Structure of the DNA Interstrand Cross-Link of 4,5′,8-Trimethylpsoralen†

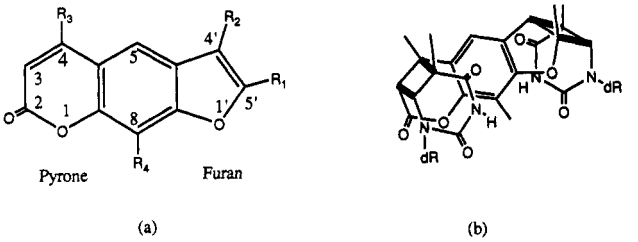
Kandallu R. Kumaresan,‡ Muthukumar Ramaswamy, and Anthony T. Yeung*

Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received October 10, 1991; Revised Manuscript Received May 1, 1992

ABSTRACT: 4,5′,8-Trimethylpsoralen (TMP) cross-links a 5′ TpA or a 5′ ApT site by photoreacting with one thymine moiety in each DNA strand. We are interested in whether psoralen interstrand cross-links all share one structure or whether there are significant differences. In this paper, we employed a rapid method for probing the structure of the cross-link by making a series of TMP cross-linked duplexes containing specific base-pair mismatches. The relative stability provided by a base pair can be correlated with neighboring base pairs by comparing the extents of gel retardation when base-pair mismatches happen in each position. From our studies, we infer that with respect to the furan-side strand, the 5′ T·A base pair of the two T·A base pairs in the TpA site is not hydrogen bonded. Immediately on each side of the cross-linked TpA site is a highly stabilized base pair. Next, a region of decreased stability occurs in each arm of a cross-linked duplex and these base pairs of least stability are located farther away from the cross-linked thymines as the lengths of the arms of the cross-linked helix increase. Finally, even in 7 M urea at 49 °C the cross-linked helix is hydrogen bonded at both ends of a duplex of 22 base pairs. We propose that the structures of interstrand cross-links in DNA vary appreciably with the DNA sequence, the length of the DNA duplex, and the structures of the DNA cross-linking agents.

Psoralens belong to a class of compounds called furocoumarins (Figure 1). They are present in a variety of vegetables and they are used with long-wave UV as an effective treatment for psoriasis. However, both psoralen–DNA monoadducts and interstrand cross-links are mutagenic (Bridges et al., 1979; Kirkland et al., 1983; Ashwood-Smith et al., 1983). Fortunately, cells can rid themselves of these DNA lesions by using a nucleotide excision DNA repair system. The UvrABC endonuclease of *Escherichia coli* is a good model for nucleotide excision repair [for reviews, see Grossman and Yeung (1990), van Houten (1990), and Sancar and Sancar (1988)]. This enzyme incises the lesion-containing DNA twice on the same DNA strand at each lesion such that an oligonucleotide about 13 bases long containing the lesion can be displaced from the DNA helix (Sancar & Rupp, 1983; Yeung et al., 1983). Although both DNA strands are modified by the psoralen molecule in the case of the psoralen interstrand DNA cross-link, the UvrABC endonuclease can only initiate repair on one of the two DNA strands. That is, the incision can be on either the furan-side or the pyrone-side strand (Jones & Yeung, 1988). Base composition near the cross-links of 4,5′,8-trimethylpsoralen (TMP)[1] seems to determine which strand is to be incised (Jones & Yeung, 1990). It follows that this choice of strand for repair by the UvrABC endonuclease suggests that the structures of psoralen cross-links in different DNA sequences are not identical with respect to lesion recognition by a repair enzyme. The work in this paper begins to develop this theme of multiple cross-link structures by describing an unexpected complexity in the stabilities of the base pairs in the DNA helix on the two sides of the psoralen cross-link.



(a) (b)

FIGURE 1: Structure of (a) psoralen derivatives and (b) the cis–syn–cis interstrand cross-link with two thymines. 4,5′,8-Trimethylpsoralen (TMP): $R_1 = R_3 = R_4 = CH_3$, $R_2 = H$. 4′-Hydroxymethyl-4,5′,8-trimethylpsoralen (HMT): $R_1 = R_3 = R_4 = CH_3$, $R_2 = CH_2OH$. 4′-Aminomethyl-4,5′,8-trimethylpsoralen (AMT): $R_1 = R_3 = R_4 = CH_3$, $R_2 = CH_2NH_2$. 8-Methoxypsoralen (8-MOP): $R_1 = R_2 = R_3 = H$, $R_4 = OCH_3$. dR = deoxyribose.

