein, N., Sinha, A. M., McGahren, W. J., & Ellestad, G. A. (1988) *Science 240*, 1198–1201.
Zein, N., McGahren, W. J., Morton, G. O., Ashcroft, J., & Ellestad, G. A. (1989) *J. Am. Chem. Soc. 111*, 6888–6890.

Binding of Triple Helix Forming Oligonucleotides to Sites in Gene Promoters†

Ross H. Durland,‡ Donald J. Kessler,‡ Sandy Gunnell,‡ Madeleine Duvic,§ B. M. Pettitt,∥ and Michael E. Hogan*,‡

Center for Biotechnology, Baylor College of Medicine, Woodlands, Texas 77381, Department of Dermatology and Internal Medicine, University of Texas Medical School, Houston, Texas 77030, Department of Chemistry, University of Houston, Houston, Texas 77204-5641

Received May 14, 1991; Revised Manuscript Received July 9, 1991

ABSTRACT: A class of triplex-forming oligodeoxyribonucleotides (TFOs) is described that can bind to naturally occurring sites in duplex DNA at physiological pH in the presence of magnesium. The data are consistent with a structure in which the TFO binds in the major groove of double-stranded DNA to form a three-stranded complex that is superficially similar to previously described triplexes. The distinguishing features of this class of triplex are that TFO binding apparently involves the formation of hydrogen-bonded G·GC and T·AT triplets and the TFO is bound antiparallel with respect to the more purine-rich strand of the underlying duplex. Triplex formation is described for targets in the promoter regions of three different genes: the human c-*myc* and epidermal growth factor receptor genes and the mouse insulin receptor gene. All three sites are relatively GC rich and have a high percentage of purine residues on one strand. DNase I footprinting shows that individual TFOs bind selectively to their target sites at pH 7.4–7.8 in the presence of millimolar concentrations of magnesium. Electrophoretic analysis of triplex formation indicates that specific TFOs bind to their target sites with apparent dissociation constants in the 10^{-7}–10^{-9} M range. Strand orientation of the bound TFOs was confirmed by attaching eosin or an iron-chelating group to one end of the TFO and monitoring the pattern of damage to the bound duplex DNA. Possible hydrogen-bonding patterns and triplex structures are discussed.

Triple-helical nucleic acid structures can be formed from synthetic polymers, as the result of internal disproportionation of polypurine arrays in duplexes, or by the binding of short oligonucleotides to purine-rich duplex DNA segments (Felsenfeld et al., 1957; Felsenfeld & Rich, 1957; Lipsett, 1963, 1964; Howard et al., 1964; Morgan & Wells, 1968; Thiele & Guschlbauer, 1971; Haas & Guschlbauer, 1976; Marck & Thiele, 1978; Lyamichev et al., 1985, 1986, 1988; Moser & Dervan, 1987; Broitman et al., 1987; Kohwi & Kohwi-Shigematsu, 1988; Htun & Dahlberg, 1988; Harvey et al., 1988; Hanvey et al., 1988; Fedorova et al., 1988; Praseuth et al., 1988). Several groups have described and studied homopolymeric DNA triplexes such as poly(dT)·poly(dA)·poly(dT) and poly(dC⁺T)·poly(dGA)·poly(dTC) (Riley et al., 1966; Arnott & Selsing, 1974; Arnott et al., 1976; Lee et al., 1979). More recently, synthetic oligonucleotides composed of C and/or T residues have been shown to form stable triple helices by binding to specific oligopurine regions in double-stranded DNA (Moser & Dervan, 1987; Lyamichev et al., 1988; Fe-

†Supported by grants from the Office of Naval Research and Triplex Pharmaceutical Corporation to M.E.H., from the NIH to M.D. and M.E.H., and from the Texas Advanced Technology Program to M.D. and M.E.H.

‡Baylor College of Medicine.
§University of Texas Medical School.
∥University of Houston.

0006-2960/91/0430-9246$02.50/0

dorova et al., 1988; Praseuth et al., 1988; Maher et al., 1989; Hanvey et al., 1989). These structures are all thought to involve binding of the third strand to the major groove of the duplex DNA. Binding results in the formation of hydrogen-bonded T·AT and C^+·GC triplets in which pyrimidines in the third strand are Hoogsteen-bonded to the purines of the duplex DNA. In the case of third-strand C residues, triplex formation requires protonation at nitrogen 3, indicated by C^+ (Rajagopal & Feigon, 1989). Triplexes containing C^+·GC triplets are highly sensitive to pH and generally require acidic conditions to form. In all cases where it has been explicitly examined, the backbone orientation of the third strand is parallel to that of the purine-rich strand of the underlying duplex (Moser & Dervan, 1987; Fedorova et al., 1988; Praseuth et al., 1988).

At present, potential targets for triplex formation by oligonucleotides are limited to regions with a high percentage of purine residues on one strand. Despite this limitation, triple helix forming oligonucleotides have tremendous potential utility because they provide a simple way to design compounds that bind to double-stranded DNA in a sequence-specific manner. Thus, triple helix forming oligonucleotides have been shown to be useful as competitors for DNA-binding proteins and as site-specific DNA damage or cleavage agents (Le Doan et al., 1987; Moser & Dervan, 1987; Fedorova et al., 1988; Praseuth et al., 1988; Maher et al., 1989; Hanvey et al., 1989). Because of the requirement for protonation of third-strand C residues, the optimum pH for triplex formation by such oligonucleotides is well below 7.0. The use of 5-methylcytosine in place of cytosine in the third strand increases the apparent pK for triplex formation about 0.5 pH units and permits detectable triplex formation up to pH 7.4 (Povsic & Dervan, 1989). However, recent work shows that the dissociation rate and the equilibrium dissociation constant for triplexes involving 5-methylcytosine-containing oligonucleotides is still very sensitive to pH over the range of 6.8–7.2 (Maher et al., 1990). Since many biological processes have an optimum in the pH range of 7.4–7.8 or higher, alternative triple helix forming compounds that are not limited to acidic pH are desirable.

In a previous study, it was shown that, at physiological pH, a discrete 27 base long oligonucleotide has the capacity to form a stable triplex at a target site within the promoter of the human c-*myc* gene, 116–142 base pairs upstream from the principle mRNA start site (Cooney et al., 1988). This region is also the binding site for several cellular factors, at least one of which is required for in vitro initiation of mRNA synthesis from the c-*myc* promotor (Boles & Hogan, 1987; Postel et al., 1989; Davis et al., 1989). In agreement with that observation, it was found that an oligonucleotide that binds to the -116/-142 site is capable of inhibiting c-*myc* transcription in vitro (Cooney et al., 1988).

In contrast to other triplexes in which the third strand consists of C and T residues, the oligonucleotides that bound the c-*myc* site were composed primarily of G residues. It was suggested that triplex formation was based on the formation of G·GC, T·AT, and A·AT triplets. Triplexes of the form poly(dG)·poly(dG)·poly(dC) have been observed previously with homopolymers and in supercoiled plasmid DNA (Lipsett, 1963, 1964; Marck & Thiele, 1978; Kohwi & Kohwi-Shigematsu, 1988). There is also evidence that triplexes such as poly(rA)·poly(rA)·poly(rU) and poly(rA)·poly(dA)·poly(dT) can form under certain circumstances (Broitman et al., 1987; Letai et al., 1988). The orientation of the third strand of the triplexes in these studies was not explicitly determined. In the work of Cooney et al. (1988), it was assumed that the third strand bound parallel to the purine-rich strand of the duplex, similar to the known orientation of triplexes containing C^+·GC and T·AT triplets. However, additional experiments with oligonucleotides that bind to the c-*myc* site indicated that the third strand in fact binds antiparallel (Durland et al., 1990). Moreover, a recent study of triplex formation involving a G-rich oligonucleotide binding to a site similar to the c-*myc* site found that the third strand bound antiparallel (Beal & Dervan, 1991).

Here we describe additional triplex-forming oligonucleotides (TFOs)[1] that bind to the c-*myc* site as well as to sites in the promoters of the human epidermal growth factor receptor (EGFR) gene and the mouse insulin receptor (IR) gene. In all cases, the TFOs consist exclusively of G and T residues and are designed to bind antiparallel to the purine-rich strand of the duplex to form G·GC and T·AT triplets. Antiparallel binding is confirmed directly by attaching reactive groups to the 3′ end of the TFOs. Binding occurs at physiological pH in the presence of 5–20 mM magnesium at 20–37 °C. Binding is sequence specific in terms of both the target duplex and the single-stranded TFO. Apparent equilibrium dissociation constants are in the 10^{-7}–10^{-9} M range. Possible hydrogen-bonding schemes for the predicted base triplets are presented.

