Intemolecular Triplex Formation Distorts the DNA Duplex in the Regulatory Region of Human Papillomavirus Type-11*

(Received for publication, August 19, 1991)

David A. Hartman‡‡, Shu-Ru Kuol, Thomas R. Broker‡, Louise T. Chow‡, and Robert D. Wells‡‡

From the $Department of Biochemistry, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, Alabama 35294, the Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, and the ‡Institute of Biosciences and Technology, Texas A&M University, Houston, Texas 77030

A conformational distortion in the DNA duplex at the regulatory region of human papillomavirus type-11 next to an intermolecular triplex, formed with a synthetic oligonucleotide, was investigated with several chemical probes. The sequence targeted for triplex formation borders on the binding sites for the regulatory proteins encoded by the viral E2 open reading frame. Dimethyl sulfate, diethyl pyrocarbonate, and OsO₄ all react to a greater extent with nucleotides in the duplex that are immediately adjacent to the triplex as compared to other bases throughout the duplex. This hypermodification was observed on both the polypurine and polypyrimidine strands of the duplex DNA. Similar hyperreactivity of bases flanking a triplex was also seen when the contiguous target polypurine tract was effectively extended by mutating interrupting pyrimidines in the human papillomavirus type-11 sequence to purines. We propose that this hyperreactivity is due to a structural distortion caused by the junction between the triplex and the duplex tracts.

Intemolecular triplexes, DNA triple helices formed between an oligonucleotide and duplex DNA, have been extensively studied in the past 5 years (1-3). One type of intermolecular triplex (H-DNA) is formed by the specific interaction of a pyrimidine-rich oligonucleotide with a duplex DNA containing a homopurine sequence. However, intermolecular triplexes with purine-rich third strands also have been reported (4). For the classic Pyr·Pyr·Pyr triplexes, the incoming third strand binds parallel to the purine strand in the major groove of the duplex via Hoogsteen hydrogen bonds. In the case of Pur·Pyr·Pyr triplexes, the orientation of the third strand in relation to the duplex purine strand is still uncertain (4, 5).

A common form of base pairing in a triplex involves a T with Hoogsteen pairing to the A of an A-T pair and a protonated C hydrogen bonded to the G of G-C pairs. Because of the necessity for C protonation, the stability of intermolecular triplexes is increased at acidic pH. However, stable triplexes have been reported at neutral pH (6-8).

The methods used to detect the formation of intermolecular triplexes include gel electrophoresis mobility shift (4), protection of the duplex from enzymatic cleavage (2, 9) and pyrimidine photodimerization (10), and sequence-specific chemical cleavage of the duplex (11). Dimethyl sulfate has been used as a chemical probe to map a triplex to the resolution of base pairs (2). Herein, we have extended the successful use of chemical probes to diethyl pyrocarbonate (DEPC) and OsO₄.

Dimethyl sulfate reacts primarily with the N7 position of DNA and DEPC reacts at the same position on A. The formation of an intermolecular triplex protects this position from reaction with these two chemical probes, a result of the Hoogsteen hydrogen bond formed between the reactive N7 and the incoming pyrimidine. OsO₄ recognizes a structural change in the polypyrimidine tract of the triplex compared to the native duplex. The protection of the polypyrimidine strand in a triplex from photodimerization was ascribed to a conformational change in the triplex target sequence when compared to the native duplex (10).

X-ray analyses of oriented fibers (12, 13) and NMR data (14, 15) indicate that the duplex DNA comprising the triplex adopts an A-like conformation. Therefore, an A-B junction should exist at either end of the triplex region. The junction between A and B DNA (16, 17) duplexes has a bend of 26° and contains 1 bp and two internucleotide linkages. Chemical probes have been used to investigate B-Z DNA junctions (18, 19), demonstrating that bases at these junctions react to greater extent than their nonjunction neighbors.

