Triplex Targets in the Human Rhodopsin Gene†

Brian D. Perkins, John H. Wilson, Theodore G. Wensel,* and Karen M. Vasquez‡

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030

Received March 6, 1998; Revised Manuscript Received June 8, 1998

ABSTRACT: We have explored the application of triplex technology to the human rhodopsin gene, which encodes a G-protein-linked receptor involved in the genetic disorder autosomal dominant retinitis pigmentosa (ADRP). Our results support the hypothesis that most human genes contain high-affinity triplex sites and further refine the rules governing identification and successful targeting of triplex-forming oligonucleotides (TFOs) to these sites. Using a computer search for sites 15 nucleotides in length and greater than 80% purine, we found 143 distinct sites in the rhodopsin gene and comparable numbers of sites in several other human genes. By applying more stringent criteria, we selected 17 potential target sites in the rhodopsin gene, screened them with a plasmid binding assay, and found 8 that bound TFOs with submicromolar affinity ($K_d = 10^{-9}-10^{-7}$ M). We compared purine (GA) and mixed (GT) TFOs at each site, and found that GA-TFOs consistently bound with higher affinity, and were less sensitive to pyrimidine interruptions in the target strand. High G-content favored high-affinity binding; only sites with >54% G-content bound TFOs with $K_d \leq 10^{-8}$ M.

A long-sought goal in molecular biology is the ability to deliver reagents to specific sites within intact genomes of living cells or organisms, to study and manipulate gene function and structure. Triplex DNA technology (*1, 2*) has been used successfully to deliver reagents that alter gene function or structure to specific sites within genes in cell-free and intracellular conditions using triplex-forming oligonucleotides (TFOs).[1] In the past decade, several laboratories have demonstrated that gene expression in cells can be inhibited by TFOs targeted to triplex forming sites in both plasmid (*3*−*6*) and chromosomal genes (*6*−*16*). TFOs with attached DNA damaging moieties can cause site-specific damage in mammalian cells and induce localized mutations (*17*−*19*) or stimulate site-specific homologous recombination (*20, 21*).

An especially challenging problem, for which triplex technology seems suited, is the development of gene-based approaches for genetic disorders with dominant inheritance. Conventional strategies for gene introduction or supplementation are not applicable to those dominant diseases caused by deleterious effects of a defective gene product, rather than by deficiency of the wild-type gene product. To treat such autosomal dominant diseases successfully, it is necessary to develop approaches to inactivate, repair, or replace the defective copy of the gene. The documented effects of triplex formation, which include transcription inhibition, mutation, and recombination, might be exploited to inactivate or correct defective genes.

One clinically important dominant disorder is autosomal dominant retinitis pigmentosa (ADRP), which affects the photoreceptor cells of the retina (*22*−*24*) and is characterized by photoreceptor cell death and progressive retinal degeneration that ultimately leads to blindness (*25*). ADRP can be caused by mutations in several different genes (*24*); however, the most common form results from mutations in the gene for rhodopsin, the G-protein-coupled receptor that is the primary photon detector in the visual signal transduction cascade (*26*). More than 70 mutations associated with ADRP have been identified in the rhodopsin gene (*27*). The idea that expression of mutant forms of rhodopsin, rather than haploinsufficiency, leads to retinal degeneration is supported by the existence of a null recessive allele that does not cause RP in haploid carriers (*28*) and by similar observations from rhodopsin "knock-out" mice (*29*). These results raise the possibility that elimination or reduction of expression of harmful forms of rhodopsin may result in the prolonged survival of retinal cells even if levels of wild-type rhodopsin are well below those in normal rods.

The work described here represents the first steps toward testing the human rhodopsin gene as a target for triplex-based approaches of gene therapy for ADRP. We have searched for and identified several potential triplex-forming sites within the rhodopsin gene. We have screened these sites and identified seven that exhibit high-affinity binding to specifically designed GA- and GT-TFOs. A comparison of K_ds at these various sites reveals critical differences between GA- and GT-TFOs, with respect to their dependence on the G-content of the target site and their sensitivity to pyrimidine interruptions in the purine-rich strand.

EXPERIMENTAL PROCEDURES

Oligonucleotide Synthesis and Purification. Oligodeoxyribonucleotides were purchased from Genosys Biotechnolo-

† This work was funded by grants from the Texas Advanced Technology Program (004949-803) and the NIH (EY11731). B.D.P. is supported by a training grant from the NIH (T32 EY07102).

* Corresponding author.

‡ Current address: Department of Therapeutic Radiology, Yale University School of Medicine, New Haven, CT 06510.

[1] Abbreviations: TFO, triplex-forming oligonucleotide; ADRP, autosomal dominant retinitis pigmentosa.

S0006-2960(98)00525-X CCC: $15.00
Published on Web 07/21/1998

gies, Inc. (The Woodlands, TX), or from Oligos Etc. (Wilsonville, OR). TFOs were dissolved in Milli-Q water and desalted on NAP-5 columns (Pharmacia LKB Biotechnology, Uppsala, Sweden). Oligonucleotides were then purified by denaturing PAGE (15% acrylamide) as described elsewhere (*30*). Oligonucleotide integrity was analyzed by PAGE of 5′-end-labeled oligos. Oligos were labeled using T4 polynucleotide kinase (New England Biolabs Inc., Beverly, MA) and [γ-^{32}P]ATP (6000 Ci/mmol) (New England Nuclear Research Products, Boston, MA) for 30 min at 37 °C. DNA concentrations were determined by UV absorbance at 260 nm.

Triplex Formation by Electrophoretic Mobility Shift Assay with Synthetic Duplexes. Oligonucleotides corresponding to gene target sites were annealed in a 1:1 molar ratio to form a target duplex. Duplex strands were 5′-end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP, and the duplex was then gel-purified by PAGE (12% acrylamide) under nondenaturing conditions. Labeled duplex ($<1 \times 10^{-10}$ M) was incubated with increasing concentrations of TFO in binding buffer (10 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$ and 10% sucrose) at 37 °C for 72 h to attain equilibrium (*30*, *31*). Triplex-induced mobility shifts were monitored by electrophoresis through native 12% polyacrylamide gels containing 89 mM Tris−borate, pH 8.0, and 10 mM $MgCl_2$. Gels were run for 6−7 h at 55 V, dried, and exposed to film for autoradiography. For quantitation, the radioactivity in each band was measured using a Betagen Beta Scope 603 blot analyzer.

