Sequence-specific targeting and covalent modification of human genomic DNA

Evgeniy S. Belousov, Irina A. Afonina, Mikhail A. Podyminogin, Howard B. Gamper, Michael W. Reed, Robert M. Wydro and Rich B. Meyer*

Epoch Pharmaceuticals, Inc., 1725 220th Street SE, #104, Bothell, WA 98021, USA

Received May 27, 1997; Revised and Accepted July 14, 1997

ABSTRACT

We compare two techniques which enable selective, nucleotide-specific covalent modification of human genomic DNA, as assayed by quantitative ligation-mediated PCR. In the first, a purine motif triplex-forming oligonucleotide with a terminally appended chlorambucil was shown to label a target guanine residue adjacent to its binding site in 80% efficiency at 0.5 µM. Efficiency was higher in the presence of the triplex-stabilizing intercalator coralyne. In the second method, an oligonucleotide targeting a site containing all four bases and bearing chlorambucil on an interior base was shown to efficiently react with a specific nucleotide in the target sequence. The targeted sequence in these cases was in the DQβ1*0302 allele of the MHC II locus.

INTRODUCTION

Site-directed mutagenesis by specific modification of genomic DNA offers a viable alternative to vector-based gene therapy in cases where single base changes are useful. We have been developing methods for DNA sequence alteration using oligodeoxynucleotides (ODNs) with reactive agents attached to one or more positions. These may cause mutation at the targeted site, induced by covalent modification (1,2), resulting in inactivation or modification of functional expression of the gene.

One method of DNA sequence targeting is with triplex-forming oligonucleotides (TFOs). If these are conjugated to electrophilic or photoreactive groups, they can recognize and efficiently modify bases on either one (1,3–6) or both (7,8) strands in targeted duplex DNA. A current difficulty with TFOs is that only purines may be in the TFO binding strand of the target sequence after complex formation.

An alternate method, useful for targeting any sequence, is use of the Escherichia coli RecA protein with single-stranded ODNs. This protein coats single-stranded DNA, and catalyzes a homology search on double-stranded DNA targets (9). This can occur when the single strand is a fairly short ODN. Oligonucleotides have been used with RecA protein to mask specific methylation of restriction sites, leaving those targeted sites sensitive to restriction after the other sites were modified (10), and to retrieve targeted plasmids from a mixture by using biotinylated ODNs, with capture of the biotinylated ODN-bound plasmids on a streptavidin column (11). In the latter case, supercoiled plasmids were required. These three-stranded complexes can be stable after deproteinization, however, if the single-stranded ODN bears a reactive group that covalently binds to the target’s complementary strand, as we have previously shown (12,13). This strand displacement method allows modification of any target site (four letter targeting) in DNA without sequence restriction by use of such a recombinase-assisted ODN (RAO).

In this study we show that both of these two gene targeting methods give highly efficient and specific site-directed electrophilic modification of a native gene in whole genomic human DNA. The target for the crosslinking TFO is a polypurine tract in the DQβ1*0302 allele of the MHC II locus, and for the RAO, a nearby site on the same gene. Quantitative ligation-mediated PCR (LMPCR) was used to monitor the extent to which the targeted nucleotide was alkylated. With either technique, efficient ODN-directed modification was shown to have occurred at a significant rate with good specificity.

MATERIALS AND METHODS

Oligonucleotide design, synthesis, conjugation and labeling

Figure 1 shows the alignment of the ODNs with the targeted sequences, the sites of modification and the structure of the oligonucleotides used in this study. The target sequence for both types of ODN was the first intron of the HLA DQβ1*0302 allele, an allele associated with predisposition to juvenile onset (Type I) diabetes (14), and the human HT-29 cells (obtained from ATCC) used here contain a single copy of this allele. We found the sequence recorded in GenBank (15) (accession no. K01499) to be incorrect; Figure 1 shows the correct sequence in the area we are targeting. When a TFO with a terminal alkylating group forms a triplex with double-stranded DNA, it reacts with nearby guanines at N7 in the major groove (4,7). In RAOs the reactive moiety, located on an internal base, preferentially reacts with guanines on the complementary strand of the target sequence after synaptic complex formation.