Materials and Methods

Oligonucleotide Synthesis and Modification. All oligodeoxyribonucleotides used in this study were synthesized on a Milligen 7500 DNA synthesizer, using standard phosphoramidite methods. Columns and reagents were obtained from Milligen/Biosearch with the exception of acetonitrile, which was obtained from Baxter. Some oligonucleotides were synthesized with the 3′-Amino Modifier (Glen Research), which results in the covalent attachment of a propanolamine linker to the 3′-hydroxyl group (see the legend to Figure 1). Oligonucleotides containing a 3′-amine group are indicated by the prefix "3′-A-". Following synthesis and deprotection, tritylated oligomers were purified from nontritylated failure sequences by reverse-phase HPLC. In some cases, denaturing polyacrylamide gel purification was also employed.

Eosin-linked TFOs were prepared by reacting 3′-amine modified TFOs with eosin-5-isothiocyanate (Molecular Probes) in sodium bicarbonate buffer at pH 9.5. Diethylenetriaminepentaacetic acid (DTPA) modified TFOs were prepared by reacting the anhydride of DTPA (Pierce) with 3′-amine TFOs under similar conditions. Eosin- and DTPA-linked TFOs are indicated by the prefixes "3′-E-" and "3′-D-", respectively.

Preparation of End-Labeled Fragments for DNase I and Chemical Mapping. For DNase I, eosin, and DTPA experiments on the human c-*myc* site, a 231 bp *Hin*dIII to *Bgl*I fragment was isolated from the plasmid pRD1000. This plasmid contains base pairs −113 to −155 of the c-*myc* site (numbered as in Figure 1A) inserted between the *Eco*RI and *Bam*HI sites of pUC19 (R. Durland, unpublished results). Experiments on the EGFR site employed a 231 bp *Hin*dIII to *Bgl*II fragment from plasmid pGER9 (Johnson et al., 1988). Experiments with the IR site used a 313 bp *Hin*dIII to *Taq*I fragment from plasmid pIRCAT+ (Sibley et al., 1989). In

[1] Abbreviations: T, deoxythymidine; A, deoxyadenosine; C, deoxycytosine; C^+, deoxycytosine protonated at N3; G, deoxyguanosine; X·YZ (where X, Y, and Z are nucleotides), a nucleotide base triplet in which Y and Z are Watson–Crick paired and X is hydrogen bonded to Y in a non-Watson–Crick manner; TFO, triplex-forming oligodeoxyribonucleotide; EGFR, human epidermal growth factor receptor gene; IR, mouse insulin receptor gene; DTPA, diethylenetriaminepentaacetic acid; EDTA, ethylenediaminetetraacetic acid; K_D, equilibrium dissociation constant; DTT, dithiothreitol; bp, base pair.

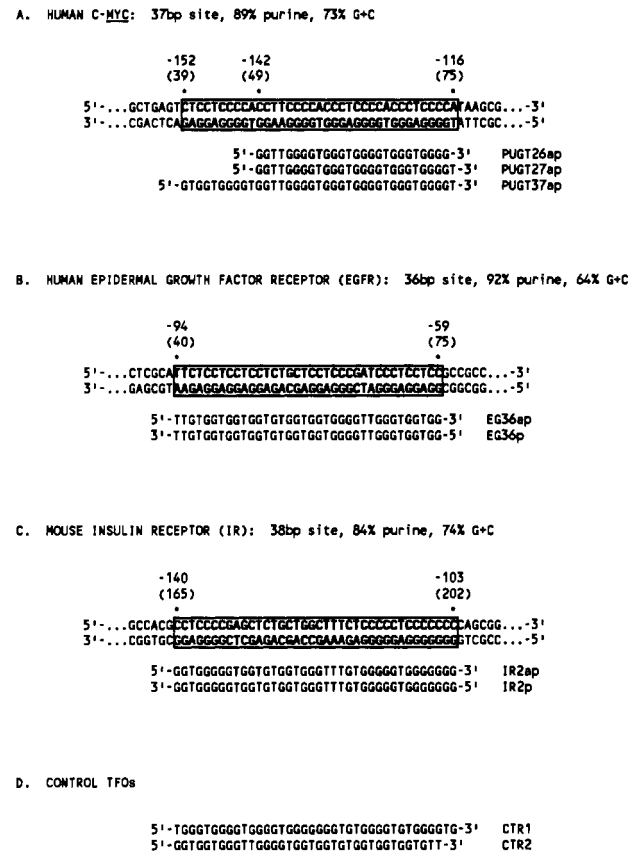


FIGURE 1: Nucleotide sequences of the triplex targets and TFOs used in this study. Targets for triplex formation are enclosed in boxes. Values for percent purine and percent G+C refer to the boxed regions. Note that some TFOs have been derivatized with propanolamine attached through a 3'-phosphodiester linkage (3'-O-PO_2-O-CH_2-CHOH-CH_2-NH_3^+). Such TFOs include the prefix 3'-A- in their designation, e.g., 3'-A-PUGT37ap, but are otherwise identical with the underivatized TFO. Attachment of eosin or DTPA to the 3'-amine group of a TFO is indicated with the prefix "3'-E-" or "3'-D-", respectively (see Materials and Methods for details). (A) Human c-*myc* gene. The double-stranded sequence is a portion of the promoter region of c-*myc*. Negative numbering refers to the distance upstream from the P1 transcriptional start site (Battey et al., 1983). Numbers in parentheses refer to the distance from the labeled end of the fragment used in the experiments in Figures 2A, 5A, and 6A. The sequences of the three c-*myc* specific TFOs are shown below the target, positioned in the expected binding orientation. (B) Human epidermal growth factor receptor (EGFR) gene. The target sequence is taken from the promoter region, with the negative numbers representing distance from the major transcription start site defined by Johnson et al. (1988). Numbers in parentheses represent distance from the labeled end of the fragment used in Figures 2B, 5B, and 6B. The sequences of EG36ap and EG36p are shown below. (C) Mouse insulin receptor (IR) gene. The target sequence is a region upstream of the promoter, as defined by Sibley et al. (1989). Negative numbers refer to distance from the downstream transcriptional start mapped in that study. Numbers in parentheses refer to the distance from the labeled end of the fragment used in Figures 2C, 5C, and 6C. Sequences of IR2ap and IR2p are shown below. (D) Sequences of control TFOs used in this study.

each case, the isolated fragment was 3'-end-labeled at the *Hin*dIII site, which labeled the purine-rich strand of the triplex target. Distances between the labeled ends and the target sequences in each fragment are indicated in Figure 1A–C. These three labeled fragments were used for the experiments in Figures 2, 5, and 6.

DNase I Footprinting Analysis. The appropriate end-labeled duplex fragment was incubated with or without TFO in a buffer consisting of 10 mM Tris-HCl, pH 7.4, 20 mM $MgCl_2$, and 10% sucrose at 37 °C for 60 min. DNase I was added at 0.125 units/mL, and the incubation was continued for 12 min. Reactions were stopped by the addition of excess EDTA and calf thymus DNA, electrophoresed on denaturing polyacrylamide gels, dried, and autoradiographed.

Band Shift Analysis. Band shift experiments were performed by incubating trace concentrations of end-labeled target duplex with increasing concentrations of TFO in a buffer consisting of 10 mM Tris-HCl, pH 7.4, 10% sucrose, and 5 or 20 mM $MgCl_2$. In some cases, 1 mM spermine was also included (see figure legends). Incubation was at 20 or 37 °C for 60 min. Samples were electrophoresed through polyacrylamide gels buffered with 89 mM Tris, 89 mM boric acid, and 5 or 20 mM $MgCl_2$, corresponding to the magnesium concentration of the sample. Gels were then dried and autoradiographed. The experiment described in the legend to Figure 4 is essentially identical, except that the TFO was end labeled and the target duplex fragments were not.

Eosin-Linked TFO Mapping of Third-Strand Orientation. Eosin-linked TFOs were incubated with end-labeled target duplexes in 10 mM Tris-HCl, pH 7.4, and 20 mM $MgCl_2$ at 37 °C for 60 min in small glass vials. Samples were then irradiated with the green output of an argon ion laser (Spectra-Physics 2020, 450–600 nm), typically at 2W/cm^2 for 2 min. During irradiation, the samples were cooled with a stream of chilled air to prevent excess heating. Samples were treated with 1 M piperidine to cleave damaged sites and electrophoresed.

DTPA-Linked TFO Mapping of Third-Strand Orientation. End-labeled target duplexes were mixed with various concentrations of DTPA-linked TFOs, followed by $FeCl_3 \cdot 6H_2O$ at one equivalent per equivalent of DTPA-TFO. Reaction buffers contained 10 mM Tris-HCl, pH 7.4, and 20 mM $MgCl_2$. Samples were incubated for 90 min at 37 °C, dithiothreitol was added to 20 mM, and incubation was continued at 37 °C overnight. Reactions were stopped by the addition of excess EDTA and calf thymus DNA, followed by ethanol precipitation, denaturation, and electrophoresis.