Plasmid Fragment Binding Assay. Plasmid p8/6, a gift from Dr. Tiansen Li of Harvard Medical School, consists of a *Sal*I genomic fragment (*32*) cloned into the multiple cloning site of pBluescript SK(−). The *Sal*I fragment contains 4.6 kb upstream sequence, the 6.9 kb rhodopsin transcription unit, and 4.9 kb downstream sequence. Plasmid p8/6 was digested with *Eco*RI (New England Biolabs Inc.) for 1 h at 37 °C, phenol-extracted, ethanol-precipitated, and resuspended in 0.1× TE buffer (10 mM Tris-HCl, 1 mM EDTA). TFOs (1×10^{-6} M) were 5′-end-labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP for 30 min and were purified through Micro Bio-Spin 6 chromatography columns (Bio-Rad Laboratories, Hercules, CA). An aliquot of each labeled TFO was subjected to electrophoresis on denaturing polyacrylamide gels, the TFO band was excised, and the radioactivity was measured by scintillation counting. For the plasmid binding assay, an equal amount of radioactivity for each TFO, at a final TFO concentration of approximately 10^{-8} M, was incubated with 1 μg of digested plasmid DNA (10^{-8} M) with binding buffer in a final volume of 10 μL. The samples were incubated at 37 °C for 18 h and then subjected to electrophoresis through a 0.8% agarose gel containing 89 mM Tris−borate, pH 8.0, and 10 mM $MgCl_2$ for 4−5 h at 60 V. The gels were stained with ethidium bromide, photographed, dried under low vacuum at 40 °C for 30 min on nylon membranes, and exposed to film for autoradiography.

Probability Calculations. The number of 15-nucleotide-long purine-rich runs expected in a given gene (assuming random sequence) was determined using the equation $2l\{[15!/m!(15 - m)!](1/2)^m(1/2)^{15-m}\}$ where $2l$ is twice the length of the gene in base pairs (since both strands must be searched) and m is the number of purine nucleotides within the string. The total number of sites expected is the sum of the values determined when using 12, 13, 14, and 15 matches (3, 2, 1, and 0 mismatches) in a string of 15 nucleotides.

To narrow the range of sites to be screened experimentally and to reduce the likelihood of multiple matching sites for each TFO, we chose a criterion based on the probability of finding a matching site with the same number of pyrimidine interruptions in a random sequence of the same length as the target site in the rhodopsin gene. This probability was calculated according to $p(n,m) = (1/4)^m(3/4)^{n-m}n!/m!(n - m)!$, where m is the number of nucleotides that match the target sequence and n is the total number of nucleotides in the target site (i.e., $n - m$ is the number of pyrimidine interruptions). A value of $< 3 \times 10^{-9}$ was chosen as the criterion for experimental screening.

RESULTS

Potential Triplex Target Sites in the Human Rhodopsin Gene. To identify potential triplex target sites within the human rhodopsin gene (6.9 kb sequence; GenBank accession no. K02281), we used the Genetics Computer Group (GCG) program to search the coding and noncoding strands for stretches of 15 nucleotides that were at least 80% purine. The initial search identified 802 partially overlapping sequences, representing 143 distinct sites. Application of our statistical criterion [$p(n,m) < 3 \times 10^{-9}$] yielded 17 sites for experimental screening. These sites and their locations within the gene are listed in Table 1, and shown schematically in Figure 1. They are fairly evenly distributed throughout the gene; however, only one was found in the coding sequence (Figure 1).

The number of partially overlapping sequences initially identified, 802, is 3.5-fold greater than the 246 sites expected for a random sequence of this length, and much greater than would be expected if there were any selective pressure against homopurine runs in the human genome. As shown in Figure 2 (solid line vs dashed line), this overrepresentation of potential triplex-forming sites is not limited to the rhodopsin gene, but extends to other human genomic sequences as well. When these genomic sequences were randomized and the resulting sequences searched for potential triplex binding sites, the results matched the probability calculations (Figure 2, open symbols vs dashed line). These results for human genomic sequences are in general agreement with a previous analysis, which showed that pure oligopurine tracts are overabundant in eukaryotic genomes (Behe, 1995).

Design of TFOs. All reported effects of TFOs in living cells, thus far, were obtained with either GA- or GT-TFOs, with the great majority designed for antiparallel binding (*2*); therefore, we designed GA- and GT-TFOs to bind to the purine-rich strand of the target sequence in an antiparallel orientation, using standard rules for triplex formation (*1*, *2*, *33*). All TFOs and their orientations relative to the purine-rich strand of the target are listed in Table 1. For each target site, we designed and tested both an antiparallel GT- and GA-TFO (designated by AT and AA, respectively, in Table 1) to compare their abilities to form triplex. At pyrimidine interruptions (PI) in the purine-rich strand of the targeted sequence, TFOs were designed to carry a T opposite C and a C opposite T (*34*). For some target sites with low