ODNs with a 5’ or 3’ aminohexyl group were prepared using N-MMT-hexanamine phosphoramidite linker (Glen Research) for the 5’-modification, or aminohexyl controlled pore glass (16) for the 3’-modifications. ODNs with an internal aminopropyl

*To whom correspondence should be addressed. Tel: +1 425 485 8566; Fax: +1 425 486 8336; Email: rmeyer@epochpharm.com
Dried and dissolved in 0.2 ml of anhydrous DMSO. After conjugation to chlorambucil, an aliquot containing 1 mg of ODN was dried overnight. The ODN peak at 20 min (flow rate = 2 ml/min) gave the column (Hamilton). A gradient of 0–60% 0.1 M acetonitrile in water and injecting it onto a PRP-1 HPLC converted to the triethylammonium (TEA) salt by dissolving the ODN in 0.5 ml of water and purified by C18 HPLC (5–85% acetonitrile in 0.1 M TEA acetate, 40 min at 1 ml/min, ~20 min retention). The product was immediately precipitated with 100 µl of 3 M sodium acetate and 4 ml of butanol for each milliliter of eluent collected. The precipitate was pelleted without delay (3000 g for 5 min), washed with 2 ml ethanol, and re-pelleted. This pellet was dried for 10 min on a centrifugal evaporator, dissolved in 0.2 ml of water, and assayed for concentration and purity by C18 HPLC. All conjugates were at least 90% pure. They were stored at −80°C, where they were stable for weeks.

**Figure 1.** Systems used in demonstration of modification of double-stranded genomic DNA by reactive oligonucleotides. (A) Schematic of the region of DNA targeted and position of LMPCR primers. TFO is designed to alkylate top (sense) strand; RAO is designed to alkylate bottom strand. The arrows marking the LMPCR primer sites schematically represent the regions where the primers for each strand are nested, such that primer 1 binds at the 3′-most position (on the copied strand) of each set, with the others closely 5′-most position (or occasionally overlapping) to the previous primer. (B) TFO1 and target sequence in a polyurine stretch of the HLA DQB1*0302 allele. Chlorambucil is appended to the 5′-end of the ODN. (C) RAO2, with chlorambucil appended to an internal dUrd, and target region on same gene. RAO1 covers the same sequence but is 50 nt long.

Group (the RAOs used here) were prepared using the protected phosphoramidite of 5′-(3-aminopropyl)-2′-deoxyuridine (17). The synthesis, HPLC purification, detritylation and butanol precipitation of both types of ODNs were carried out using standard procedures as previously described (18). Drying of the ODN solutions was performed under vacuum using a centrifugal evaporator.

For conjugation, ODN (from a 2′-end of the ODN) was extracted three times with ether and the DNA precipitated as 50% glycerol, along with 10 µl of 10 mM MgCl2 and 1 mM spermine was added 10 µl of a 10x stock of TFO1 or controls to give a final ODN concentration of 10⁻⁶–10⁻⁹ M. After mixing and incubation 3 h at 37°C, the DNA was pelleted by addition of 10 µl of 3 M NaOAc, pH 7.0, and 300 µl of ice-cold 100% ethanol, chilling (−70°C), centrifugation at 12 000 r.p.m. for 15 min at 4°C, washed with EtOH, and dried.

**Reaction of triple-forming ODNs with genomic DNA**

Genomic DNA from HT29 cells was prepared with a Wizard Genomic DNA Purification Kit (Promega). To 5–10 µg of genomic DNA in 90 µl 20 mM HEPES, pH 7.2, 140 mM KCl, 10 mM MgCl₂ and 1 mM spermine was added 10 µl of a 10x stock of TFO1 or controls to give a final ODN concentration of 10⁻⁶–10⁻⁹ M. After mixing and incubation 3 h at 37°C, the DNA was pelleted by addition of 10 µl of 3 M NaOAc, pH 7.0, and 300 µl of ice-cold 100% ethanol, chilling (−70°C), centrifugation at 12 000 r.p.m. for 15 min at 4°C, washed with EtOH, and dried.