A good deal is known about the photochemistry of psoralens. The chemical reactivity of psoralen with DNA is highly specific and controllable (Song & Tapley, 1979; Cimino et al., 1985). The pyrimidine bases of DNA involved in cross-link formation are thymine moieties >99% of the time (Tessman et al., 1985; Yeung et al., 1987). The formation of a thymine–psoralen–thymine cross-link in a DNA duplex occurs in three steps. Psoralen first intercalates into the DNA helix noncovalently. With the absorption of a quantum of light, either the 3,4-pyrone double bond or the 4′,5′ furan double bond of psoralen can photoreact with the 5,6 double bond of a pyrimidine to form a pyrimidine–psoralen monoadduct in cis–syn conformation (Kanne et al., 1982b; Peckler et al., 1982). Reaction of the 3,4-pyrone double bond destroys the coumarin nucleus. However, if the 4′,5′ furan double bond is the first to react, the coumarin nucleus is intact and therefore can absorb light at 365 nm for its 3,4-pyrone double bond to react with the 5,6 double bond of another pyrimidine to form a cross-link with cis–syn–cis stereochemistry (Kanne et al., 1982a; Parsons, 1980; Figure 1). The 8-methoxy group of

† These studies were supported, in part, by a grant from the National Science Foundation (DMB88-02091) to A.Y., an institutional instrumentation grant (DIR-8812108) from the National Science Foundation, institutional grants from the National Institute of Health to the Fox Chase Cancer Center (CA06927, RR05539), an appropriation from the Commonwealth of Pennsylvania, and a grant from the Glenmede Trust to the FCCC.

* Author to whom correspondence should be addressed.

‡ Present address: Biology Department, Brookhaven National Laboratory, Upton, Long Island, NY 11973.

[1] Abbreviations: TMP, 4,5′,8-trimethylpsoralen; AMT, 4′-aminomethyl-4,5′,8-trimethylpsoralen; 8-MOP, 8-methoxypsoralen; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; TBE, 50 mM Tris–borate and 1 mM EDTA, pH 8.3; oligo, oligonucleotide; bp, base pairs.

0006-2960/92/0431-6774$03.00/0

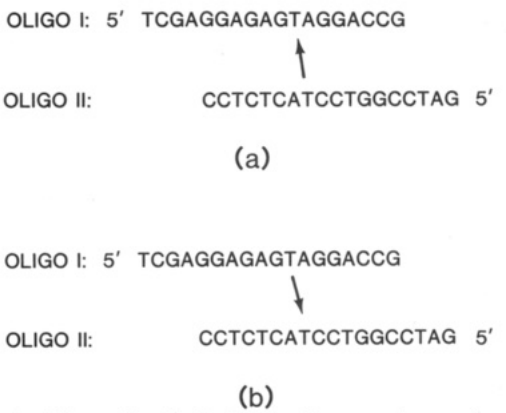


FIGURE 2: Oligonucleotide duplex used in a previous study. Oligo I and oligo II are complementary except for the two 4-base sticky ends. There is a single 5′ TpA site in the middle of the duplex and TMP can intercalate into this site and react in either of two orientations to form two cross-linked isomers. The arrows indicate the orientation of the psoralen molecules from the furan ring to the pyrone ring. Panel a is the cross-link of lower electrophoretic mobility (upper cross-link). Panel b is the cross-link of the higher electrophoretic mobility (lower cross-link). The oligonucleotide duplexes used in Figures 4–9 are the same as this duplex except that one or both of the sticky ends are filled in to produce duplexes up to 22 base pairs long.

8-methoxypsoralen (8-MOP) and the 8-methyl group of 4,5′,8-trimethylpsoralen (TMP) point into the minor groove of the B-DNA helix. Photoreaction of 8-MOP, but not TMP, with DNA produces significant amounts of pyrone-side monoadducts that will not react further to form cross-links.