Table 1: TFO Binding at Potential Target Sites in the Rhodopsin Gene

Site[a]	Sequence[b]	%G[c]	PI[d]	TFO[e]	K_d(M)[f]
746	**5′ GACAGGACAGGAGAAGGGAGAAGG 3′***	50	2		
(24)	3′ GTTTGGTTTGGTGTTGGGTGTTGG 5′			1AT	$>10^{-6}$
	3′ GATAGGATAGGAGAAGGGAGAAGG 5′			1AA	$>10^{-6}$
1206	**5′ GAGAGGGGAGGCAGAGGA 3′***	61	1		
(18)	3′ GTGTGGGGTGGTTGTGGT 5′			2AT	**$>10^{-6}$**
	3′ GAGAGGGGAGGTAGAGGA 5′			2AA	**1×10^{-9}**
1411	**5′ AGGGAGAGGGGAAGAGA 3′**	59	0		
(17)	3′ TGGGTGTGGGGTTGTGT 5′			3AT	**1×10^{-8}**
	3′ AGGGAGAGGGGAAGAGA 5′			3AA	**1×10^{-9}**
1876	**5′ AGAGGAAGAAGAAGGAAA 3′**	38	0		
(18)	3′ TGTGGTTGTTGTTGGTTT 5′			4AT	$>10^{-6}$
	3′ AGAGGAAGAAGAAGGAAA 5′			4AA	$\geq3\times10^{-6}$
1965	**5′ GAGAAGGGAGAGGGAGGAAGGA 3′**	59	0		
(22)	3′ GTGTTGGGTGTGGGTGGTTGGT 5′			5AT	**3×10^{-9}**
	3′ GAGAAGGGAGAGGGAGGAAGGA 5′			5AA	**1×10^{-9}**
2265	**5′ GGGAGGAGGGGGAAGGGGCAGAGGG 3′***	75	1		
(25)	3′ GGGTGGTGGGGGTTGGGGTTGTGGG 5′			6AT	**5×10^{-9}**
	3′ GGGAGGAGGGGGAAGGGGTAGAGGG 5′			6AA	**2×10^{-9}**
2619	**5′ GAGCAGAAGGGAAGAAG 3′**	47	1		
(17)	3′ GTGTTGTTGGGTTGTTG 5′			7AT	$>10^{-6}$
	3′ GAGTAGAAGGGAAGAAG 5′			7AA	$>10^{-6}$
2750	**5′ AGGGCAGAAGAAGAAACAGA 3′**	35	2		
(20)	3′ TGGGTTGTTGTTGTTTTTGT 5′			8AT	$>10^{-6}$
	3′ AGGGTAGAAGAAGAAATAGA 5′			8AA	$>10^{-6}$
3135	**5′ GAGGAGGAAAGGGAAGTGAGAGTGGGAA 3′***	54	2		
(28)	3′ GTGGTGGTTTGGGTTGCGTGTGCGGGTT 5′			9AT	**$>10^{-6}$**
	3′ GAGGAGGAAAGGGAAGCGAGAGCGGGAA 5′			9AA	**1×10^{-8}**
3217	**5′ GGAGAAGATGGGGAAAAAG 3′***	47	1		
(19)	3′ GGTGTTGTCGGGGTTTTTG 5′			10AT	$>10^{-6}$
	3′ GGAGAAGACGGGGAAAAAG 5′			10AA	$>10^{-6}$
	5′ GGTGTTGTGGGGGTTTTTG 3′			10PT	$>10^{-6}$
4129	**5′ AGAAGGCAGAGAAGGAGG 3′**	50	1		
(18)	3′ TGTTGGTTGTGTTGGTGG 5′			11AT	$>10^{-6}$
	3′ AGAAGGTAGAGAAGGAGG 5′			11AA	$>10^{-6}$
4706	**5′ GAGGGTAAAGGGGTAAAGGG 3′**	55	2		
(20)	3′ GTGGGCTTTGGGGCTTTGGG 5′			12AT	$>10^{-6}$
	3′ GAGGGCAAAGGGGCAAAGGG 5′			12AA	$>10^{-6}$
4857	**5′ AGAAAGGTAGAAAGGGCAAAA 3′**	33	2		
(21)	3′ TGTTTGGCTGTTTGGGTTTTT 5′			13AT	$>10^{-6}$
	3′ AGAAAGGCAGAAAGGGTAAAA 5′			13AA	$>10^{-6}$
	5′ TGTTTGGGTGTTTGGGTTTTT 3′			13PT	$>10^{-6}$
5064	**5′ GAGGGGCAGAAGCAGGCAAAGGG 3′**	52	3		
(23)	3′ GTGGGGTTGTTGTTGGTTTTGGG 5′			14AT	$>10^{-6}$
	3′ GAGGGGTAGAAGTAGGTAAAGGG 5′			14AA	$>10^{-6}$
5313	**5′ GGGGGAAGGTGTAGGGGATGGGAGA 3′***	64	3		
(25)	3′ GGGGGTTGGCGCTGGGGTCGGGTGT 5′			15AT	**1×10^{-6}**
	3′ GGGGGAAGGCGCAGGGGACGGGAGA 5′			15AA	**3×10^{-9}**
5662	**5′ AGGAAGGAATGGAGGAATGAATGGGAAGGGAGAA 3′**	47	3		
(34)	3′ TGGTTGGTTCGGTGGTTCGTTCGGGTTGGGTGTT 5′			16AT	**$>10^{-6}$**
	3′ AGGAAGGAACGGAGGAACGAACGGGAAGGGAGAA 5′			16AA	**1×10^{-7}**
	5′ TGGTTGGTTGGGTGGTTGGTTGGGGTTGGGTGTT 3′			16PT	**$>10^{-6}$**
6158	**5′ GGGGGAGGGGGACGGTGAAGG 3′**	71	2		
(21)	3′ GGGGGTGGGGGTTGGCGTTGG 5′			17AT	**8×10^{-9}**
	3′ GGGGGAGGGGGATGGCGAAGG 5′			17AA	**2×10^{-9}**

[a] Numbers refer to the most 5′ nucleotide of the target sequence in the mRNA identical strand. Numbers in parentheses indicate the length of the target site. [b] Target sequences are shown in boldface type. An asterisk indicates sequences that are the complement to the mRNA identical strand. The orientation of the TFOs is shown relative to the purine-rich strand of target sequence. [c] % G refers to the G-content of the purine-rich strand of the target site. [d] Numbers refer to the number of pyrimidine interruptions in the purine-rich strand. [e] Numbers correspond to the target sites in the rhodopsin gene (see Figure 1). AT stands for antiparallel GT-TFO, AA for antiparallel GA-TFO, and PT for parallel GT-TFO. [f] K_ds shown in boldface were determined by electrophoretic mobility shift analysis (TFOs 2, 3, 5, 6, 9, 15, 16, 17). For target sites where binding was not detected, TFOs were assigned K_ds $>10^{-6}$ M.

FIGURE 1: Rhodopsin gene transcription unit. Exons are shown as boxes and introns as lines. Untranslated regions are shown as filled boxes. Vertical lines indicate potential triplex binding sites screened for TFO binding.