Human papillomaviruses (HPVs) infect cutaneous and mucosal epithelium causing hyperproliferation. More than 60 distinct types of HPV have been identified (20). HPV-11 is the etiologic agent for genital warts and laryngeal papillomatosis. The sequence targeted for the formation of the intermolecular triplexes in this study is located at the 3’ end of the upstream regulatory region (URR) of HPV-11, preceding the E6 promoter and E6 gene (Fig. 1, top). The sense strand contains a polypurine stretch 22 nucleotides long (genome position 21-42) interrupted by CC at position 36-37. This particular sequence occurs only once in the HPV-11 genome and it overlaps a putative Sp1 transcription factor-binding site (21), as well as the first of a pair of binding sites (ACCNGT) for the several regulatory proteins encoded by the viral E2 open reading frame. Together these sites are critical for the regulation of the E6 promoter (Refs. 22 and 23, and references therein). In addition, the target sequence is within the segment of DNA defined to be the binding site for the viral E1 replication protein and the origin of DNA
replication (Refs. 24–26). Therefore, it is of considerable interest to investigate the possible influence of triplex formation on viral transcription and DNA replication.

We report the initial investigation of the properties of triplex formed in wild-type and mutated HPV target sequences. Hypermodification of both the pyrimidine-rich and purine-rich strands of DNA duplex bases of HPV-11 were observed. This reactivity may be due to A-B DNA conformational junctions immediately adjacent to the intermolecular triplex.

**MATERIALS AND METHODS**

**General—Oligonucleotides** were prepared on an Applied Biosystems 380A synthesizer and purified by high performance liquid chromatography. Further purification, when necessary, was achieved by denaturing gel electrophoresis. DNA restriction and DNA modification enzymes were obtained from New England Biolabs and Boehringer Mannheim, respectively. Other reagents were obtained from Aldrich or Sigma.

**Plasmids—HPV-11** has a double-stranded circular genome 7933 bp long. Plasmid pUR 23-3 Δ5-3 contains nucleotides 7808–7983/1–99 from the 3' end of the HPV-11 upstream regulatory region (23). The contiguous purine sequence for triplex formation from nucleotides 1–73 are shown at the top of Fig. 1. The contiguous pyrimidine tract was extended to its 3' side by introducing site-directed mutations by a modification of the method of Zoller and Smith (27) (Fig. 1, bottom). Briefly, the neomycin (and kanamycin) resistance gene (encoding aminoglycoside 3'-phosphotransferase from the bacterial transposon Tn5) containing a frameshift mutation at the NcoI site and the entire HPV-11 URR spanning nucleotides 7072–99 were cloned sequentially into the polylinker region of the Stratagene vector pBS-SK*.

The single-stranded M13 phage DNA template was then prepared after helper phage infection. Two oligonucleotides were used in conjunction to synthesize the complementary strand to the M13 template using T4 DNA polymerase. One oligonucleotide restored the mutated neomycin resistance gene to wild-type, while the other introduced the desired mutations in the HPV-11 URR. The DNA was first transformed into repair-deficient Escherichia coli strain BMH 71–18 mat S, under kanamycin selection. Plasmid DNA recovered from resistant colonies was then transformed into E. coli DH5α, also under kanamycin selection. Mutant plasmids recovered from the DH5α transformants were confirmed by DNA sequencing using the Sanger diodeoxynucleotide chain termination method (28).

**RESULTS**

**DMS Modification of HPV-11 URR DNA Fragment—The** 222-bp long PstI-HindIII restriction fragment containing the triplex target sequence was excised from the plasmids described above. To label the polypurine-containing strand, the recessed end of the restriction fragment was filled with 32P-nucleotides using the Klenow fragment of a DNA polymerase, and the 5'-flanking base for oligo a or the 3'-flanking base for oligo b is G. Both of these flanking bases were modified to a greater extent than other Gs (Fig. 2). Significant hyperreactivity was observed at an oligonucleotide concentration that resulted in good chemical protection. The level of modification is greater at the 3' end of the DNA fragment (Fig. 2, lanes 5–8) when compared to the 5' end (Fig. 2, lanes 1–4). This hyperreactivity is probably due to the increased exposure of the N7 position of G to reaction with dimethyl sulfate, because of a conformational change at the junction between the triplex and the duplex.