Some of the properties of a psoralen cross-link are known. Interstrand cross-linking of two DNA strands prevents the strands from separating and increases the apparent melting temperature of the DNA duplex (Shi & Hearst, 1986). A structure has been proposed for a psoralen cross-link formed by 4′-aminomethyl-4,5′,8-trimethylpsoralen (AMT) in the palindromic sequence 5′ GGGTACCC in solution (Tomic et al., 1987). This structure is based on cross-link modeling starting with the crystallographic data on the 8-MOP-thymine monoadduct (Peckler et al., 1982) and nuclear magnetic resonance (Tomic et al., 1987). It was concluded that cross-linking by AMT produces a 53° bend into the major groove at the site of psoralen addition and a 56° unwinding that spans the 8-base-pair duplex. With respect to the furan-side strand, the helix on the 3′ side was more unstable than the helix on the 5′ side. A striking feature of the AMT cross-link was a lack of hydrogen bonding of the 3′ T·A base pair in the TpA site, with respect to the furan-side strand.

The structure of psoralen-cross-linked helix may vary with the DNA sequence. A molecule of TMP can intercalate at a 5′ TpA site in two orientations and photoreact to form two different cross-linked duplexes (Figure 2). With respect to the orientation of one strand of this cross-linked sequence, the two cross-linked duplexes have been called orientational isomers. Another way of looking at these cross-linked duplexes is to hold the psoralen molecule constant in orientation. Then, with respect to the psoralen molecule, these cross-linked duplexes are just two TMP cross-links in two different DNA sequences. These two cross-linked duplexes migrate differently in a DNA sequencing gel (Yeung et al., 1988; Figure 3), again suggesting that TMP cross-links in different DNA sequences differ in conformation. The nature of the difference in the conformation has not been elucidated. However, from the observation that increasing the gel temperature by 16 °C decreased the mobility of the faster moving band (Yeung et al., 1988), it can be inferred that the two cross-links differ in

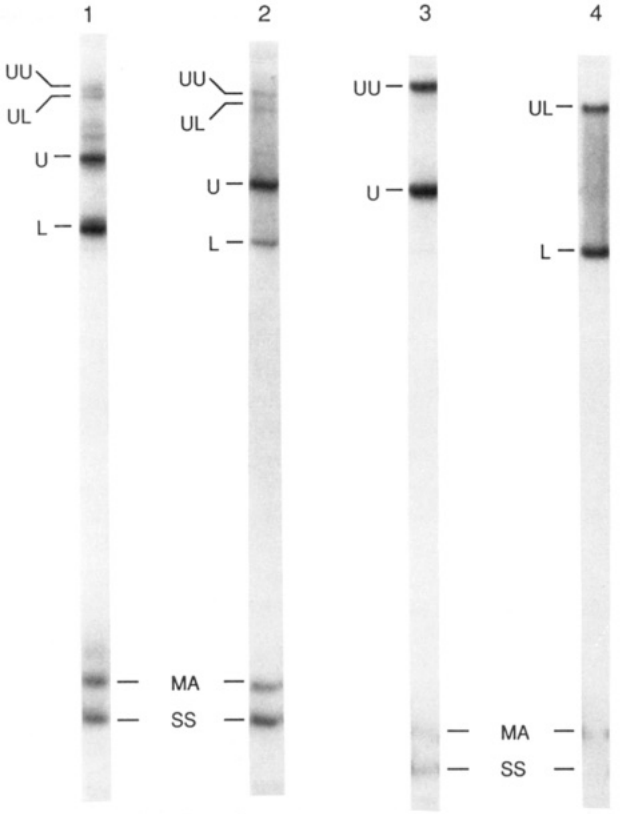


FIGURE 3: Resolution of psoralen cross-links in DNA sequencing gels. Oligo I of Figure 2 was 5′ labeled with ^{32}P and hybridized to oligo II. The duplex was reacted with TMP (lane 1) or 8-MOP (lanes 2–4). Lane 1 was photoreacted for 5 min and lane 2 was for 1 h. The upper cross-link (U) and lower cross-link (L) of lane 2 was purified from the gel and analyzed by electrophoresis again in another gel in lanes 3 and 4, respectively. The U-band split and produced UU (upper-upper cross-link band). The L-band split to produce UL (upper-lower cross-link band). UU and UL were also visible in smaller amounts in lanes 1 and 2. The differences in relative intensities of the UU and U bands and the UL and L bands in lanes 3 and 4 versus lane 2 may be due to small variations in the temperature of the gel reservoir during sample application, leading to more melting in lanes 3 and 4. SS = single-stranded; MA = monoadduct oligonucleotide.

the stability of some base pairs. Yet, the whereabouts of those base pairs and why the mobility decreases upon their melting is unclear. This phenomenon is examined in more detail in the current work.