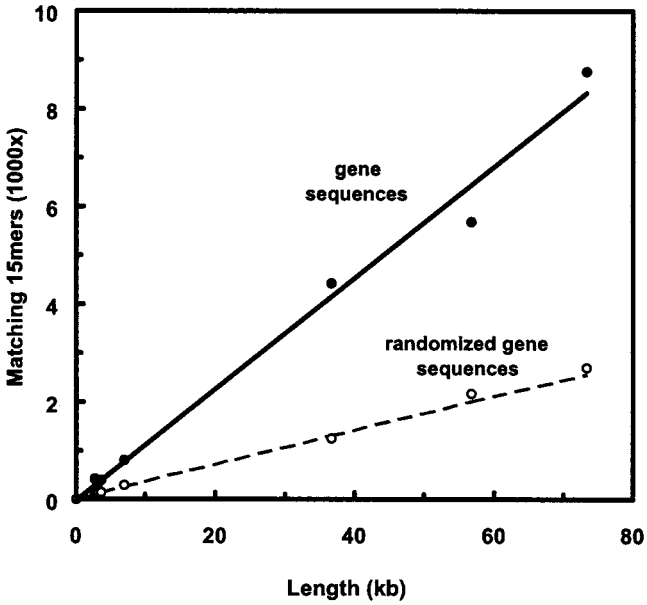

FIGURE 2: Purine-rich sequences in human genes. The number of 15mers that match the computer search criteria of 80% purine in one strand is plotted as a function of length (filled circles) for the following human genomic sequences: human growth hormone (2657 bp), adenine phosphoribosyltransferase (3016 bp), ubiquitin (3107 bp), β actin (3646 bp), rhodopsin (6953 bp), adenosine deaminase (36 741 bp), hypoxanthine phosphoribosyltransferase (56 737 bp), and the β globin locus (73 308 bp). After sequence randomization, the computer search was repeated 3 times. The average was plotted in the same manner (open circles). The dashed line represents the equation: $2l\{[15!/m!(15-m)!](1/2)^m(1/2)^{15-m}\}$, summed for 0, 1, 2, and 3 mismatches as a function of length (see Experimental Procedures).

G-content (<50%), a GT-TFO was also designed to bind parallel (designed by PT in Table 1) to the purine-rich strand (*10*, *35*−*37*). At pyrimidine interruptions, parallel GT-TFOs were designed to carry a T opposite C (*38*) and a G opposite T (*39*).

Sensitivity of the Plasmid Binding Assay. A plasmid binding assay offers a quick and efficient method to screen for triplex formation (*30*). To determine the sensitivity of this assay, we used TFO5AA and TFO5AT, which we had found in preliminary experiments to bind to their target site with high affinity (see Table 1). To generate TFOs with a range of binding affinities, we synthesized a variety of mixed TFOs (containing G, A, and T nucleotides) and TFOs in which one or two mismatched nucleotides were introduced that were not expected to form productive base-triplets. The K_ds of these TFOs were measured by quantitative electrophoretic mobility shift assays (band-shift assays). In general, mixed TFOs bound with about the same affinity as TFO5AA and TFO5AT, while TFOs with mismatches bound with lower affinity (Table 2). We chose TFO5H and TFO5I to include in the plasmid binding assays because they exhibited intermediate ($K_d = 10^{-8}$ M) and very low ($K_d = 10^{-6}$ M) binding affinities, respectively, and thus could help characterize the sensitivity of the assay.

Results of a typical plasmid fragment binding assay are shown in Figure 3. Plasmid p8/6, which contains a genomic copy of the rhodopsin gene, was digested with *Eco*RI to generate five fragments, two of which were derived from the transcription unit of the gene. The ethidium-stained gel and the corresponding autoradiograph for several TFOs are shown. As can be seen in Figure 3B, TFO5AT (lane 1), TFO5C (a mixed TFO, lane 3), and TFO5H (lane 4) gave strong signals at normal exposure, while TF05I (lane 5) was barely detectable. Upon longer exposure of this gel, the signal from TF05I was also clearly detectable (data not shown). These results suggest that the plasmid binding assay can be used to detect triplex formation for TFOs with K_ds as high as 10^{-6} M, a conclusion confirmed by subsequent results with TFO4AA and TFO15AT (see Table 1).

Screening for Triplex-Forming Sites within the Rhodopsin Gene. The plasmid binding assay was used to screen for triplex formation by 37 TFOs designed to bind to 17 target sites in the human rhodopsin gene. In all cases where triplex formation was detected, the TFO bound to the fragment of the rhodopsin gene that carried the target site, as expected. Triplex formation was detected for 14 GA- or GT-TFOs at 9 target sites (Table 1). In no instance was triplex formation detected for a parallel GT-TFO.

K_d Determination by Quantitative Electrophoretic Mobility Shift Assays. To characterize fully the binding affinities of individual TFOs, band-shift assays were performed for all TFOs directed at target sites that yielded positive results in the plasmid binding assay. Synthetic duplexes were formed by annealing oligonucleotides that contained the target sites plus 10 additional nucleotides of flanking DNA on both sides. Radiolabeled duplexes were incubated with their corresponding TFOs under standard triplex-forming conditions for 72 h to allow binding to approach equilibrium (*30*, *31*). Triplex formation was observed as an upward shift in the mobility of radiolabeled duplex on native polyacrylamide gels.

Typical band-shift analyses of four TFOs are shown in Figure 4. By quantitating the amount of radiolabel migrating as duplex and triplex at each concentration of TFO and averaging the results from several experiments, the dissociation constant (K_d) can be accurately calculated (*30*). In those cases where only the GA-TFO bound in the plasmid binding assay, the corresponding GT-TFO was also tested by band-shift analysis. Results from these experiments are shown in Table 1. At eight of the nine target duplexes that were tested, at least one TFO formed a triplex with a K_d of 1×10^{-7} M or lower. A total of 11 TFOs formed very high-affinity triplexes at 7 targeted sites, with a K_d of 1×10^{-8} M or lower. At each target site, the TFO composed only of purines (GA-TFOs) bound with higher affinity than the TFO composed of purines and pyrimidines (GT-TFOs). The G-content of the target, and therefore of the corresponding TFO, also appeared to affect binding affinities. In Figure 5, the logarithms of the K_d value are plotted as a function of percent G in the purine-rich target strand. In general, the binding affinity of TFOs tended to increase as the G-content of the target site increased.