**Reaction of ODNs with genomic DNA with RecA protein catalysis**

To 10 µl of an ice cold 1 µM chlorambucil–ODN (RAO1 or RAO2) solution in 10 mM Tris–HCl, pH 7.5, and 1 mM EDTA was added 10 µl of 10x RecA buffer [100 mM Tris-acetate, pH 7.5, 500 mM NaOAc, 120 mM Mg(OAc)₂, 10 mM dithiothreitol (DTT) and 50% glycerol], along with 10 µl of 10 mM ATPγS and water to give a final volume of 87 µl, which was kept ice cold. After addition of 3 µl of a cold solution of RecA protein (New England Biolabs #249L) solution (2 mg/ml), 5–10 µg of genomic DNA was added and the mixture was incubated for 6 h at 37°C. After addition of 82 µl of 10 mM Tris–HCl, pH 7.5, containing 1 mM EDTA, 10 µl of 10% SDS and 8 µl of Proteinase K (5 mg/ml), and incubation for 30 min at 37°C, the reaction mixture was extracted with an equal volume of phenol–chloroform. The aqueous phase was extracted three times with ether and the DNA precipitated as above.

**Quantitative ligation-mediated PCR (LMPCR)**

Most steps of this technique were performed as described (20,21), although certain modifications (22) were made for this work. Since we required a method to quantitate the amount of site-specific alkylation and cleavage, we generated an internal control site by cleavage of the DNA with a restriction enzyme after the chlorambucil–ODN reaction. Selection of a restriction endonuclease was based on the enzyme having a recognition site downstream (3′) sequence complementary to the first primer. A BamHI site for each strand was identified as shown in the schematic in Figure 1. Initial experiments confirmed completion of digestion at these sites, using Southern blot hybridization and a PCR-generated probe, and also confirmed the lack of dependence of the ratio of target
site cleavage to restriction site cleavage on amount of treated DNA (0.5–5 µg tested) used in the LMPPCR protocol.

The second modification to the protocol was heating the DNA to 95°C in the buffer used for first strand synthesis, pH 8.9, for 10 min prior to annealing and extending the first primer. This is slightly longer than the standard 3 min of the protocol (21), and causes quantitative depurination and cleavage of the DNA at the site of alkylation with the generation of the 5'-phosphate required for the ligation step. All chlorambucil alkylation sites in this work were designed to be on N-7 of guanines.

For each experiment an aliquot of the genomic DNA was treated with dimethylsulfate and amplified along with the rest of the samples to provide a G-ladder (20). The DNA samples were restricted with BamHI to completion by incubating 3 h under optimal conditions with a 3-fold excess of restriction enzyme. The volume was adjusted to 100 µl with water and the DNA was precipitated as described above. The pellet was resuspended in 10 mM Tris–HCl, pH 7.5, 1 mM EDTA, to give a DNA concentration of ~0.5 µg/µl. To 5 µl of this chilled solution in a PCR tube was added 25 µl of the primer 1 solution (21). First strand synthesis and ligation of the universal linker (see above), was performed as described by Mueller and co-workers (21), except that the 95°C heating step was extended to 10 min. Nested PCR was then performed as described by Pfeifer and Riggs (20).

Phosphorimaging was used to analyze the polyacrylamide gel electropherograms of the LMPPCR results, and the ratio of intensities of the bands due to the restriction and alkylation sites gave the efficiency of targeted alkylation.

**RESULTS**

**Targeted DNA modification by a reactive TFO**

Figure 2 shows the results of incubation of TFO1 with isolated genomic DNA from human HT29 cells in physiological buffer, as analyzed by LMPPCR. The BamHI (internal standard) site was located 36 bases 5’ from the target site. Since we showed by Southern blot that 100% of the DNA was cleaved by BamHI, the efficiency of cleavage at the targeted site is given by the ratio of the density in the cleavage site to the sum of the density in the cleavage and the restriction site, assuming negligible background. Table 1 gives this quantitation of specific targeting.

The incubations were performed in the presence or absence of the triplex-stabilizing agent coralyne, which had a significant effect on the efficiency of alkylation. Lee and co-workers (23) have shown that coralyne stabilized pyrimidine motif triplexes, and we have observed that coralyne stabilizes triplexes formed using G/A and G/T motif TFOs (H.Gamper and J.Lampe, unpublished results). In the presence of coralyne, the sequence-specific modification of the target nucleotide is clearly seen at concentrations as low as 4 nM (7%). At increasing concentrations, the selective alkylation of the target site increases, with 90% efficiency seen at a concentration of 500 nM. Even without the triplex stabilization by coralyne, the efficiency of modification is high. At 500 nM 80% modification was achieved with ODN alone.