It was observed that when a cross-linked DNA molecule was isolated from a DNA sequencing gel and then analyzed again by another DNA sequencing gel, it can split and produce a band of even slower mobility (Figure 3, lanes 3 and 4). Both of these bands came from the same psoralen orientational isomer. Although the enhanced retardation is consistent with the idea of two levels of melting of the DNA helix of a cross-linked duplex, the exact nature of the change has not been proven. It is a purpose of the current study to test whether stepwise melting of the DNA duplex in the structure of a TMP cross-link can account for that amount of decrease in mobility. Because the cross-linked duplex is a relatively small molecule, it is unexpected that two species of the same cross-link isomer can simultaneously form a band each during electrophoresis in a denaturing DNA sequencing gel. In this paper, we examined one of the possible explanations, that the base pairs near a psoralen cross-link may be unequal in stability and that the presence of an unusually stable base pair may hinder the melting or the renaturation of the helix. We will probe the stability of the base pairs within a cross-linked DNA duplex. Our method is to force the duplexes to melt by systematically introducing different numbers of base-pair

mismatches into various locations of the cross-linked duplex and then observe the gel mobilities of the mismatched duplexes in denaturing polyacrylamide gels.

It is not known whether cross-links formed by psoralen derivatives of different chemical structures produce the same distortion in the DNA helix. For example, the unpaired T·A base pair in the cross-linked 5′ TpA site of the AMT cross-link may be a major feature of a psoralen cross-link. We wish to test whether this unpaired base pair is located at the same location for TMP. The electrophoretic analysis of mismatched cross-links described here provides a rapid method of obtaining some information to facilitate decisions in future spectroscopic and high-resolution structural studies.

Psoralen interstrand cross-linking has been useful for locating double-stranded regions in nucleic acids [for a review, see Cimino et al. (1985)]. A secondary goal of the current study is to determine how much sequence homology is necessary for the formation of psoralen interstrand cross-links.

We report here that perfect sequence homology is not necessary for the formation of psoralen cross-links. Our study illustrates an unexpected complexity in the base-pair stability of a psoralen-cross-linked helix.

EXPERIMENTAL PROCEDURES

Preparation of Oligonucleotides. Oligonucleotides were synthesized by Applied Biosystems DNA synthesizers, using the economical FC3 program we previously described (Yeung & Miller, 1991). The oligonucleotides were purified by resolving 5 A_{260} units of each oligonucleotide in a 0.5 mm × 30 mm × 40 cm, 20% acrylamide gel, 37.5:1 acrylamide: bis(acrylamide), 7 M urea, and 1× TBE. All gels were run between 43 and 49 °C at 2000 V for 3.5–4 h. The tracking dye was 90% formamide, 0.1% xylene cyanol, and 0.25% bromophenol blue. The full-length oligonucleotides were visualized with the least amount of UV shadowing against the highly fluorescent background of a Lightning-Plus enhancer screen and then extracted from the gel slice as previously described (Maxam & Gilbert, 1980). ^{32}P-Labeling of oligonucleotides was at the 5′ termini, using γ-labeled [^{32}P]ATP at 6000 Ci/mmol and T4 polynucleotide kinase as described (Maxam & Gilbert, 1980) except that unlabeled ATP was used to chase the phosphorylation to completion. The labeled oligonucleotides were resolved again on 20% acrylamide denaturing DNA sequencing gels and purified for the TMP photoreactions.

Preparation of Oligonucleotide Containing a Specific TMP Furan-Side Monoadduct. A typical photoreaction consists of about 10 μg of each of the two complementary oligonucleotides, annealed to form a duplex by heating to 85 °C and slowly cooling to room temperature. The photoreaction consists of the oligonucleotide duplex, 5 μg/mL TMP in 100 μL of 5 mM Tris-HCl, 0.2 mM EDTA, and 50 mM NaCl, pH 7.6, irradiated at 25 °C for the indicated times by a 500-W filtered light source, at 130 J m^{-2} s^{-1} at the sample. Details of these photoreaction conditions were previously described (Yeung et al., 1988). For the experiments in Figures 7–10, we used a duplex of structure 10 of Figure 5 to form the monoadduct structure 11 of Figure 5 (same as structure 1 of Figure 6). Typically, 3 min of irradiation was used for monoadduct formation. After the photoreaction, the TMP was removed by two extractions with chloroform followed by ethanol precipitation of the DNA. The monoadduct-containing oligonucleotide was resolved from the nonadducted oligonucleotide on a DNA sequencing gel and purified as described above.