Table 2: Effects of Mismatches on Binding Affinities of GA- and GT-TFOs

TFO	Sequence[a]	K_d (M)[b]	GA, GT, GAT[c]	MM[d]
	5′ GAGAAGGGAGAGGGAGGAAGGA 3′			
5AT	3′ GTGTTGGGTGTGGGTGGTTGGT 5′	$3x10^{-9}$	GT	0
5AA	3′ GAGAAGGGAGAGGGAGGAAGGA 5′	$1x10^{-9}$	GA	0
5B	3′ GTGTTGGGTGAGGGTGGTTGGT 5′	$2x10^{-9}$	GAT	0
5C	3′ GAGAAGGGAGTGGGAGGAAGGA 5′	$3x10^{-9}$	GAT	0
5D	3′ GAGTAGGGAGAGGGAGGTAGGA 5′	$1x10^{-8}$	GAT	0
5E	3′ GTGTTGGGTGTG**A**GTGGTTGGT 5′	$>10^{-6}$	GT	1
5F	3′ GTGTTGGGTGTG**C**GTGGTTGGT 5′	$>10^{-6}$	GT	1
5G	3′ GTGTTGGGTG**C**GGGTGGTTGGT 5′	$1x10^{-8}$	GT	1
5H	3′ GAGAAGGGAGAG**T**GAGGAAGGA 5′	$1x10^{-8}$	GA	1
5I	3′ GAGAAG**C**GAGAG**C**GAGGAAGGA 5′	$1x10^{-6}$	GA	2
5J	3′ GTGTTGGGTGTG**T**GTGGTTGGT 5′	$>10^{-6}$	GT	1

[a] The target sequence is shown in boldface on the first line. Mismatched nucleotides are shown as boldface letters within the TFO sequence. [b] K_ds were determined by electrophoretic mobility shift analysis. [c] TFOs were designed to bind antiparallel to the purine-rich strand of the target using either G and A (GA), G and T (GT), or G, A, and T (GAT) nucleotides. [d] Numbers represent the number of mismatched nucleotides in the TFO.

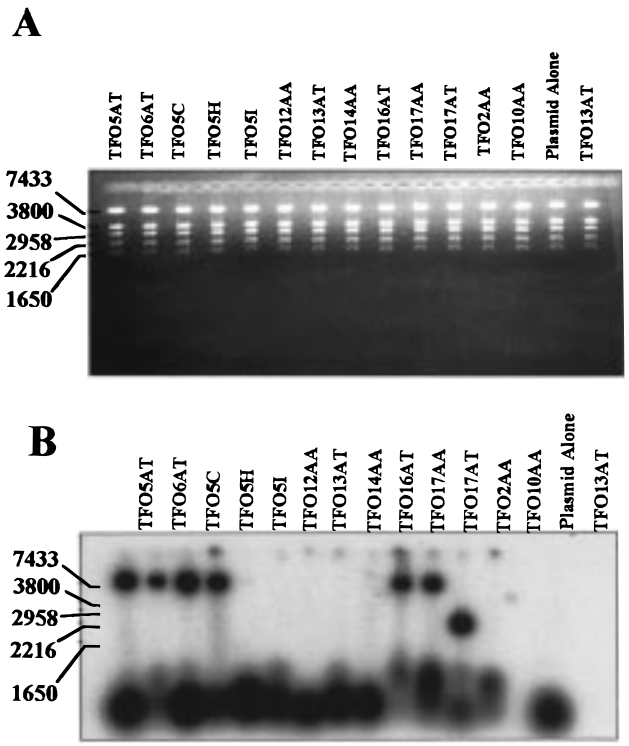

FIGURE 3: Plasmid binding assay. (A) Ethidium bromide stained gel of samples containing *Eco*RI-digested plasmid and radiolabeled TFOs that had been incubated under conditions suitable for triplex formation. (B) Corresponding autoradiograph of gel from part A. A TFO that forms a triplex with the plasmid migrates with the band containing the target site (7.4 and 2.2 kb fragments).

DISCUSSION

A prerequisite for application of triplex technology to gene therapy is to identify triplex-forming sites in target genes. We surveyed the 6.9 kb rhodopsin gene for potential sites and characterized all that were at least 17 nucleotides long. Seven sites, about one per kbp, bound TFOs with high affinity (K_d 10^{-8} M). These results are consistent with the observation that long homopurine sequences are significantly more abundant in human genes than expected (*40, 41*).

Application to ADRP. Application of triplex technology to dominant disorders such as ADRP requires that the defective gene be inactivated or repaired by TFOs designed to stimulate recombination, inhibit transcription, or introduce mutations. The most effective treatment would likely be with TFOs modified to allow covalent modification of the targeted gene, for example, by photo-cross-linking with psoralens (*42*). While delivery of oligonucleotides to specific tissues is a technical challenge, successful delivery to the retina via intravitreal injection has been reported (*43, 44*), with low toxicity and inflammation. In the case of the rhodopsin gene, in which more than 70 distinct mutations are associated with ADRP, it may advantageous to develop generic, gene-specific TFOs, such as the intron-directed TFOs described here, that can serve in general therapies for ADRP.

The most attractive treatment possibility is the use of modified TFOs to sensitize the rhodopsin gene for targeted recombination. If a damage-inducing TFO were accompanied by the appropriate segment of the rhodopsin gene, targeted correction of the mutant gene could be achieved. In tissue culture, the introduction of double-strand breaks has been shown to stimulate recombination several 1000-fold in the region around the break (*45, 46*). For this approach it is not critical that recombination be directed specifically to the defective allele because gene targeting at the wild-type allele would be without consequence.

Gene inactivation by inhibition of transcription may be an alternative treatment for ADRP. While generic TFOs could not discriminate between wild-type and mutant genes, inactivation of, e.g., 50% of all rhodopsin genes in a portion of the retina could be expected to produce a statistical distribution of cells whose transcriptionally active rhodopsin genes consist of one good copy (likely a rescued cell), one bad copy (at least as bad as the starting situation, but possibly no worse), no copies (possibly no worse than the starting situation), and one bad copy and one good copy (identical to the starting situation). If this change in distribution could slow the pace of retinal degeneration as much as 20%, it would likely be of great therapeutic benefit.