These assays were run in buffers that approximate physiological concentrations of major cations. The effect of potassium ion on triplex formation is variable, depending on the motif of the TFO being used. We (H.Gamper, unpublished results) and others (24,25) have found that potassium ion significantly destabilizes triplexes using purine motif (G/A or G/T) TFOs. In all of the hybridizations conducted in this study using genomic DNA, we used 140 mM KCl in the buffer to achieve physiological ionic strength and approximate conditions within the nuclei of cells (26). These cell-free results, therefore, may be more predictive of the cell culture activity of TFOs than studies performed in buffers with no monovalent cations.

<table>
<thead>
<tr>
<th>Lane in Figure 2</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TFO1 (nM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No coralyne</td>
<td>4</td>
<td>20</td>
<td>100</td>
<td>500</td>
<td>4</td>
<td>20</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>8 µM coralyne</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Targeted alkylation and cleavage</td>
<td>1</td>
<td>5</td>
<td>36</td>
<td>80</td>
<td>7</td>
<td>28</td>
<td>75</td>
<td>90</td>
</tr>
</tbody>
</table>

Figure 2. Alkylation of the HLA DQβ 1*0302 allele in human genomic DNA by TFO1, a chlorambucil–ODN conjugate. A homopurine run in the first intron of this single copy gene, present in DNA isolated from HT-29 cells, was targeted using a reactive 21mer (5’-chlorambucil-AGGAGAAAGGAGGGAGAGAG). Following incubation for 3 h at 37°C with addition of 8 µM coralyne (lanes 6–9) or without any coralyne (lanes 2–5), the DNA was ethanol precipitated and restricted with BamHI. Quantitative LMPPCR was performed using the following primers (sense strand): 5’-CCATAAATTTAGTCCAGGC (primer 1), 5’-GTCTTGAGGGACGCCATAATGAA (primer 2), and 5’-32P-CTGCTG-TGTCTGAGGGAAATTCGTATCTC (primer 3). Final concentrations of reactive ODN were: lane 1, no ODN; lanes 2 and 6, 4 nM; lanes 3 and 7, 20 nM; lanes 4 and 8, 100 nM; lanes 5 and 9, 500 nM; lane G, G-ladder.
in the case of the 50mer RAO1 (~50%) and moderately efficient in the case of the 30mer RAO2. We believe that the alkylation reaction occurs within a synaptic complex (28) after strand exchange, and that the RAO is in a complex resembling a Watson–Crick duplex with the complementary strand, such that adjacent major groove nucleophilic sites on the complementary strand of target DNA are accessible to reactive side chains on the RAOs (12,13). The fairly long crosslinking side chain used here can span at least two bases to either side of such a duplex. The complete lack of reaction at the target site in the absence of RecA protein is evidence that the reaction is not occurring by duplex formation of RAO with denatured DNA.

DISCUSSION

Two versatile techniques by which naked genomic DNA may be covalently modified with high efficiency and at specifically targeted nucleotides have been described here. Previous demonstrations of efficient targeting at this level of complexity have been limited to sites engineered in chromosomal DNA (8) and to integrated viral sequence in cells which were photochemically modified with lower efficiency (29). We now show the ability to target a site in a native gene with a triplex forming ODN with an appended electrophilic group with physiological salt concentrations. We further show the ability to target any desired sequence, by design, using reactive oligonucleotides and RecA protein catalysis.

Use of TFOs continues to be limited by their ability to ‘read’ only two letters of the genetic code. Although work in other laboratories has attempted to expand this code by incorporating in a TFO the ability to recognize pyrimidine interruptions in the polypurine strand (30–32) or to recognize two short purine tracts on alternate strands of the DNA (33), a general solution to recognition of DNA by triplex forming oligonucleotides is not likely to be available in the near term. Given these limitations, however, one can recognize and label targeted sites in genomic DNA at low concentrations in very high efficiency under physiological conditions. The addition of a triplex-specific intercalator that stabilizes the triplex (23,34,35) gave an increase in this efficiency and demonstrates the applicability of these agents to complex nucleic acid systems. With a view to cell culture work with these ODNs, we have found that the concentration of the intercalator we used, 8 µM coralyne, is non-toxic to cells in culture (data not shown).