RESULTS

Triplex target sites and the relevant TFOs are described in Figure 1. All three sites are highly enriched for purine residues on one strand, ranging from 84 to 92% purines. In addition, all three sites are relatively GC rich, with G+C values of 64 to 74%. It should be noted that all three sites are taken from the natural promoter sequences of the indicated genes, and although they are purine-rich, they are not homopurine runs. These sites were selected in part to test the ability of the TFOs described here to recognize and bind to naturally occurring sequences in potentially important gene regions. All assays of triplexes in this study were performed at pH 7.4–7.8 in the presence of 5 or 20 mM magnesium chloride, which was previously shown to be essential for triplex formation (Cooney et al., 1988).

Human c-myc. Triplex formation by G-rich oligonucleotides was first described by Cooney et al. (1988) for a target in the promoter region of the human c-*myc* gene (Figure 1A). This target site has been previously shown to be important in c-*myc* expression in vivo on the basis of deletion analysis (Hay et al., 1987) and is known to be the binding site for a protein factor that is essential for c-*myc* transcription in vitro (Postel et al., 1989). Binding of TFOs to the c-*myc* target can be assessed by DNase I footprinting in a manner similar to DNA-binding proteins. This is shown for the TFOs 3'-A-PUGT27ap and 3'-A-PUGT37ap in Figure 2A. 3'-A-PUGT37ap is designed to bind to the 37 bp target sequence antiparallel to the purine-rich strand, while 3'-A-PUGT27ap is targeted to a 27 bp portion of that sequence. Note that the designation 3'-A-

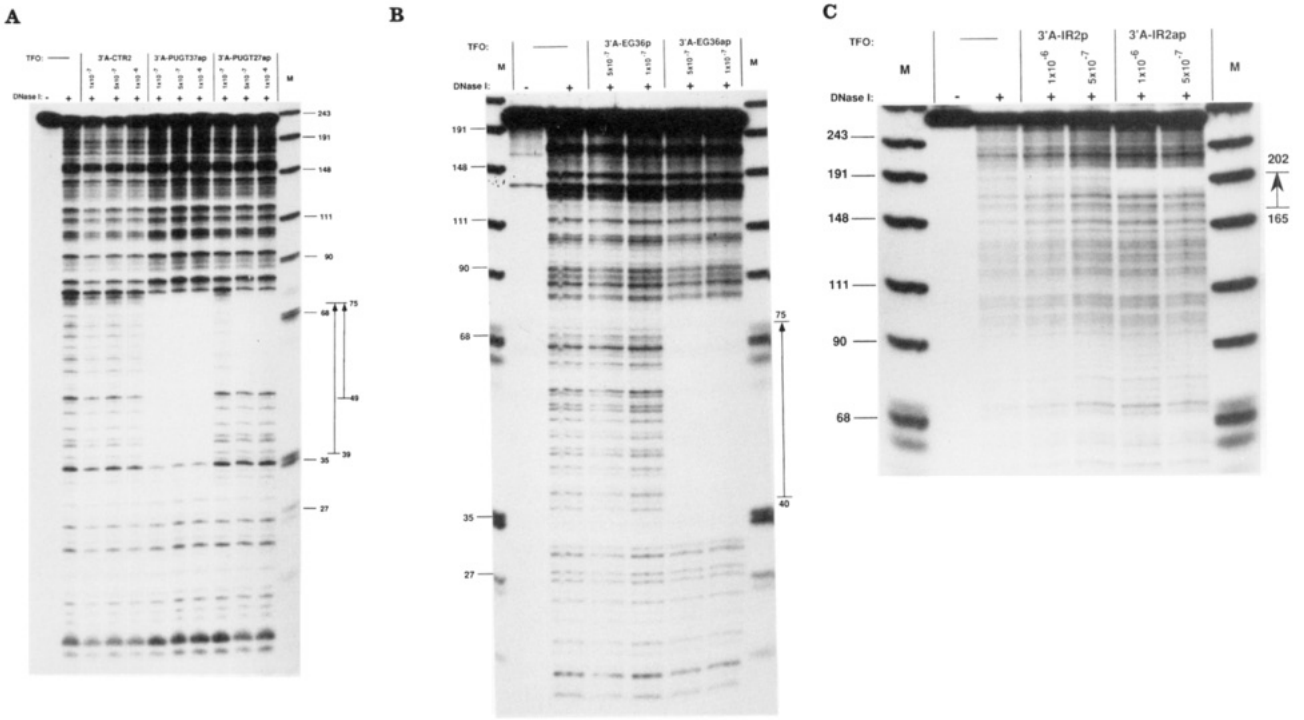


FIGURE 2: DNase I footprinting studies of triplex formation. End-labeled double-stranded DNA fragments were mixed with the indicated TFOs in a standard buffer containing 20 mM magnesium and treated with DNase I. Cooney et al. (1988) have shown previously that DNase I cuts triplex DNA at much lower efficiency than duplex DNA, yielding detectable footprints. Expected binding sites for the TFOs are indicated by arrows, with the arrowheads representing the 3′ ends of antiparallel bound TFOs. (A) Footprinting of the human c-*myc* site. The TFOs 3′-A-PUGT37ap, 3′-A-PUGT27ap, and 3′-A-CTR2 at the indicated concentrations (in strand equivalents) were mixed with a 231 bp fragment containing the c-*myc* target in binding buffer and treated with DNase I. Control lanes lacking TFO, with and without DNase I, are also shown. Lane M contains markers of the indicated sizes in base pairs. The arrows to the right show the expected binding sites for PUGT37ap (39–75) and PUGT27ap (49–75). (B) EGFR site. A 232 bp end-labeled fragment containing the EGFR triplex target was mixed with either 3′-A-EG36p or 3′-A-EG36ap and treated with DNase I. The arrow shows the expected binding site for EG36ap. (C) Mouse IR site. A 313 bp end-labeled fragment was mixed with 3′-A-IR2p or 3′-A-IR2ap at the indicated concentrations and treated with DNase I. The arrow represents the expected binding site for IR2ap.

indicates that these TFOs have a propanolamine linker added to the 3′-OH group (see Materials and Methods for details). The DNase I footprints of these two TFOs correspond in size and location to their expected binding sites. 3′-A-PUGT37ap binding results in a footprint over the entire 37 bp target, while 3′-A-PUGT27ap binding results in a footprint that is approximately 10 bp shorter. These data indicate that, within the limits of accuracy for DNase I footprinting of triplexes, 3′-A-PUGT37ap and 3′-A-PUGT27ap bind as predicted. Figure 2A also shows that there is no apparent interaction of either TFO with any sequences outside of the predicted binding site. An unrelated control TFO, 3′-A-CTR2, gives no discernable footprint, indicating that binding to the target sequence is TFO specific.

In order to assess the strength of TFO binding to the c-*myc* site, we have examined triplex formation using an electrophoretic assay that we refer to as band shift analysis. As originally shown by Cooney et al. (1988), triplexes remain intact during electrophoresis in magnesium-containing polyacrylamide gels. Furthermore, the migration of the triplex is significantly reduced relative to duplex or single-stranded DNA. Thus, the extent of triplex formation in a given sample can be monitored by this assay. Figure 3A shows a band shift analysis of PUGT27ap binding to a radiolabeled 53 bp fragment containing the c-*myc* target site. By using trace quantities of the target fragment (approximately 10^{-11} M in binding site equivalents) and titrating with increasing concentrations of PUGT27ap, we can examine the concentration dependence of triplex formation. In this case, triplex formation appears to be a two-state process, with free binding site plus free TFO forming a single discrete complex. Assuming a two-state equilibrium, the TFO strand concentration at the midpoint of the titration is equivalent to the apparent equilibrium dissociation constant (K_D) for triplex formation. For PUGT27ap binding to the c-*myc* site, that point is approximately 10^{-9} M. Although it is likely that triplex formation is a more complex process than these calculations assume, we have found that the apparent K_D is useful for comparing triplex formation by different TFOs at a given site. In particular, we find that the addition of a propanolamine linker to the 3′-terminus of PUGT27ap to form 3′-A-PUGT27ap has no measurable effect on the apparent K_D of triplex formation (data not shown). Similar data have been obtained for PUGT37ap, giving an apparent K_D of about 5×10^{-10} M. In contrast, band shift analysis using a random sequence isomer of PUGT37ap shows no triplex formation at TFO concentrations up to 10^{-6} M (not shown). These data indicate that TFO binding to the c-*myc* site is specific for the TFO sequence.