Formation of Oligonucleotide Duplexes, Containing Mismatched Bases, Cross-Linked with the Psoralen Molecule in a Single Orientation between the Two DNA Strands. The bottom strand containing a furan-side monoadduct (structure 11, Figure 5) was hybridized to each complementary top strand, containing various amounts of sequence changes, to form the duplexes described in Figures 5 and 6. Upon irradiation of these duplexes by 365-nm UV for 10 min, interstrand cross-links were formed, all with the furan ring attached to the bottom DNA strand. The base-pair mismatches used to disrupt normal hydrogen bonding at each base pair included A·C, G·T, and C·C mismatches. We have not included purine–purine mismatches in this work because they are likely to produce greater steric hindrance and may complicate data interpretation.

Formation of Oligonucleotide Duplexes, Containing Mismatched Bases, Cross-Linked with the Psoralen Molecule in Both Orientations between the Two DNA Strands. To obtain both psoralen orientations for each structure in Figures 5 and 6, we hybridized pairs of oligonucleotides containing various amounts of base-pair mismatch, as described in Figures 5 and 6, and photoreacted them with TMP randomly under standard conditions as described above. After the photoreaction, the DNA samples were dried in a Speed-Vac, dissolved in tracking dye, and analyzed by DNA sequencing gels.

RESULTS AND DISCUSSION

Perfect DNA Sequence Complementation Is Not Required for Psoralen Interstrand Cross-Linking. The approach used in this study relies on the possibility of forming TMP cross-links between oligonucleotides containing mismatched base pairs. The results in Figures 7–10 illustrate that duplexes containing as much as 27% of the sequence as mismatched bases, including some with a mismatch within one of the two base pairs involved in the cross-linking site, can form interstrand cross-links with good efficiency. However, for the unfavorable cross-linking sites (such as those in Figure 7, lanes 12–15, and Figure 8, lane 16) containing mismatches at or next to the psoralen molecule, the oligonucleotide with the preformed covalently bound furan-side monoadduct produced interstrand cross-links more efficiently (Figures 7 and 8) than in the random reactions that require two steps to form the cross-links (Figures 9 and 10). Our results highlight the need for caution in the interpretation of experiments in which the formation of psoralen interstrand cross-links is used to indicate DNA sequence homology at target sites.

The Complementary Bases at the Termini of Both Cross-Link Isomers Are Hydrogen Bonded. Psoralen can intercalate in a 5′ TpA site in two orientations and photoreact to form two cross-links with opposite orientations of the psoralen molecule with respect to the DNA sequence. These two cross-links are called psoralen orientational isomers. We previously showed that the psoralen orientational isomers in an oligonucleotide duplex resolved in a DNA sequencing gel into two major bands of significantly different mobilities (Yeung et al., 1988) and therefore reflecting conformational differences between these cross-links. Some possible conformations are described in Figure 4. Intuitively, one may think that in a DNA sequencing gel, the 7 M urea at 49 °C would melt the DNA base-pairing of one or both arms of the cross-linked duplex (Figure 4B,C). Then the single-stranded arms would unfold and may retard the electrophoretic mobilities of the cross-linked duplex. However, as we show below, this description is incorrect. The complementary bases at the termini of each cross-link band are essentially hydrogen bonded

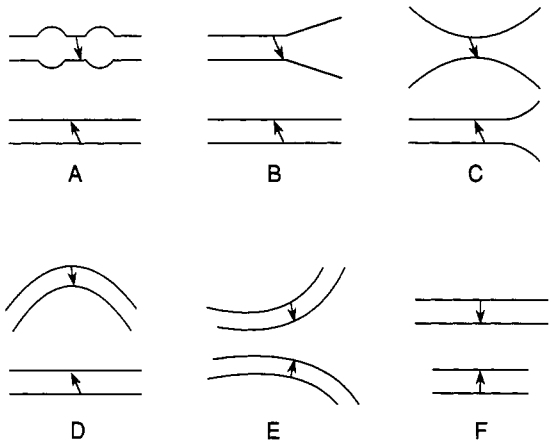


FIGURE 4: Putative structures of psoralen-cross-linked oligonucleotide duplexes during denaturing gel electrophoresis. Each of the six pairs of structures represents the upper cross-link (top structure) and the lower cross-link (lower structure). The arrows indicate the orientation of the psoralen molecule in the cross-links, from the furan ring toward the pyrone ring. Panel A represents what we currently believe is the major difference in cross-links of different mobilities. Panels B and C are variations of structures in which whole arms of the cross-linked duplex are melted. Panel D illustrates that the upper cross-link may possess a sharp kink in the DNA, leading to gel retardation. Panel E illustrates the possibility that gel retardation may be the result of the handedness of the supercoiled structure as suggested by Drak and Crothers (1991). Panel F illustrates the possibility that cross-linked duplexes may differ in length.