The locations of the seven high-affinity binding sites within introns make it unlikely that TFOs could be designed to introduce inactivating mutations into the rhodopsin gene,

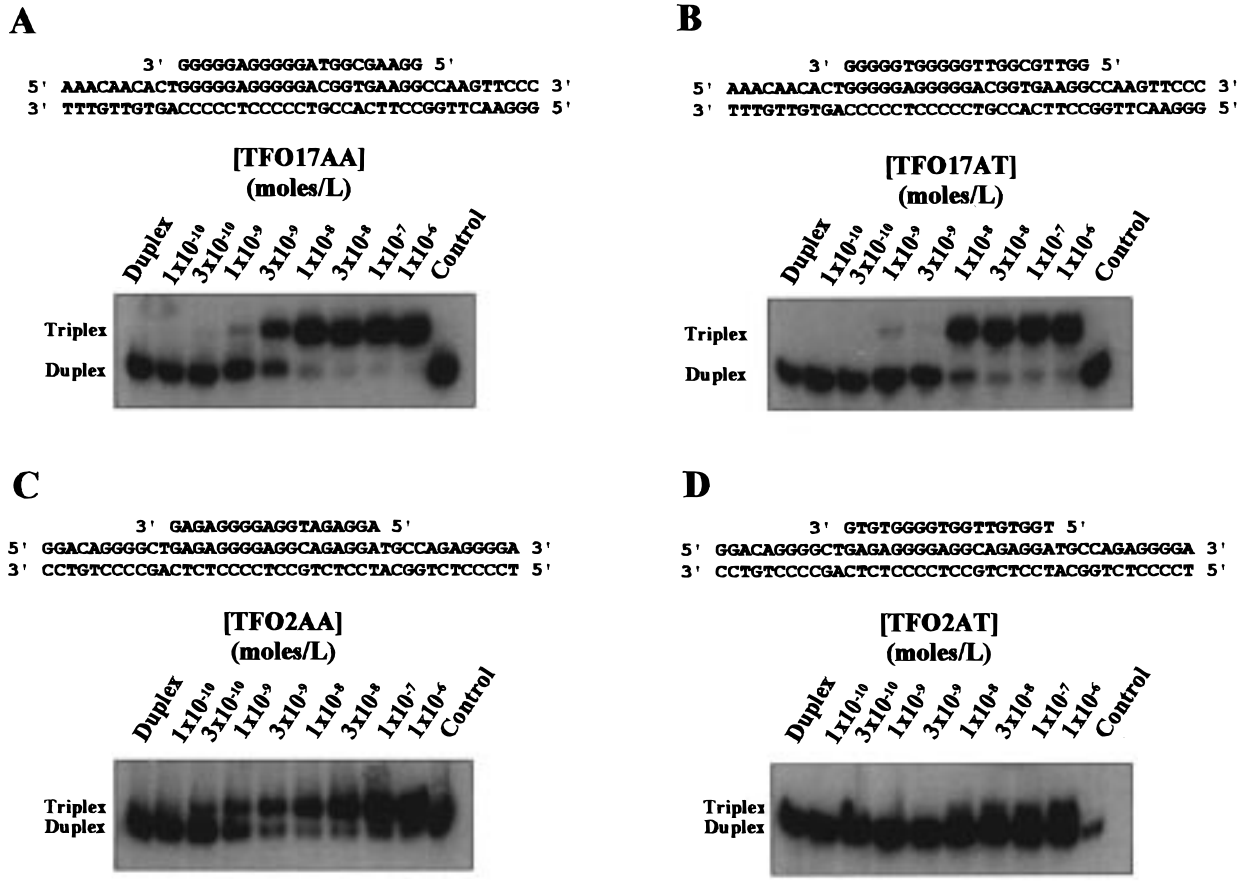

FIGURE 4: Band-shift analysis of TFO binding to synthetic duplexes. (A) TFO17AA; (B) TFO17AT; (C) TFO2AA; and (D) TFO2AT. Target sequences with the corresponding TFO are shown above each autoradiograph. "Duplex" indicates radiolabeled duplex alone with no added TFO. "Control" indicates a TFO that is not specific for the target duplex. Molar concentrations of TFOs in the incubations are indicated above the appropriate lane.

much less to correct particular ADRP mutations. Therefore, recombinational and transcriptional approaches are the methods of choice for further study. Fortunately, both approaches are readily testable in tissue culture and animal models.

High G-Content of Target Site Facilitates Triplex Formation. As noted previously (*34*, *47*, *48*), G-content is important for high-affinity triplex formation (Figure 5). All the sites that bound TFO with high affinity ($K_d \leq 1 \times 10^{-8}$ M) had G-contents of at least 54% (Table 1), and only two TFOs (4AA and 16AA) bound detectably to sites with G-contents below 50%. Binding affinities tended to increase as the G-richness of the target increased, and all targets with 58% or greater G-content bound GA-TFOs with high affinity (Figure 5).

GA- vs GT-TFOs. Previous studies comparing binding of GA- and GT-TFOs at individual sites have reported that GA-TFOs bind much better (*49*−*53*), within a factor of 3−4 (*34*, *50*, *54*, *55*), or much worse (*56*) than GT-TFOs. Some NMR results have been interpreted as indicating that A-AT triples may be weaker than T-AT triples (*57*, *58*). In our studies at 17 different target sites, we have observed a consistent result; at each of the 9 sites where triplex formed, the GA-TFO bound with higher affinity than the corresponding GT-TFO. In our studies, all experiments were carried out at 37 °C, to avoid self-association of some GA-TFOs (*59*), which can reduce their apparent triplex-forming activity (*60*, *61*).

The physical basis for the higher affinities of GA-TFOs (Table 1) is unclear, but there are several possibilities (*53*). Irregularities introduced into the backbone of the third strand when adjacent nucleotides are different (*36*, *62*) may be energetically less favorable at G−T and T−G junctions than at G−A and A−G junctions. In addition, Ts may make weaker hydrogen bonds to the underlying duplex than do As, or they may stack less favorably than As with adjacent nucleotides in the third strand.

GT-TFOs May Bind with Higher Specificity than GA-TFOs. In general, target sites with pyrimidine interruptions in the purine-rich strand formed lower affinity triplexes than sites with comparable G-contents and no pyrimidine interruptions. GT-TFOs may be more sensitive to such interruptions, since two of the four GT-TFOs that bound with reasonably high affinities did so at sites with no pyrimidine interruptions (Table 1: TFO3 and TFO5; Figure 5). When mismatches were deliberately introduced into TFOs that bound with very high affinity (Table 2: TFO5AT, TFO5AA), the GT-TFO was more drastically affected. The GA-TFO with one mismatch still bound with high affinity, although reduced 10-fold (Table 2: TFO5H), while the same mismatch in a GT-TFO completely eliminated binding, implying a greater than 300-fold reduction in affinity (Table 2: TFO5J).