We have also performed the inverse of the band shift analysis, in which radiolabeled TFO is mixed with unlabeled duplex DNA and electrophoresed to assay triplex formation. In this experiment, plasmid pMHX DNA, which contains 3.8 kb of c-*myc* sequences including the triplex target (Boles & Hogan, 1987), was digested with *Taq*I amd *Pst*I, resulting in 10 fragments. The triplex target is contained in a 479 bp fragment. This DNA also contains several other G-rich, purine-rich sequences, such as the sequence 5′-GGGAGGCGTGGGGGTGGGACGGTGGGG-3′ contained in the 844 bp fragment. Note that this sequence has a base composition very similar to the intended triplex target (Figure 1A). The digested plasmid DNA was mixed with a 100-fold molar excess of radiolabeled PUGT27ap and electrophoresed

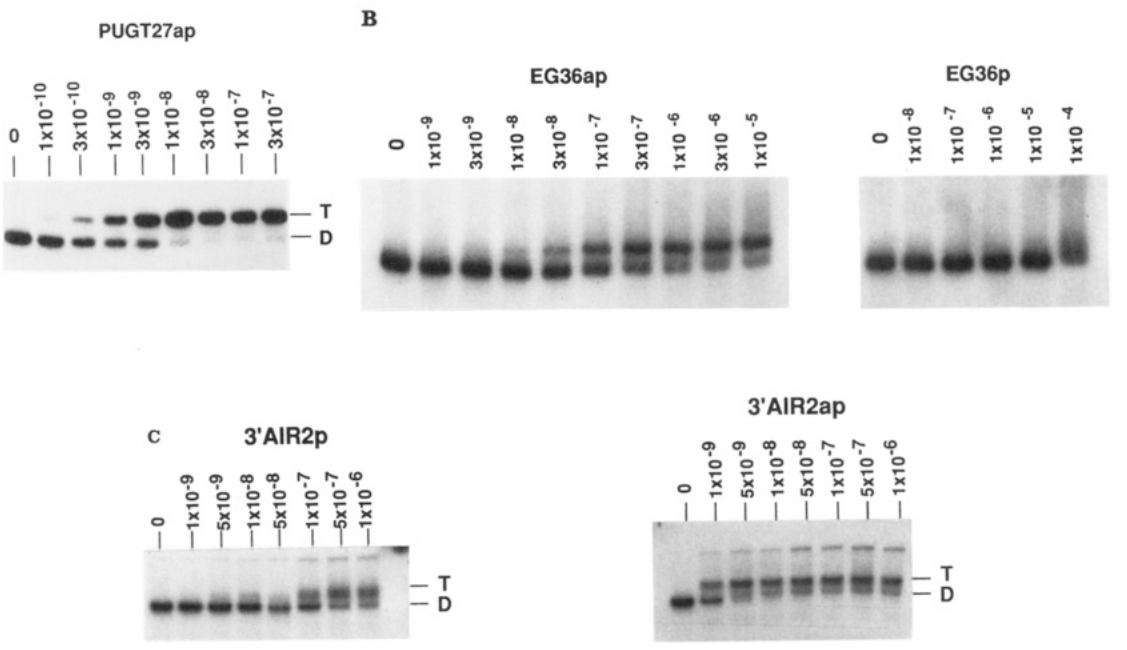


FIGURE 3: Band shift analysis of triplex formation. Binding of TFOs to a given target was assayed by electrophoresing the mixtures through a polyacrylamide gel in a buffer containing 89 mM Tris, 89 mM boric acid, and 5 or 20 mM magnesium chloride (TBM_5 or TBM_{20} buffer). Under these conditions, triplexes migrate significantly slower that duplexes (Cooney et al., 1988). (A) Human c-*myc*. An end-labeled 53 bp fragment containing the c-*myc* target (base pairs −155 to −112 numbered as in Figure 1A) was used at approximately 3×10^{-11} M in binding site equivalents. The indicated concentrations of PUGT27ap (strand equivalents) were added, followed by a buffer consisting of 10 mM Tris-HCl, pH 7.8, 5 mM magnesium chloride, 1 mM spermine, and 10% sucrose (final concentrations). Binding proceeded for 1 h at room temperature, followed by electrophoresis in a TBM_5-buffered polyacrylamide gel. The gel was dried and subjected to autoradiography. The positions of duplex and triplex DNA are indicated by "D" and "T", respectively. (B) Human EGFR. A synthetic 77 bp duplex consisting of base pairs −96 to −20 (numbered as in Figure 1) at 10^{-10} M was mixed with the indicated concentrations of EG36ap (left) or EG36p (right). Buffer was added as for part A, above, except the magnesium chloride concentration was 20 mM and spermine was omitted. Electrophoresis was in a TBM_{20}-buffered gel. Duplex and triplex are marked "D" and "T". (C) Mouse IR. A synthetic 51 bp duplex consisting of base pairs −145 to −95 at a concentration of about 10^{-10} M was mixed with the indicated strand concentrations of IR2p (left) or IR2ap (right), followed by buffer containing 20 mM magnesium (without spermine). Electrophoresis was in a TBM_{20}-buffered gel, and duplex and triplex are marked as above.

through a magnesium-containing gel (Figure 4). Lane 1 shows the migration of free TFO, and lane 2 shows the mixture of TFO plus duplex DNA. Lane 2 clearly shows that the TFO has complexed with the 479 bp fragment. No other duplex fragments, including the 844 bp fragment, bind detectable levels of TFO. While this is not a stringent test of the specificity of triplex formation, it does indicate that a target site that is merely G-rich and purine-rich is not sufficient to bind the TFO and that target sequence is critical.

For triplexes containing C^+·GC and T·AT triplets, it is known that the third strand binds parallel to the purine-rich strand of the duplex (Moser & Dervan, 1987; Fedorova et al., 1988; Praseuth et al., 1988). In the initial description of triplex formation at the c-*myc* site, it was implied in a figure that the TFO was bound parallel (Cooney et al., 1988). However, further examination suggested that the TFO in that study, PU1, could potentially bind either parallel or antiparallel and make similar base contacts with the duplex. This is due to the fact that the c-*myc* target is pseudopalindromic. Thus, the third strand orientation was unclear. DNase I footprinting data indicate that binding may in fact be antiparallel. Note that PUGT37ap is equivalent to PUGT27ap with an additional 10 bases on the 5′ end (Figure 1A). Examination of the footprints in Figure 2A indicates that the additional bases on the 5′ end of PUGT37ap protect sequences at the 3′ end of the purine-rich target strand. These data suggest that these TFOs bind antiparallel.

In order to address the question of third-strand orientation directly, we have modified the ends of selected TFOs with reactive groups capable of damaging double-stranded DNA. We reacted eosin-5-isothiocyanate with 3′-A-PUGT26ap to

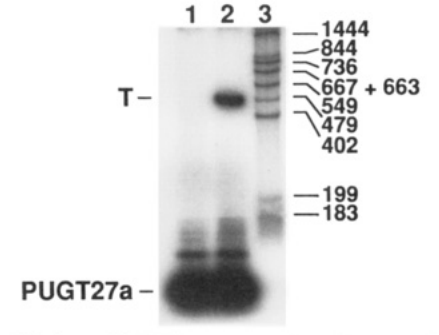


FIGURE 4: Binding of PUGT27ap to plasmid DNA fragments. Plasmid pMHX (Boles & Hogan, 1987) was digested with *Pst*I and *Taq*I, generating 10 fragments, including a 479 bp fragment containing the triplex target. 10^{-7} M end-labeled PUGT27ap was incubated in a buffer of 10 mM Tris-HCl, pH 7.8, 5 mM $MgCl_2$, 1 mM spermine, and 10% sucrose, without (lane 1) and with (lane 2) 10^{-9} M binding site equivalents of unlabeled pMHX DNA fragments. After 1 h at room temperature, the mixtures were electrophoresed in a TBM_5-buffered 5% polyacrylamide gel having a 20% polyacrylamide "plug" at the bottom to retain unbound PUGT27ap. Lane 3 shows the migration of end-labeled pMHX fragments in the absence of TFO, with the sizes in base pairs given to the right. Binding of PUGT27ap to the triplex target contained in the 479 bp pMHX fragment is evident in lane 2 and is indicated by "T".

form 3′-E-PUGT26ap, in which the eosin is covalently attached to the 3′ end of the TFO. Such modifications have little if any effect on the apparent equilibrium dissociation constant of the triplex (not shown). Triplexes formed with 3′-E-PUGT26ap and end-labeled duplex DNA were irradiated with the green output of an Argon ion laser, stimulating the eosin to produce singlet oxygen. Singlet oxygen is capable of re-