Ultimately, the targeting of user-designed sites in DNA should have no limitations regarding sequence. The method shown here is an effective method for accomplishing this: the use of RecA protein to catalyze the homology search and strand exchange by ODNs, enabling the covalent crosslinking to the targeted site. Our previous demonstration (13) of efficiency and selectivity of this method in plasmid DNA is now shown applicable to whole genomic DNA. We had shown that the length of the oligonucleotide needed to be at least 30 bases (13). In the more complex system used here, efficiency may benefit by using an ODN a few more bases longer, such as the 50mer RAO1 in Figure 3.

The utility of agents that specifically label DNA is broad. These reactive ODNs could see immediate use in the isolation of rare sequences in complex dsDNA. For example, using chlorambucil-bearing ODNs with a biotin tag, it would be straightforward to affinity-capture complementary sequences of dsDNA which are hybridized and crosslinked to the ODN. Release of the captured DNA could be effected by depurination and strand scission of the

**Targeted modification of genomic DNA with reactive four-letter ODNs**

Figure 3 shows the results of targeting a specific site in the HLA DQβ1*0302 allele in human genomic DNA by chlorambucil–ODN conjugates. A site in the first intron of this single copy gene was targeted using a RecA protein coated, homologous 30mer RAO2 [5′-GAAGATCGAATTTCCAGAGACACA-GCA] or 50mer RAO1 [5′-GGTATTGAGAGATGCGATTTCCAGA-GAGACACAGGGATTTGCTA], where U is 5′-[3-{4-([5-[(2-chloro-ethyl)amino]phenyl]butyramido}propyl]deoxyuridylate. After 6 h incubation RecA protein was removed by phenol extraction and the DNA was restricted with BamHI. LMPCR was performed using the following primers (antisense strand): 5′-ATCCCCATCTACAGGCT (primer 1), 5′-GCTGGAGAGAGA-AGGAGAGAGGAG (primer 2) and 5′-32P-GAGGACACAAGTGACAT-TTACTACAGGTG (primer 3). Lane 1, no ODN or RecA protein; lane 2, 50mer RAO1 minus RecA protein; lane 3, RAO1 plus RecA protein; lane 4, 30mer RAO2 minus RecA protein; lane 5, RAO2 plus RecA protein; lane G, G-ladder.

![Figure 3.](image)
adducted guanine adduct using the mild, non-denaturing conditions previously described by Povsic et al. (8). Another application of these ODNs might be their use as artificial restriction enzymes. Double-stranded cleavage of DNA would require two sites of crosslinking from the ODN to the target site on DNA, which we have shown with TFOs (7) and RAOs (M. Podyminogin, unpublished results).

The potential of this approach for modification of genes in cells for therapeutic purposes is most interesting. Covalent modification of targeted sites in cells would be expected to inhibit transcription by physical blockage (36). More importantly, permanent alteration of targeted sites in cells would be expected to inhibit transcription for therapeutic purposes is most interesting. Covalent modification of the supF gene by photoreactive TFOs elicits mutations when processed by mammalian cells, although the efficiency reported to date is low (37,38). Modification of both strands of the target site has the potential to produce much higher rates of mutation, as we have found in preliminary studies. This powerful ability to directly change the code of a gene may be applied to correction of genetic defects due to single point mutations or to inactivation of gene function. The latter might find use, for instance, in inactivation of genes responsible for certain autoimmune diseases like type I diabetes (14). Ultimately, the genetic modification principles described herein are preliminary steps to a novel type of gene therapy based on synthetic oligonucleotides that modify the function of endogenous genes.

ACKNOWLEDGEMENTS

We wish to thank Dr Vladimir Gorn for synthesis of oligonucleotides, Deborah Lucas for conjugation reactions, and Drs Igor Kutyavin, David Brown and Joe Hedgpeth for helpful discussions.

REFERENCES