To the extent that these results can be generalized, they have important implications for therapeutic applications of GA- and GT-TFOs. The relative insensitivity of GA-TFOs to pyrimidine interruptions in the purine-rich strand of the

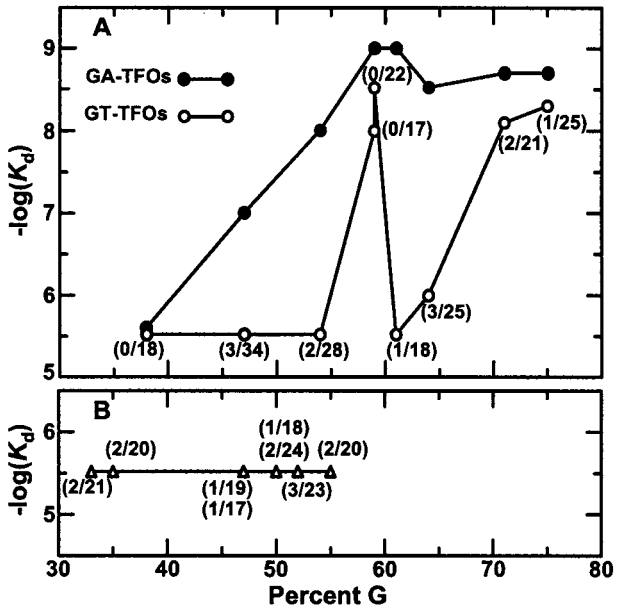

FIGURE 5: Effects of G-content on triplex binding. The negative logarithm of the dissociation constant is plotted as a function of G-content for each target site that exhibited triplex formation with at least one TFO by the plasmid binding assay. Filled circles represent GA-TFOs; open circles represent GT-TFOs. The first number in parentheses is the number of inversion sites in the purine-rich strand of the target, while the second number is the length of the target site. TFOs whose binding was too weak to be detected are arbitrarily plotted with a K_d value of $10^{-5.5}$ M.

target suggests that it will be easier to find target sites for high-affinity GA-TFOs than for high-affinity GT-TFOs. For the rhodopsin gene, we found seven high-affinity sites for GA-TFOs but only four bound GT-TFOs with high affinity. The greater apparent sensitivity of GT-TFOs to mismatches in the TFO suggests that GT-TFOs may discriminate better than GA-TFOs between target sites and closely related nontarget sites elsewhere in the genome. However, single mismatch discrimination has been reported for high-affinity triplex formation by GA-TFOs (*63*). The issue of specificity is important, as TFO interactions at nontarget sites could lead to unacceptable side effects. When the human genome sequence is known, a rational choice of GA- or GT-TFOs may be possible based on the sequences of related, nontargeted sequences elsewhere in the genome.

ACKNOWLEDGMENT

We thank Dr. Tiansen Li (Harvard Medical School) for the plasmid used in these experiments. We also thank Dr. Geoff Sargent for useful comments on the manuscript.

REFERENCES

1. Frank-Kamenetskii, M. D., and Mirkin, S. M. (1995) *Annu. Rev. Biochem. 64,* 65−95.
2. Vasquez, K. M., and Wilson, J. H. (1998) *Trends Biochem. Sci. 23,* 4−9.
3. Ing, N. H., Beekman, J. M., Kessler, D. J., Murphy, M., Jayaraman, K., Zendegui, J. G., Hogan, M. E., BW, O. M., and Tsai, M. J. (1993) *Nucleic Acids Res. 21,* 2789−2796.
4. Kovacs, A., Kandala, J. C., Weber, K. T., and Guntaka, R. V. (1996) *J. Biol. Chem. 271,* 1805−1812.
5. Svinarchuk, F., Debin, A., Bertrand, J. R., and Malvy, C. (1996) *Nucleic Acids Res. 24,* 295−302.
6. Kochetkova, M., and Shannon, M. F. (1996) *J. Biol. Chem. 271,* 14438−14444.
7. Orson, F. M., Thomas, D. W., McShan, W. M., Kessler, D. J., and Hogan, M. E. (1991) *Nucleic Acids Res. 19,* 3435−3441.
8. Postel, E. H., Flint, S. J., Kessler, D. J., and Hogan, M. E. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88,* 8227−8231.
9. McShan, W. M., Rossen, R. D., Laughter, A. H., Trial, J., Kessler, D. J., Zendegui, J. G., Hogan, M. E., and Orson, F. M. (1992) *J. Biol. Chem. 267,* 5712−5721.
10. Scaggiante, B., Morassutti, C., Tolazzi, G., Michelutti, A., Baccarani, M., and Quadrifoglio, F. (1994) *FEBS Lett. 352,* 380−384.
11. Okada, T., Yamaguchi, K., and Yamashita, Y. (1994) *Growth Factors 11,* 259−270.
12. Thomas, T. J., Faaland, C. A., Gallo, M. A., and Thomas, T. (1995) *Nucleic Acids Res. 23,* 3594−3599.
13. Tu, G. C., Cao, Q. N., and Israel, Y. (1995) *J. Biol. Chem. 270,* 28402−28407.
14. Porumb, H., Gousset, H., Letellier, R., Salle, V., Briane, D., Vassy, J., Amor-Gueret, M., Israel, L., and Taillandier, E. (1996) *Cancer Res. 56,* 515−522.
15. Rininsland, F., Johnson, T. R., Chernicky, C. L., Schulze, E., Burfeind, P., Ilan, J., and Ilan, J. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94,* 5854−5859.
16. Aggarwal, B. B., Schwarz, L., Hogan, M. E., and Rando, R. F. (1996) *Cancer Res. 56,* 5156−5164.
17. Sandor, Z., and Bredberg, A. (1994) *Nucleic Acids Res. 22,* 2051−2056.
18. Wang, G., Levy, D. D., Seidman, M. M., and Glazer, P. M. (1995) *Mol. Cell. Biol. 15,* 1759−1768.
19. Wang, G., Seidman, M. M., and Glazer, P. M. (1996) *Science 271,* 802−805.
20. Faruqi, A. F., Seidman, M. M., Segal, D. J., Carroll, D., and Glazer, P. M. (1996) *Mol. Cell. Biol. 16,* 6820−6828.
21. Sandor, Z., and Bredberg, A. (1995) *Biochim. Biophys. Acta 1263,* 235−240.
22. Shastry, B. S. (1994) *Am. J. Med. Genet. 52,* 467−474.
23. Berson, E. L. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93,* 4526−4528.
24. Dryja, T. P., and Li, T. (1995) *Hum. Mol. Genet. 4,* 1739−1743.
25. Heckenlively, J. R. (1988) *Retinitis Pigmentosa*, J. B. Lippincott, Philadelphia.
26. Nathans, J. (1992) *Biochemistry 31,* 4923−4931.
27. Rao, V. R., and Oprian, D. D. (1996) *Annu. Rev. Biophys. Biomol. Struct. 25,* 287−314.
28. Rosenfeld, P. J., Cowley, G. S., McGee, T. L., Sandberg, M. A., Berson, E. L., and Dryja, T. P. (1992) *Nat. Genet. 1,* 209−213.
29. Humphries, M. M., Rancourt, D., Farrar, G. J., Kenna, P., Hazel, M., Bush, R. A., Sieving, P. A., Sheils, D. M., McNally, N., Creighton, P., Erven, A., Boros, A., Gulya, K., Capecchi, M. R., and Humphries, P. (1997) *Nat. Genet. 15,* 216−219.
30. Vasquez, K. M., Wensel, T. G., Hogan, M. E., and Wilson, J. H. (1995) *Biochemistry 34,* 7243−7251.
31. Xodo, L. E., Pirulli, D., and Quadrifoglio, F. (1997) *Eur. J. Biochem. 248,* 424−432.
32. Joseph, R. M., and Li, T. (1996) *Invest. Ophthalmol. Vis. Sci. 37,* 2434−2446.
33. Helene, C. (1991) *Anti-Cancer Drug Des. 6,* 569−584.
34. Chandler, S. P., and Fox, K. R. (1996) *Biochemistry 35,* 15038−15048.
35. Sun, J. S., De Bizemont, T., Duval-Valentin, G., Montenay-Garestier, T., and Helene, C. (1991) *C. R. Acad. Sci., Ser. III 313,* 585−590.
36. Giovannangeli, C., Rougee, M., Garestier, T., Thuong, N. T., and Helene, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 8631−8635.
37. de Bizemont, T., Duval-Valentin, G., Sun, J. S., Bisagni, E., Garestier, T., and Helene, C. (1996) *Nucleic Acids Res. 24,* 1136−1143.
38. Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., and Weis, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89,* 3840−3844.