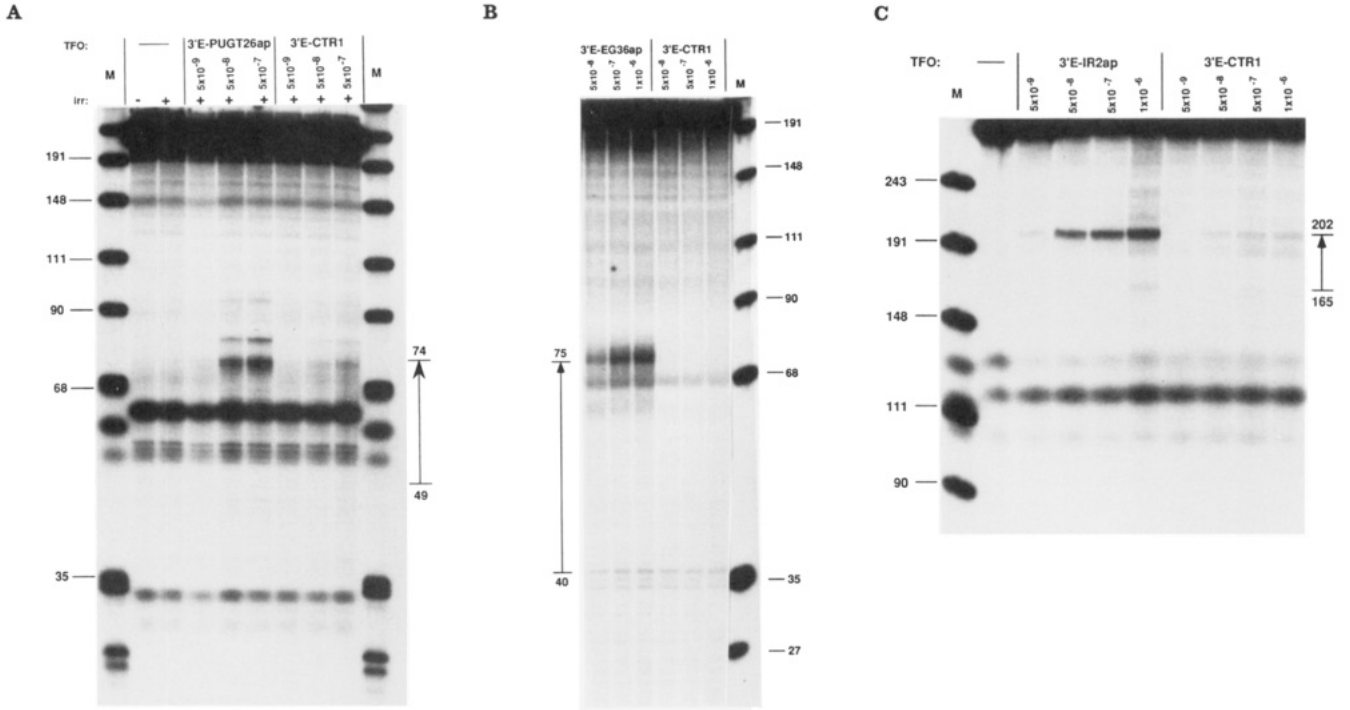


FIGURE 5: Eosin–TFO mapping of the third strand orientation. TFOs having eosin covalently linked to their 3′ ends were bound to end labeled target duplexes and irradiated with the green output of an argon ion laser. Excitation of the eosin moiety results in the formation of singlet oxygen, which can cause DNA damage. Following irradiation, the samples were treated with piperidine to cleave damaged sites and electrophoresed on denaturing polyacrylamide gels (see Materials and Methods for details). Arrows indicate the expected binding sites of the TFOs, with the arrowheads representing the 3′ ends of the TFOs. (A) Human c-*myc*. The 231 bp fragment containing the c-*myc* site was 3′-end-labeled on the purine-rich strand. 3′-E-PUGT26ap (e.g., PUGT26ap with eosin attached to the 3′ end) or 3′-E-CTR1 at the indicated strand concentrations were incubated with the fragment and treated as described above and under Materials and Methods. The arrow indicates the expected PUGT26ap binding site. "irr:" indicates samples that were unirradiated (−) or irradiated (+). (B) Human EGFR. The 232 bp fragment containing the EGFR site was 3′-end-labeled on the purine-rich strand and incubated with 3′-E-EG36ap or 3′-E-CTR1 at the indicated molar strand concentrations. The arrow indicates the expected EG36ap binding site. (C) Mouse IR. The 313 bp fragment 3′-end-labeled on the purine-rich strand was incubated with the indicated molar strand concentrations of 3′-E-IR2ap or 3′-E-CTR1. The arrow represents the expected binding site for IR2ap.

acting with DNA to form cleavable modifications at specific bases. The reaction is diffusion dependent and occurs at high rates only near the eosin source (Hogan et al., 1987). Thus, the location of the 3′ end of bound TFO can be determined by examining the pattern of damage to the duplex target.

Figure 5A shows the results of such an experiment. The vertical arrow at the right of the figure indicates the expected position of PUGT26ap (Figure 1A) bound to the c-*myc* target, with the arrowhead representing the 3′ end of the TFO (assuming it binds antiparallel). Reaction with 3′-E-PUGT26ap induces cleavages in the target consistent with the proposed location of the 3′ end (position 74). A control TFO, 3′-E-CTR1, shows low levels of cleavage at the highest concentrations that may indicate weak or transient interaction with the target site. No cleavage is observed at positions corresponding to parallel binding of the 3′-E-PUGT26ap (position 49). Additional experiments using TFOs with eosin attached to the 5′ end show cleavages only in the position expected for antiparallel binding (R. Durland, unpublished observations). Furthermore, cleavage requires covalent attachment of the eosin to the TFO, since a mixture of TFO and free eosin yields no apparent reaction. Omitting magnesium from the reaction also abolishes cleavage (not shown). These data indicate that the eosin-modified TFOs are providing accurate information on the backbone orientation of the third strand in the triplex.

To further confirm the strand orientation in the c-*myc* triplexes, we have modified the amine group of 3′-A-PUGT26ap by reacting it with diethylenetriaminepentaacetic acid anhydride (DTPA), which results in an EDTA-like group covalently attached to the 3′ end of the TFO (designated 3′-D-PUGT26ap). When chelated with iron and treated with a reducing agent such as dithiothreitol (DTT), hydroxyl radicals are produced that can cleave DNA (Dryer & Dervan, 1985; Moser & Dervan, 1987). In Figure 6A, increasing concentrations of the 3′-D-PUGT26ap–iron chelate were added to end-labeled duplex DNA containing the c-*myc* target and incubated to permit triplex formation. Subsequent reaction with DTT resulted in specific DNA cleavage at positions that cluster about the predicted location of the 3′ end of antiparallel bound TFO. Maximum cleavage rates occur at two sites on either side of the predicted 3′ terminus of the TFO. Similar observations have been noted previously for cleavage of DNA by iron EDTA covalently attached to oligonucleotides (Dryer & Dervan, 1985; Moser & Dervan, 1987).

Human Epidermal Growth Factor Receptor (*EGFR*). The target for triplex formation in the promoter region of the EGFR gene is shown in Figure 1B. It consists of the sequence from 59 to 94 base pairs upstream of the major transcription start site, as described by Johnson et al. (1988). It overlaps a region found to be sensitive to S1 nuclease in supercoiled plasmid DNA and to bind the transcription factor Sp1 and an unidentified protein present in nuclear extracts (Johnson et al., 1988). Deletion analysis indicates that this region is required for optimal expression of the EGFR promoter.

DNase I footprints of EG36ap and EG36p are shown in Figure 2B. These two TFOs have identical sequence composition, but they are mirror image isomers, i.e., the sequence of EG36ap read 5′ to 3′ is the same as the sequence of EG36p read 3′ to 5′. Unlike the c-*myc* site, the EGFR site is not pseudopalindromic. Thus, it is expected that EG36ap can only