under the conditions of the DNA sequencing gel. To show this, we examined one arm of the cross-linked duplex at a time, holding the psoralen molecule in a constant orientation with the furan-side bonded to the bottom strand of the oligonucleotide duplex (Figure 5, structure 1). To make it easier to follow the data, the structure numbers in Figure 5 correspond to the lane numbers in Figures 7 and 9. These structures have an identical bottom strand that is ^{32}P labeled at each 5′ terminus. These structures were prepared in order to examine the base-pairing when an arm of the cross-linked duplex contains only six base pairs. Similarly, the structure numbers in Figure 6 correspond to the lane numbers in Figures 8 and 10. These structures were prepared in order to examine the base-pairing when an arm of the cross-linked duplex contains 10 base pairs.

For the perfectly base-paired duplexes in Figures 5 and 6, the cross-link isomer with the furan ring on the bottom DNA strand is called the upper cross-link because it is the psoralen orientational isomer of the slower electrophoretic mobility for this duplex (Figure 5, structures 1 and 10). The isomer of the faster electrophoretic mobility, the lower cross-link, is formed when the furan ring is attached to the top DNA strand. Upper cross-link (U) is seen in lanes 1 and 10 of Figure 7. Lanes 1–3 and 4–15 were run in two different gels. However, the relative mobilities of the upper cross-link bands for lanes 1 and 10 are almost the same. In Figure 5, structures 2 and 3 are similar to structure 1 except for the mispairing of two or four base pairs, respectively. One can see in lanes 2 and 3 of Figure 7 that the unpairing of the leftmost base pairs in these two structures resulted in a further decrease in the mobility of the upper cross-link band. The decrease in mobility is greater for the melting of the first two base pairs than the second two base pairs. Structures 4–10 address the base-pairing of the right arm of a cross-linked duplex when there are only six base pairs in that arm. Starting with the upper cross-link band of structure 10, structures 9–4 progressively increase the mispairing of the right arm one base at a time from the terminus, with corresponding decreases in gel