Similar investigations were conducted with oligonucleotides 9 and 13 nucleotides in length that were hybridized to the HPV-11 duplex from nucleotide 21 to 29 and from nucleotide 21 to 33, respectively. In both cases, hypermodification was observed for the Gs flanking the triplex at the 3' end. Also, the hypermodification of the nearest neighbor on the polypurine strand at the 3' end of an intramolecular triplex has been reported (30), but was only found for some triplexes.

**DEPC Hypermodifies Adenines 3' to an Intramolecular Triplex—The** DEPC was employed to determine if the reactivity observed above was specific for G. Although DEPC will react with all purines, it primarily reacts with the N7 position of As and has been employed to probe intramolecular triplexes (30) and cruciforms (31). Therefore, DEPC complements dimethyl sulfate for the mapping of intramolecular triplexes and can also be used to probe A-rich areas of triplex targets such as those present in this sequence.

**Triplexes were formed between the wild-type HPV-11 target sequence and either an 11-mer (Fig. 1, oligo c) or a 20-mer (Fig. 1, oligo d) (Fig. 3). Oligomer c forms a triplex which extends into the last two of the six As at the 5' end of the polypurine tract. The other oligonucleotide spans the first 20 bases of the purine run ending before the last two As and includes two T·C·G mismatches at positions 36 and 37. The
Hypermodification of a DNA Duplex Adjacent to a Triplex

**Fig. 1.** Oligonucleotides and wild-type and mutated HPV-11 URR sequence targeted for triplex formation. The relevant portion of the 222-bp fragment from the wild-type HPV-11 URR is shown at the top with the target polypurine sequence underlined. The oligonucleotides (a–g), shown in the middle, are written from 5' to 3' and pair with the target sequence by Hoogsteen hydrogen bonding. The binding sites for the HPV E2 proteins, the host TFIID, as well as a binding site for host transcription factor Spl, are overlined. Site-directed mutations of the URR sequences, I–III, are shown at the bottom. All mutated bases are designated by lower case letters.

---

**Fig. 2.** Dimethyl sulfate probing of the polypurine containing strand. The 222-bp wild-type HPV-11 DNA fragment (Fig. 1, top) (3 nM) was labeled with the Klenow fragment of the E. coli DNA polymerase I. The dimethyl sulfate probing reactions were performed as described under “Materials and Methods.” The final concentrations of the oligonucleotides added are shown above each lane. Oligonucleotide a was used in lanes 1–4 and oligonucleotide b in lanes 5–8. The first two lanes at the left are the G and G + A Maxam-Gilbert sequencing lanes, respectively. Hypermodified bases are marked by arrows to the side of the lanes. The concentrations of oligonucleotides and DNA are reported before the addition of dimethyl sulfate.

---

**Fig. 3.** DEPC probing of the purine-rich strand. The DEPC probing reactions were performed as described under “Materials and Methods.” The final concentrations of the oligonucleotides added to the 222-bp wild-type HPV DNA fragment (Fig. 1, top) (3 nM) labeled with the Klenow fragment of DNA polymerase I are shown above each lane. Oligonucleotide c was used in lanes 1–4 and oligonucleotide d in lanes 5–8. The first two lanes at the left are the G and G + A Maxam-Gilbert sequencing lanes, respectively. The hyperreactive base (lanes 5–8) is marked by an arrow. The concentrations of oligonucleotides and DNA are reported before the addition of DEPC.

Oligonucleotides protected the duplex purines from reacting with DEPC.

The polypurine target was protected equally well from reaction with DEPC or dimethyl sulfate at the same concentration of oligonucleotide. These data suggest that DEPC is just as sensitive as dimethyl sulfate in determining the amount of oligonucleotide needed to achieve significant triplex formation.

Once again, increased modification of the A flanking the 3’ end of the triplex was observed (Fig. 3, lanes 5–8). However, the A flanking the 5’ end of the triplex formed by oligo c did not react to a greater extent than the neighboring As (Fig. 3, lanes 1–4). This hyperreactivity of the 3’-flanking A residue,
but not the 5'-flanking A, may be due to an increased accessibility of the DEPC to the A at the 3' end. The electrophilic carbonyl of DEPC is more sterically hindered than the methyl group of dimethyl sulfate. The conformation at the 3' junction might be such that the steric hindrance is less than at other positions outside the triplex.