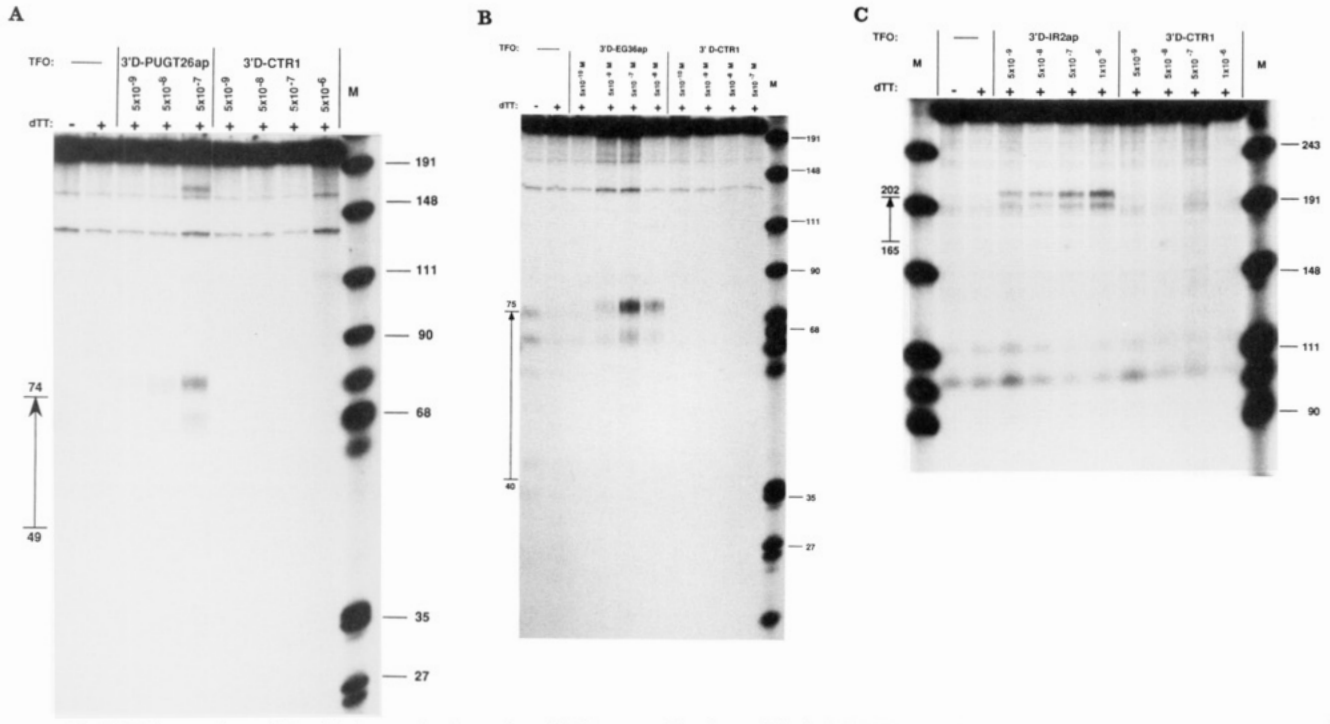


FIGURE 6: DTPA mapping of the third strand orientation. TFOs were 3′-end-modified with DTPA and incubated with the relevant end-labeled duplex in binding buffer containing iron and dithiothreitol (DTT; see Materials and Methods for details). The DTPA linked to the 3′ ends of the TFOs chelates iron, which reacts with DTT to produce hydroxyl radicals that can cleave duplex DNA. Control lanes incubated without TFO in the absence (−) and presence (+) of DTT are also shown. Arrows indicate expected TFO-binding sites, with the arrowheads representing the 3′ ends of the TFOs. (A) Human c-*myc*. 3′-D-PUGT26ap or 3′-D-CTR1 were incubated at the indicated strand concentrations with the end-labeled fragment used in Figures 2A and 5A. The arrow shows the expected binding site for PUGT26ap. (B) Human EGFR. 3′-D-EG36ap or 3′-D-CTR1 at the indicated concentrations were incubated with 3′-end-labeled DNA as in Figures 2B and 5B. The arrow represents the binding site for EG36ap. (C) Mouse IR. 3′-D-IR2ap or 3′-D-CTR1 were incubated at the indicated concentrations with the 3′-end-labeled fragment used in Figures 2C and 5C. The arrow shows the binding site for IR2ap.

bind antiparallel and EG36p can only bind parallel (if at all) while maintaining the desired base triplets. Only EG36ap yields a detectable footprint, suggesting that only antiparallel binding can occur at this site.

Band shift analysis of the EGFR site is shown in Figure 3B. The target site in this experiment was a synthetic 77 bp duplex containing the EGFR target site as well as flanking sequences from the EGFR promoter region. By a similar analysis to that used with c-*myc*, the apparent equilibrium dissociation constant for EG36ap binding to the synthetic EGFR site is (1–2) × 10^{-7} M. Under identical conditions, the sequence isomer EG36p shows no triplex formation at concentrations up to 10^{-4} M.

The strand orientation of EG36ap when bound to the EGFR site was examined by covalent attachment of eosin (Figure 5B) and DTPA (Figure 6B) to the 3′ end of 3′-A-EG36ap, by using procedures essentially identical with those used with c-*myc*. Figure 5B shows that triplexes formed from 3′-E-EG36ap result in site-specific DNA damage to a target duplex consistent with an antiparallel orientation. A control TFO, 3′-E-CTR1, shows no apparent cleavage. Similar results are obtained with 3′-D-EG36ap (Figure 6B). These results confirm that TFOs binding to the EGFR site bind antiparallel to the purine-rich target strand.

Mouse Insulin Receptor (IR). The mouse IR target is shown in Figure 1C and corresponds to a purine-rich region located upstream of the two major transcription start sites defined by Sibley et al. (1989). We are not aware of any function ascribed to this sequence at present.

DNase I footprint analysis of IR2ap and IR2p binding to the IR target sequence is shown in Figure 2C. These two TFOs are sequence isomers designed to bind antiparallel and parallel, respectively (Figure 1C). DNase I footprinting indicates that only IR2ap binds tightly enough to give a footprint at concentrations up to 10^{-6} M. However, the extent of the footprint is less than predicted from the size of the TFO, and it appears that only the 3′ end of the TFO is interacting strongly with the target. An examination of the target sequence shows that the 3′ end of the TFO may bind to a perfect homopurine region of 19 bp. However, the 5′ end of the TFO is targeted to a region that is only 68% purine on one strand and includes a stretch of three consecutive pyrimidines. This may prevent the 5′ end of IR2ap from binding to give a footprint in this region.

Band shift analysis for IR2ap and IR2p binding to the IR site is presented in Figure 3C. The target for this assay consisted of a 55 bp synthetic duplex containing the IR-binding site as well as flanking sequences. IR2ap binding to the target yields complexes with several different mobilities. As suggested above, this may be due to the relatively high number of pyrimidine residues at one end of the purine-rich target. Thus, if only the 3′ end of IR2ap binds strongly to its target, different conformations of the 5′ end may give rise to the multiple complexes observed. These effects prevent estimation of an apparent equilibrium dissociation constant for IR2ap, but it may be noted that greater than 90% of the target duplex has been shifted into a lower mobility complex at IR2ap concentrations of 5 × 10^{-9} M and above.

Band shift analysis indicates that the parallel isomer IR2p also binds to the target in this system. However, binding requires substantially higher concentrations of TFO (10^{-7} M and above). It is not possible for IR2p to bind to the target and maintain similar or identical base interactions over its entire length. We have not yet determined whether the observed binding indicates weak parallel binding of IR2p to the purine-rich target strand or antiparallel binding of a portion

of the TFO (i.e., the G-rich 5′ end) to a segment of the target.

Eosin- and DTPA-modified 3′-A-IR2ap was used to confirm the strand orientation. Both 3′-E-IR2ap and 3′-D-IR2ap demonstrate that binding is antiparallel (Figure 5C and 6C, respectively). In both experiments, a similarly modified control TFO, CTR1, gave reduced but detectable cleavage at sites close to those induced by IR2ap. The reason for this is unclear.

Discussion

The data described in this work show that G-rich oligodeoxyribonucleotides can bind to certain duplex DNAs to form triplexes. This confirms and extends the previous observations of Cooney et al. (1988). On the basis of analogy to previously described triple helices such as poly(dT)·poly(dA)·poly(dT), we propose that binding is through the major groove of the duplex and that specific hydrogen-bonded base triplets are formed. The sequences of the TFOs and the targets in this study suggest that triplex formation depends on the formation of T·AT and G·GC triplets. Such triplets have been described in a number of contexts previously. T·AT triplets are well known from extensive studies of triple helices containing polypyrimidine third strands. In particular, fiber diffraction studies on poly(dT)·poly(dA)·poly(dT) provide evidence that the third-strand T residues are Hoogsteen bonded to the A residues of the duplex (Arnott & Selsing, 1974; Arnott et al., 1976). These data are supported by the results of nuclear magnetic resonance studies on triplexes containing a mixture of T·AT and C^+·GC triplets (Rajagopal & Feigon, 1989; de los Santos et al., 1989). However, the third strand in those triple helices is bound parallel to the purine-rich duplex strand, while the TFOs described here bind antiparallel. G·GC triplets are known to occur for poly(G)·poly(G)·poly(C) in homopolymers and in supercoiled plasmid DNA (Marck & Thiele, 1978; Kohwi & Kohwi-Shigematsu, 1988). However, the only high-resolution data for G·GC triplets comes from X-ray crystallography of single triplets in transfer RNA molecules (Jack et al., 1976; Sussman et al., 1978; Hingerty et al., 1978). Thus, we can only speculate on the detailed structures of the triplexes observed here.