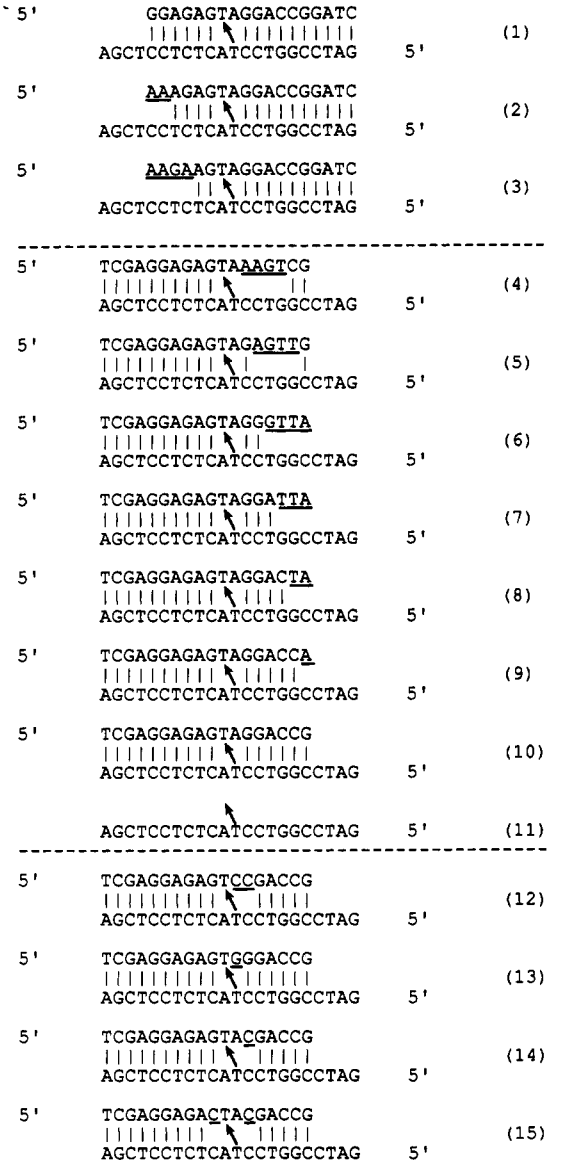


FIGURE 5: Sequences and structures of the DNA duplexes for the experiments in Figures 7 and 9, examining the base-pairing of the six base pairs nearest to the cross-linked TpA site. The structure numbers correspond to lane numbers in Figures 7 and 9. Vertical lines represent hydrogen bonding except for the central TpA site, where the vertical lines are omitted for clarity. All the duplexes in Figures 5 and 6 have the same bottom strand. The bases that are changed in the top strand in order to break hydrogen bonding by mismatch are underlined. The structures are divided into three groups: those with the mismatched bases on the left side of the duplex, those with the mismatched bases on the right side, and those with the mismatched bases located centrally at or next to the psoralen-modified base pairs. Arrows between the duplexes indicate TMP molecules from the furan side to the pyrone side for the experiments in Figure 7. However, for the experiments in Figure 9, the arrows do not indicate the orientation of the psoralen molecules because random reaction of TMP was used to produce two orientational isomers in each duplex.

mobilities seen in lanes 9–4 of Figure 7. Comparing lanes 9 and 10, one sees the great helix stability imposed by an interstrand cross-link on a DNA duplex. Even the last base pair at the terminus was base-paired under the conditions of DNA sequencing gels.

Similarly, the above examination of the base-pairing properties of cross-linked duplex was repeated for cross-linked arms of 10 base pairs. Figure 6, structure 1, is the same furan-side monoadducted bottom strand illustrated by structure 11 of Figure 5. Structure 2, Figure 6 (same as structure 10, Figure 5), is the upper cross-link formed when this monoad-

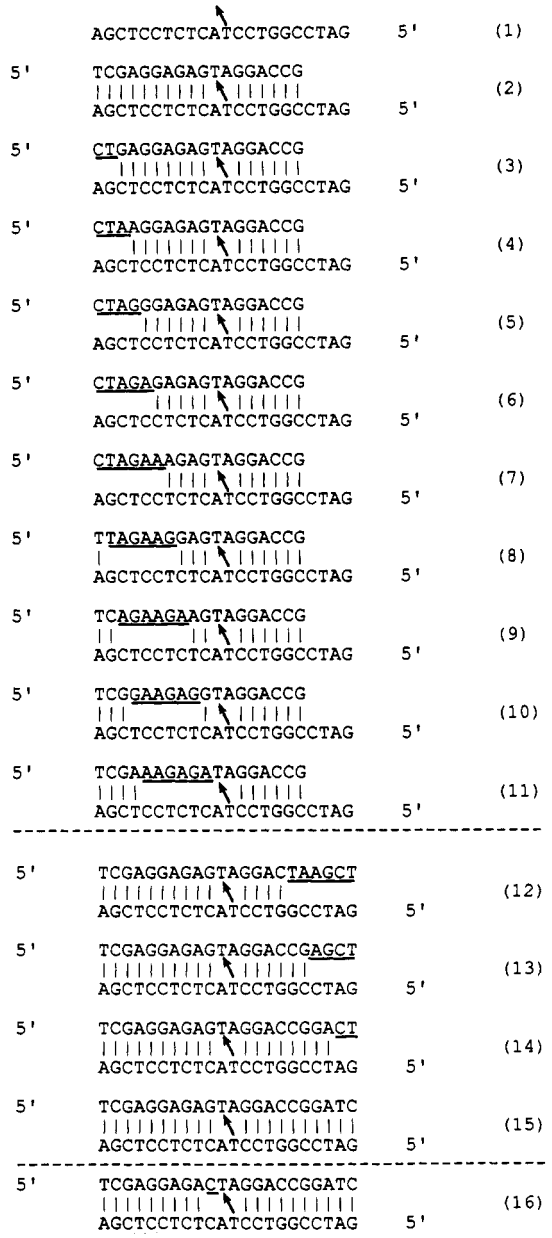


FIGURE 6: Sequences and structures of the DNA duplexes for the experiments in Figures 8 and 10, examining the base-pairing of the 10 base pairs nearest to the cross-linked TpA site. The structure numbers correspond to lane numbers in Figures 8 and 10. The representation of this figure is the same as for Figure 5. Arrows between the duplexes indicate TMP molecules from the furan side to the pyrone side for the experiments in Figure 8. However, for the experiments in Figure 10, the arrows do not indicate orientation of the psoralen molecules because random reaction of TMP was used to produce two orientational isomers in each duplex.