**Comparative Ability of Oligonucleotides Containing Two Mismatches to Form Triplexes—Naturally occurring sequences rarely contain extended runs of purines. To generalize triplex technology, it would be desirable to determine the consequences of interrupting pyrimidines on the stabilities of triplexes**

For these experiments, but the polypyrimidine strand was labeled at the 5' end with T4 polynucleotide kinase. The 32P-labeled DNA was incubated with increasing amounts of oligomer c and probed with OsO4. Within the polypyrimidine target, OsO4 reacts only with the three Ts at nucleotides 28, 31, and 35.

Two interesting phenomena were observed as the oligonucleotide concentration was increased (Fig. 5). The two reactive Ts, in the middle of the polypyrimidine tract, were protected from modification with OsO4, (Fig. 5, bottom). This protection occurred at the same concentration as observed for the DEPC probing of the polypurine strand. However, the boundary T at nucleotide 35 was not protected. The presence of the triplex

---

**Fig. 5. OsO4 probing of the pyrimidine-rich strand.** The wild-type HPV DNA fragment (222 bp, 3 nM), labeled on the polypyrimidine strand with T4 polynucleotide kinase, was incubated with oligomer c at 30, 300, and 3000 nM. The log of the fold excess of the oligomer concentrations is shown on the abscissa. After 30 min, the samples were probed with OsO4 as described. The autoradiographs were scanned using a Bio-Rad 620 densitometer equipped with a Hewlett-Packard 3392A recording integrator. The upper panel represents the intensity of the bands for flanking nucleotides 21-24. Values shown are relative to the intensity of the bands for 3000 nM of oligo c at the 100% point. In the lower panel, all values are shown relative to the densitometric scans of nucleotides 28 and 31 without oligo c which was assigned a value of 100% and the 3000 nM point a value of 0%.
or the adjacent duplex-triplex junction did not appear to affect the reactivity of this T. Also the presence of the third strand did not protect the other Ts in the fragment outside of the triplex from reaction with OsO₄.

In addition, the four Ts immediately 5' to the end of the triplex also were modified to a greater extent, as the oligonucleotide concentration was increased (Fig. 5, top). Thus, OsO₄ appeared to detect a conformational change in the polypuridine strand upon the formation of a triplex. We attribute this hypermodification to a change from the conformation adopted by the d(A)₄-d(T)₄ tract to an alternate conformation induced by the formation of an intermolecular triplex. After triplex formation, these 6 Ts no longer act in concert (see “Discussion”).

No change in the OsO₄ modification pattern of the bases which flank a triplex formed by oligomers b and d on the polypurine or polypurine strand was observed. However, when the polypurine strand of the triplex was probed, the reactive Ts were protected from chemical modification.³

**DEPC and Dimethyl Sulfate Probing of Purines Adjacent to the Triplex on the Polypurine Strand**—The purines on the polypurine strand immediately next to the triplexes formed by the four oligonucleotides (a-d), and the target wild-type DNA were subjected to chemical probing. For oligomers c and a, Gs are 5' of the triplex and thus dimethyl sulfate was employed. In the cases of b and d, As are the 3' adjacent bases and DEPC was used as a probe.

For oligomer a, no protection of the Gs at nucleotides 36 and 37 from chemical modification was observed (Fig. 6, lane 6). Dimethyl sulfate does not appear, at least in this case, to detect any structural change in the polypurine strand upon triplex formation. Hypermodification was seen only for the second G 3' from the triplex formed by oligomer c (Fig. 6, lane 5). However, bases which flank the triplexes formed by oligonucleotides a, b, and d were modified to the same extent as all other duplex purines (lanes 6, 2 and 3, respectively).

**Chemical Probing of Intermolecular Triples Formed with Mutants of the E2 Responsive Sequence 3**—To ascertain whether the observed modification patterns were in some cases caused by the presence of the two Cs in the polypurine tract, plasmids containing mutations in the HPV-11 URR were prepared and the appropriate fragments utilized. These mutations are shown in Fig. 1 with the substituted bases shown in lower case. In each instance, except for nucleotide 47, A or G replaced 4, 5, or 6 pyrimidines within the tract from nucleotide 21 to 50 to give polypurine runs of 30 bases with 2, 1, and 0 interruptions for mutations I–III, respectively.