Figure 7 shows two possible hydrogen-bonding schemes for T·AT and G·GC triplets: labeled Hoogsteen and reverse Hoogsteen. (While this nomenclature historically refers only to the hydrogen-bonding schemes shown for the T·AT triplets, we refer to the analogous G·GC triplets with the same designations for convenience.) The T·AT triplets observed in poly(dT)·poly(dA)·poly(dT) are of the Hoogsteen type. Yeast $tRNA^{Phe}$ has been shown to contain a reverse Hoogsteen G·GC triplet. There are as yet no data to suggest which triplets are formed in the triplexes described here. It should be noted that the antiparallel orientation of the third strand requires that if Hoogsteen-type triplets are formed, the third-strand bases must have syn glycosidic angles, while reverse Hoogsteen triplets would permit the third-strand bases to adopt the normal anti conformation. We and others have found that it is possible to construct computer models in either conformation that avoid steric problems and maintain reasonable torsional angles for all three strands (van Vlijmen et al., 1990). Experiments to determine the structural details of a representative triplex are in progress.

An important feature of triplex formation in this study is that it occurs at physiological pH levels. The assays described here were typically buffered to a pH between 7.4 and 7.8. More detailed studies with the c-*myc* target have indicated that triplex formation is essentially insensitive to pH over a range of at least 5.5–8.3 (R. Durland, unpublished observations). This is not unexpected, since the proposed triplets in

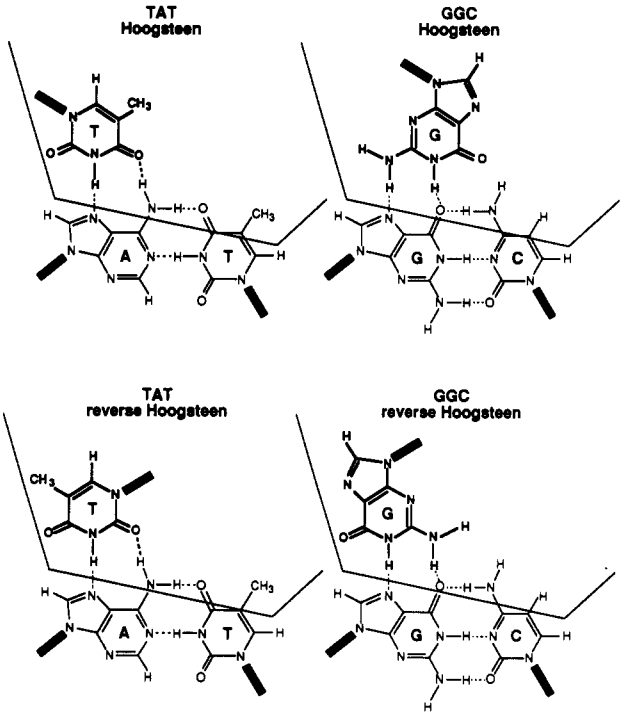


FIGURE 7: Possible hydrogen-bonding schemes for G·GC and T·AT triplets. Triplets are labeled Hoogsteen and reverse Hoogsteen depending on the orientation of the third-strand base. Hydrogen bonds are indicated by dotted lines. The location of the major groove of the duplex-binding site is indicated by the solid lines. Note that in order to maintain the experimentally determined antiparallel orientation of the third strand, the glycosidic bonds of third-strand bases in Hoogsteen-bonded triplets must be in the syn conformation, while those of reverse Hoogsteen-bonded triplets must be anti.

these structures do not involve ionized bases. Triplex formation requires multivalent cations such as magnesium at concentrations around 5–20 mM. In addition, we have observed that some triplexes, but not others, are stabilized by spermine in low millimolar amounts. The role of these cations in triplex formation remains to be determined.

The effect of monovalent ions such as sodium and potassium on the triplexes described here has not yet been examined in detail. Preliminary work with the c-*myc* site suggests that triplex formation is inhibited by high ionic strength. Similar results have been reported for triplexes containing C^+·GC and T·AT triplets (Maher et al., 1990). Several investigators have demonstrated that sodium and potassium can promote the formation of four-stranded helices by G-rich oligonucleotides (Arnott et al., 1974; Zimmerman et al., 1975; Henderson et al., 1987; Sen & Gilbert, 1988; Williamson et al., 1989; Sundquist & Klug, 1989). Formation of such structures by the G-rich TFOs in this study would be expected to interfere with triplex formation. Further experimentation in this area is necessary.

TFO binding in this work is very strong. Apparent dissociation constants are as low as 10^{-9} M, which is comparable to the binding of the *Escherichia coli* Trp repressor protein to its target sequence (Rose & Yanofsky, 1974; Haydock & Somerville, 1984). Triplex formation is also relatively specific. Individual TFOs bind tightly to their targets and weakly or not at all to unrelated sequences.

The target sites for triplex formation in this study all have a strong preference for purine residues on one strand. This is consistent with the proposed formation of T·AT and G·GC triplets as depicted in Figure 7. An ideal target would be a perfect homopurine·homopyrimidine sequence. However, we

have chosen to examine triplex formation at targets that are not 100% purine so that naturally occurring targets that may be involved in gene expression may be studied. When pyrimidines are present in the purine-rich target strand, we have arbitrarily chosen to design TFOs that place G residues at CG pairs and T residues at TA pairs. Work is in progress to investigate other possible accommodations of pyrimidines in the purine-rich strand, and it is likely that this design can be improved. Nevertheless, our data show that triplexes will form at naturally occurring sites that are not perfect homopurine·homopyrimidine regions.

A second feature of the target sites studied here is that they are all relatively G-rich. Thus, the observed characteristics of these triplexes are likely to be dominated by the G·GC triplets. In particular, the antiparallel strand orientation is likely to be a consequence of G·GC triplet formation, since oligo(dT) is known to bind parallel to poly(dA)·poly(dT). It is interesting to consider how the proposed T·AT triplets in the current triplexes may differ from those in poly(dT·dA·dT).

Reported attempts by other investigators to confirm the results of Cooney et al. (1988) were unsuccessful (Lyamichev et al., 1988, 1990; Maher et al., 1989). Such failures may have been due to the lack of magnesium or to the unrecognized requirement for antiparallel binding. However, Beal and Dervan (1991) recently demonstrated binding of a G-rich 19-mer to a G-rich homopurine target and found that binding was antiparallel. Our observations demonstrate that G-rich TFOs such as those described in this work can bind to naturally occurring sites in gene promoters. Thus, triplex formation by G-rich TFOs may be a useful tool. Work with triplexes containing T·AT and C^+·GC triplets has shown that they can be used as site-specific DNA cleavage reagents and can interfere with the actions of DNA-binding proteins. However, they are limited by the requirement for relatively low pH, generally less than 7.0. Triplex formation by G-rich TFOs is likely to have similar uses without being limited to low pH conditions. In fact, Cooney et al. (1988) showed that a TFO that binds the c-*myc* site is capable of inhibiting in vitro transcription of the c-*myc* gene. The ability of cultured mammalian cells to spontaneously internalize oligonucleotides suggests that similar possibilities may exist in vivo (Zamecnik et al., 1986; Vlassov et al., 1986). Recent work in which a TFO that binds to the c-*myc* site was administered to cultured HeLa cells has provided evidence that triplex formation can occur in the nuclei of living cells and can affect cellular processes such as transcription (Postel et al., 1991).

References

Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509–521.

Arnott, S., Chandrasekaran, R., & Marttila, C. M. (1974) *Biochem. J. 141*, 537–543.

Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. C. (1976) *Nucleic Acids Res. 3*, 2458–2470.

Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., & Leder, P. (1983) *Cell 34*, 779–787.

Beal, P. A., & Dervan, P. B. (1991) *Science 251*, 1360–1363.

Boles, T. C., & Hogan, M. E. (1987) *Biochemistry 26*, 367–376.

Broitman, S. L., Im, D. D., & Fresco, J. R. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5120–5124.

Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.

Davis, T. L., Firulli, A. B., & Kinniburgh, A. J. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9682–9686.

de los Santos, C., Rosen, M., & Patel, D. (1989) *Biochemistry 28*, 7282–7289.

Dryer, G. B., & Dervan, P. B. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 968–972.

Durland, R. H., Kessler, D. J., Duvic, M., & Hogan, M. (1990) in *Molecular Basis of Specificity in Nucleic Acid-Drug Interactions* (Pullman, B., & Jortner, J., Eds.) pp 565–578, Kluwer Academic, Boston, MA.

Fedorova, O. S., Knorre, D. G., Podust, L. M., & Zarytova, V. F. (1988) *FEBS Lett. 228*, 273–276.

Felsenfeld, G., & Rich, A. (1957) *Biochim. Biophys. Acta 26*, 457–468.

Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.

Haas, B. L., & Guschlbauer, W. (1976) *Nucleic Acids Res. 3*, 205–218.

Hanvey, J. C., Shimizu, M., & Wells, R. D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6292–6296.

Hanvey, J. C., Shimizu, M., & Wells, R. D. (1989) *Nucleic Acids Res. 18*, 157–161.

Harvey, S. C., Luo, J., & Lavery, R. (1988) *Nucleic Acids Res. 16*, 11795–11809.

Hay, N. J., Bishop, M., & Levens, D. (1987) *Genes Dev. 1*, 659–671.

Haydock, P. V., & Somerville, R. L. (1984) *Biochem. Biophys. Res. Commun. 119*, 926–932.

Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., Jr., & Blackburn, E. H. (1987) *Cell 51*, 899–908.

Hingerty, B., Brown, R. S., & Jack, A. (1978) *J. Mol. Biol. 124*, 523–534.

Hogan, M. E., Rooney, T. F., & Austin, R. H. (1987) *Nature 328*, 554–557.

Howard, F. B., Frazier, J., Lipsett, M. N., & Miles, H. T. (1964) *Biochem. Biophys. Res. Commun. 17*, 93–102.

Htun, H., & Dahlberg, J. E. (1988) *Science 241*, 1791–1796.

Jack, A., Ladner, J. E., & Klug, A. (1976) *J. Mol. Biol. 108*, 619–649.

Johnson, A. C., Jinno, Y., & Merlino, G. T. (1988) *Mol. Cell. Biol. 8*, 4174–4184.

Kohwi, Y., & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3781–3785.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.-L., Thuong, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749–7760.

Lee, J. S., Johnson, D. A., & Morgan, A. R. (1979) *Nucleic Acids Res. 6*, 3073–3091.

Letai, A. G., Palladino, M. A., Fromm, E., Rizzo, V., & Fresco, J. R. (1988) *Biochemistry 27*, 9108–9112.

Lipsett, M. N. (1963) *Biochem. Biophys. Res. Commun. 11*, 224–228.

Lipsett, M. N. (1964) *J. Biol. Chem. 239*, 1256–1260.

Lyamichev, V. I., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1985) *J. Biomol. Struct. Dyn. 3*, 327–338.

Lyamichev, V. I., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn. 3*, 667–669.

Lyamichev, V. I., Mirkin, S. M., Frank-Kamenetskii, M. D., & Cantor, C. R. (1988) *Nucleic Acids Res. 16*, 2165–2178.

Lyamichev, V. I., Frank-Kamenetskii, M. D., & Soyfer, V. N. (1990) *Nature 344*, 568–570.

Maher, L. J., III, Wold, B., & Dervan, P. B. (1989) *Science 245*, 725–730.

Maher, L. J., III, Dervan, P. B., & Wold, B. J. (1990) *Biochemistry 29*, 8820–8826.

Marck, C., & Thiele, D. (1978) *Nucleic Acids Res. 5*, 1017–1028.

Morgan, A. R., & Wells, R. D. (1968) *J. Mol. Biol. 37*, 63–80.

Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.

Postel, E. H., Mango, S. E., & Flint, S. J. (1989) *Mol. Cell. Biol. 9*, 5123–5133.
Postel, E. H., Flint, S. J., Kessler, D. J., & Hogan, M. E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* (in press).
Povsic, T. J., & Dervan, P. B. (1989) *J. Am. Chem. Soc. 111*, 3059–3061.
Praseuth, D., Perrouault, L., Le Doan, T., Chassignol, M., Thuong, N., & Hélène, C. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1349–1353.
Rajagopal, P., & Feigon, J. (1989) *Nature 339*, 637–640.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Rose, J. K., & Yanofsky, C. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 3134–3138.
Sen, D., & Gilbert, W. (1988) *Nature 334*, 364–366.
Sibley, E., Kastelic, T., Kelly, T. J., & Lane, M. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9732–9736.
Sundquist, W. I., & Klug, A. (1989) *Nature 342*, 825–829.
Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M., & Kim, S.-H. (1978) *J. Mol. Biol. 123*, 607–630.
Thiele, D., & Guschlbauer, W. (1971) *Biopolymers 10*, 143–157.
van Vlijmen, H. W., Rame, G. L., & Pettitt, B. M. (1990) *Biopolymers 30*, 517–532.
Vlassov, V. V., Gorokhova, O. E., Ivanova, E. M., Kutyavin, I. V., Yurchenko, L. V., Yakubov, L. A., Abdukayumov, M. N., & Skoblov, Y. S. (1986) *Biopolim. Kletka 2*, 323–327.
Williamson, J. R., Raghuraman, M. K., & Cech, T. R. (1989) *Cell 59*, 871–880.
Zamecnik, P. D., Goodchild, J., Taguchi, Y., & Sarin, P. S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4143–4146.
Zimmerman, S. B., Cohen, G. H., & Davies, D. R. (1975) *J. Mol. Biol. 92*, 181–192.

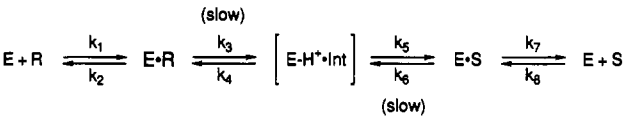
Mechanism of the Reaction Catalyzed by Mandelate Racemase. 1. Chemical and Kinetic Evidence for a Two-Base Mechanism†

Vincent M. Powers,‡ Carolyn W. Koo, and George L. Kenyon*
Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143
John A. Gerlt* and John W. Kozarich*
Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742
Received April 4, 1991; Revised Manuscript Received June 14, 1991

ABSTRACT: The fate of the α-hydrogen of mandelate in the reaction catalyzed by mandelate racemase has been investigated by a mass spectroscopic method. The method entails the incubation of (*R*)- or (*S*)-[α-^{1}H]mandelate in buffered D_2O to a low extent of turnover (about 5–8%), esterification of the resulting mixture of mandelates with diazomethane, derivatization of the methyl esters with a chiral derivatizing agent, and quantitation of the isotope content of the α-hydrogen of both substrate and product by gas chromatography/mass spectrometric analysis. No significant substrate-derived α-protium was found in the product for racemization in either direction. In addition, in the (*R*) to (*S*) direction almost no exchange (≤0.4%) of the α-hydrogen in the remaining (*R*) substrate pool occurred, but in the (*S*) to (*R*) direction 3.5–5.1% exchange of the α-hydrogen in the remaining substrate (after 5.1–7.2% net turnover) was found. Qualitatively similar results were obtained in the (*S*) to (*R*) direction in H_2O when (*S*)-[α-^{2}H]mandelate was used as substrate. In other experiments, an overshoot in the progress curve was observed when the racemization of either enantiomer of [α-^{1}H]mandelate in D_2O was monitored by following the change in ellipticity of the reaction mixture; the magnitude of the overshoot was greater in the (*R*) to (*S*) than in the (*S*) to (*R*) direction. All of the available data indicate that the reaction catalyzed by mandelate racemase proceeds by a two-base mechanism, in contrast to earlier proposals.

Mandelate racemase (MR)[1] (EC 5.1.2.2) from *Pseudomonas putida* ATCC 12633 catalyzes the interconversion of the (*R*) and (*S*) enantiomers of mandelic acid via abstraction of the α-hydrogen as a proton. Catalysis is thought to involve an intermediate that has at least partial resonance-stabilized carbanionic character[2] on the basis of the effect of electron withdrawing/donating groups on the velocity of the reaction

Scheme I



and the elimination of halide ions from *p*-(halomethyl)-mandelates (Kenyon & Hegeman, 1979; Lin et al., 1988,

† This work was supported by Grants GM-40570 (J.A.G., J.W.K., and G.L.K.), AR-17323 (G.L.K.), GM-34572 (J.A.G.), and GM-37210 (J.W.K.). V.M.P. was supported by a predoctoral fellowship from the American Foundation for Pharmaceutical Education, and C.W.K. was supported by a National Institutes of Health Predoctoral Traineeship. This is paper 13 in a series on mandelate racemase; paper 12 is Neidhart et al. (1990).
‡ Present address: Department of Molecular and Cell Biology, W. M. Stanley Hall, University of California, Berkeley, CA 94720.

[1] Abbreviations: CD, circular dichroism; FID, flame ionization detector; GC, gas chromatograph(y); H297N, the mutant of mandelate racemase in which histidine 297 has been replaced with asparagine; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; MR, mandelate racemase; MTPA, α-methoxy-α-(trifluoromethyl)phenylacetic acid; MW, molecular weight; NMR, nuclear magnetic resonance spectroscopy; PLP, pyridoxal phosphate.

0006-2960/91/0430-9255$02.50/0