39. Griffin, L. C., and Dervan, P. B. (1989) *Science 245*, 967−971.
40. Behe, M. J. (1987) *Biochemistry 26*, 7870−7875.
41. Behe, M. J. (1995) *Nucleic Acids Res. 24*, 689−695.
42. Vasquez, K. M., Wensel, T. G., Hogan, M. E., and Wilson, J. H. (1996) *Biochemistry 35*, 10712−10719.
43. Rakoczy, P. E., Lai, M. C., Watson, M., Seydel, U., and Constable, I. (1996) *Antisense Nucleic Acid Drug Dev. 6*, 207−213.
44. Flores-Aguilar, M., Besen, G., Vuong, C., Tatebayashi, M., Munguia, D., Gangan, P., Wiley, C. A., and Freeman, W. R. (1997) *J. Infect. Dis. 175*, 1308−1316.
45. Rouet, P., Smih, F., and Jasin, M. (1994) *Mol. Cell. Biol. 14*, 8096−8106.
46. Sargent, R. G., Brenneman, M. A., and Wilson, J. H. (1997) *Mol. Cell. Biol. 17*, 267−277.
47. Clarenc, J.-P., Lebleu, B., and Leonetti, J.-P. (1994) *Nucleosides Nucleotides 13*, 799−809.
48. Malkov, V. A., Voloshin, O. N., Veselkov, A. G., Rostapshov, V. M., Jansen, I., Soyfer, V. N., and Frank-Kamenetskii, M. D. (1993) *Nucleic Acids Res. 21*, 105−111.
49. Faucon, B., Mergny, J. L., and Helene, C. (1996) *Nucleic Acids Res. 24*, 3181−3188.
50. Gamper, H. B. J., Kutyavin, I. V., Rhinehart, R. L., Lokhov, S. G., Reed, M. W., and Meyer, R. B. (1997) *Biochemistry 36*, 14816−14826.
51. Lacoste, J., Francois, J. C., and Helene, C. (1997) *Nucleic Acids Res. 25*, 1991−1998.
52. Roy, C. (1994) *Eur. J. Biochem. 220*, 493−503.
53. Scaria, P. V., and Shafer, R. H. (1996) *Biochemistry 35*, 10985−10994.
54. Beal, P. A., and Dervan, P. B. (1991) *Science 251*, 1360−1363.
55. Mayfield, C., Squibb, M., and Miller, D. (1994) *Biochemistry 33*, 3358−3363.
56. Washbrook, E., and Fox, K. R. (1994) *Nucleic Acids Res. 22*, 3977−3982.
57. Radhakrishnan, I., and Patel, D. J. (1993) *Structure 1*, 135−152.
58. Radhakrishnan, I., de los Santos, C., and Patel, D. J. (1993) *J. Mol. Biol. 234*, 188−197.
59. Noonberg, S. B., Francois, J. C., Garestier, T., and Helene, C. (1995) *Nucleic Acids Res. 23*, 1956−1963.
60. Alunni-Fabbroni, M., Manfioletti, G., Manzini, G., and Xodo, L. E. (1994) *Eur. J. Biochem. 226*, 831−839.
61. Alunni-Fabbroni, M., Pirulli, D., Manzini, G., and Xodo, L. E. (1996) *Biochemistry 35*, 16361−16369.
62. Radhakrishnan, I., and Patel, D. J. (1994) *Structure 2*, 395−405.
63. Svinarchuk, F., Bertrand, J. R., and Malvy, C. (1994) *Nucleic Acids Res. 22*, 3742−3747.

BI980525S