Extension of the polypurine tract led to a significant increase in the stability of the triplexes formed with oligonucleotide a. All mutations led to an equivalent protection from dimethyl sulfate modification at a 10-fold lower concentration than needed for oligomer a and the wild-type sequence.³ While two adjacent interruptions of the polypurine tract reduce the stability of the formed triplex, one mismatch, at least in this instance, does not destabilize the triplex. In addition, similar hypermodification patterns were observed for the mutants. These results suggest that the two Cs do not play a significant role in the phenomena observed here.

Further studies of triplexes with these mutated HPV-11 sequences were conducted with two 17-mers (not shown) which covered nucleotides 28–44 and 26–42. For the first oligomer, protection of the Gs on the polypurine strand from dimethyl sulfate modification and the hypermodification of the Gs at nucleotides 27 and 45 was observed. For the second 17-mer, only normal modification by DEPC was observed for the A at nucleotide 25 on the polypurine strand which flanks the 5' end of the triplex as described above.

In summary, studies with the mutated duplex target sequences revealed that the presence of two pyrimidines in the polypurine tract significantly destabilized an intermolecular triplex. However, only one C in the polypurine run did not lead to an appreciable change in the concentration of oligonucleotide needed to form a triplex.

**DISCUSSION**

The formation of an intermolecular triplex increases the accessibility of duplex nucleotides immediately adjacent to the triplex to chemical probes. All studies were conducted using a wild-type or site-mutated 222-bp fragment (Fig. 1) located at the 3' end of the human papillomavirus upstream regulatory region which includes an imperfect run of 22 purines interrupted by two adjacent Cs. Fig. 7 summarizes the results with oligonucleotides a–d shown hybridized to their complementary sequences in parts a–d, respectively.

In most cases, the Gs adjacent to or one removed (part c) from the triplex are modified to a greater extent by dimethyl sulfate than the other Gs throughout the duplex. However, the G in line b was more reactive with dimethyl sulfate than the Gs in parts a and c. These results suggest that the N7 position of the Gs on the polypurine strand next to the 3' end of the triplex are more exposed than in other Gs. In two other instances, parts a and c, the Gs next to the triplex on the polypurine strand did not exhibit increased reactivity to dimethyl sulfate.

With DEPC as a probe, only the A in part d which flanks the triplex at the 3' end of the polypurine strand exhibited increased reactivity. It is not surprising that DEPC did not react to a greater extent with As that border the triplex when
Hypermodification of a DNA Duplex Adjacent to a Triplex

5493

The relevant region of the wild-type HPV target sequence is shown in Figure 7. The hypermodification patterns are represented as follows. Hyperreactive bases are in bold face. Hyperreactive bases are in large bold-face type. Dimethyl sulfate was used to modify Gs; DEPC, As; and OsO₄, Ts.

compared with other As throughout the rest of the duplex. DEPC is larger than dimethyl sulfate; the conformational change that evokes the accessibility may not be sufficiently large to observe via DEPC probing. In addition, the structural distortion at the 5' end of the triplex may not be as great as that at the 3' end.

Cs within the region of interest on either strand did not react to a significant extent with OsO₄, either in the presence or absence of an intermolecular triplex. Ts are more reactive with OsO₄ and did exhibit a change in modification both within the triplex and flanking the triplex with one exception (Fig. 5). However, only the four Ts on the polypurine strand in part c 3' to the triplex were hypermodified by OsO₄, probably due to the interruption of the six Ts by the triplex.

Interestingly, homopolymeric d(A)n - d(T)n tracts (n ≥ 4) are resistant to modification by OsO₄ (36) and thermal melting, and NMR studies have shown that these sequences are conformationally very stable (37-39). This stability has been proposed to lead to decreased reactivity with OsO₄. When the stretch of A·T pairs is reduced to 3 or less, the Ts will react with OsO₄ (36).

Thus, we propose that this tract of 6 Ts can be considered as three separate regions after triplex formation. Starting from the 3' end of the tract, the first 3 Ts are in an unstacked B-DNA conformation, as demonstrated by reaction with OsO₄. Because the next T is at the junction between the B-DNA duplex and the A'-DNA duplex of the triplex, it is also modified by OsO₄. Thus, the presence of the junction prevented the d(A)n - d(T)n tract from adopting the more stable conformation which is resistant to OsO₄ modification. The last two bases are, of course, under the triplex and, because they are on the other side of the junction, would not contribute base stacking cooperativity to the external poly(A) tract.

We have demonstrated, as part of these studies, that OsO₄ can be used as a probe for intermolecular triplex formation. The normally reactive Ts on the polypyrnidimide strand were protected from reaction with OsO₄ as the oligonucleotide concentration was increased. This effect is similar to that previously observed for the photofootprinting of intermolecular triplexes.

Intermolecular triplex formation prevents the formation of [6-4]-pyrimidine photodimers (10). This effect was explained by a reduction in the range of motion of pyrimidines upon triplex formation. Photoinduced DNA damage is decreased after the transition from B- to A-DNA. The decrease in OsO₄ reactivity observed here is also probably due to the transition from B-DNA to A'-DNA as the triplex is formed. In addition, the presence of the third strand is expected to reduce the accessibility of the reagent to the 5,6-double bond of T. The use of OsO₄, however, allowed, in this case, for the detection of triplexes even on the polypyrnidimide strand.

The use of DEPC to detect the formation of intermolecular triplexes has not been previously described. The protection from reaction with DEPC confirmed the hypothesis of Sullivan and Lebowitz (36) that the N7 position of A is exposed to a greater extent in polymeric A·T tracts (n ≥ 4) when compared to other duplex As. The Ts in the third strand bind to the duplex A via Hoogsteen hydrogen bonds involving the N7 position, protecting it from reaction with DEPC. The use of DEPC as a chemical probe therefore provides a method to map triplexes when long (dA)n - (dT)n tracts are present within the polypyrnidimide sequence.

We propose that the increased accessibilities to chemical probes are due to a bend in the DNA duplex at the junction between the duplex and triplex. This bend results in the increased exposure of the N7 position of purines, as exhibited by the increased modification by dimethyl sulfate and DEPC. The original duplex strands contributing to the triplex are in an A'-DNA conformation (12-15). The junction between A-DNA and B-DNA is bent by 26°, whereas an angle of 19° was determined for an A'-B DNA junction (17).

The observed greater modification of bases immediately 3' to the end of the triplex is similar to that reported for the bending of (dA)n - (dT)n tracts. A larger angle was observed at the 3' end of the A tract as compared to the 5' end (40). This difference could explain the increase in the modification of the bases at that end. The widening of the major groove by this change in the angle would lead to a more exposed N7 position.

The lack of hypermodification observed on the polypyrnidimide strand for the bases flanking the triplex at the 5' end is also the result of this bend. The N7 position of these Gs were probably sterically hindered by the neighboring bases from modification due to the junction-induced bend. The bend is such that the next base is hypermodified. In addition, the lack of protection of the T at nucleotide 35 on the polypyrnidimide strand also would suggest a conformational change between triplex and duplex.

The possibility of forming an intermolecular triplex and its accompanying change in the overall duplex conformation has implications for possible experimental and therapeutic interference of the regulation of HPV-11 gene expression. Within the URR of HPV-11 there are four binding sites for the E2 proteins (41), two of which are found at nucleotides 35-46 and 50-61. The full-length E2 protein can transactivate or under certain conditions repress (23) the E6 promoter located near the tandem E2-binding sites. The carboxyl portion of the protein, E2-C, and an E1M·E2C fusion protein can also bind to the responsive sequences and repress both E2-dependent and independent transcription (23, 42).

An intermolecular triplex formed in this region may mediate binding of Sp1, E1, and E2 to their respective binding sites by either direct competition for binding to the recognition sequence or by changing the conformation of the DNA.
to a structure which is not recognized by the proteins. These modulations might lead to a strategy for the regulation of viral DNA replication and transcription.

REFERENCES


34. Deleted in proof.