duct is hybridized to a complementary strand and further irradiated by 365-nm UV. Structures 3–11 of Figure 6 progressively have mismatched base pairs introduced at the leftmost terminus. Structures 12–14 of Figure 6 contain mismatched base pairs on the rightmost six base pairs of the duplex. The denaturing gel analysis of these structures is shown in Figure 8. Figure 8 extends the conclusion of Figure 7 to cross-linked duplexes with arms of 10 base pairs and illustrates that the two termini of a cross-linked duplex are hydrogen bonded even under the denaturing conditions of a DNA sequencing gel. This observation raises three points. First, by using base-pair mismatches to make sure that the two strands in an arm of a cross-linked duplex cannot form base-pairing, we show that retardation of the gel mobility of a cross-link can be the result of the melting of the base pairs in the helix. Second, if the two ends of each cross-linked duplex are correctly base-paired, then the difference in mobility of the upper cross-link and the lower cross-link must come from some feature between the two termini. This concern will be addressed further below. Third, the great stability of the cross-linked helix suggests caution in experiments in which psoralen is used to determine the amount of double-strandedness in nucleic acid structures. The great stability of the cross-linked helix may be a point of nucleation whereupon the length of the double-stranded region becomes exaggerated.

The Stabilities of the Base Pairs in a Cross-Linked Duplex Are Not Even. From the mobilities of the mismatched cross-link duplexes in Figures 7 and 8, the general observation that mobility decreases as the number of melted base pairs increases seems to be true. However, in order to use mobility as a probe of the structure of the cross-linked duplex in different DNA sequences, one should refine the observation and determine whether the melting of each base pair has the same incremental effect on mobility retardation. Conversely, an unusual increase in the amount of mobility retardation due to the introduction of a mismatch at a specific base pair may reflect destabilization of more than one base pair by the single mismatch specified, or the next base may be one of unusually low stability. In Figure 7, examining arms of six base pairs in the cross-linked duplex, one can see that the largest jump in the retardation of mobility in the left arm of the cross-link is between lanes 1 and 2 and in the right arm is between lanes 7 and 8. The jump between lanes 7 and 8 was due to the mismatch of a single base. This may suggest that the new T·G mispairing in structure 7 leads to partial melting of the next A·T base pair on the left such that the structure resembles that of structure 6. That the destabilized base pair here is an A·T base pair that is inherently less stable than a G·C base pair is not the reason for this enhanced destabilization because other examples later would be observed at G·C base pairs.

Similarly, in Figure 8, for 10-base-pair long arms of TMP cross-linked duplex, one can see that the largest decrease in mobility for the left arm is between structures 2 and 3 (lanes 2 and 3). For the right arm, the largest jump in mobility decrease is between structures 12 and 13 (lanes 12 and 13). In both cases, the next base is a G·C base pair. Thus whether the base pair is A·T or G·C does not contribute significantly to the location of the base pairs of lowest stability in a cross-linked duplex. The decrease in stability in those base pairs must be substantially greater than the stability of the single hydrogen-bond difference between a G·C and an A·T base pair. From these data, one can tentatively conclude that, in each arm of a cross-linked duplex, there exist one or more base pairs of lower stability than their neighbors. Quantitative measurements of these stabilities in the cross-link must await spectroscopy experiments. While our focus is on the hypothesis that these mismatches mainly disrupt hydrogen bonding, we have not ruled out the possibility that some mismatches at particular locations may dramatically alter the conformation of the cross-linked duplex in unforeseen ways.

The Location of the Bases with the Least Stability Changed with the Length of the Duplex. In an experiment to elucidate the structure of a cross-link, whether using X-ray crystallography or nuclear magnetic resonance spectroscopy, one is faced with the limitations of the methods and would prefer to use the minimum number of nucleotides in the helix. Thus it is unexpected that we observed a dependence of the location of the base pairs of the least stability on the length of the oligonucleotide duplex. Perhaps the length of the DNA duplex should be a concern in the structural determination of in-

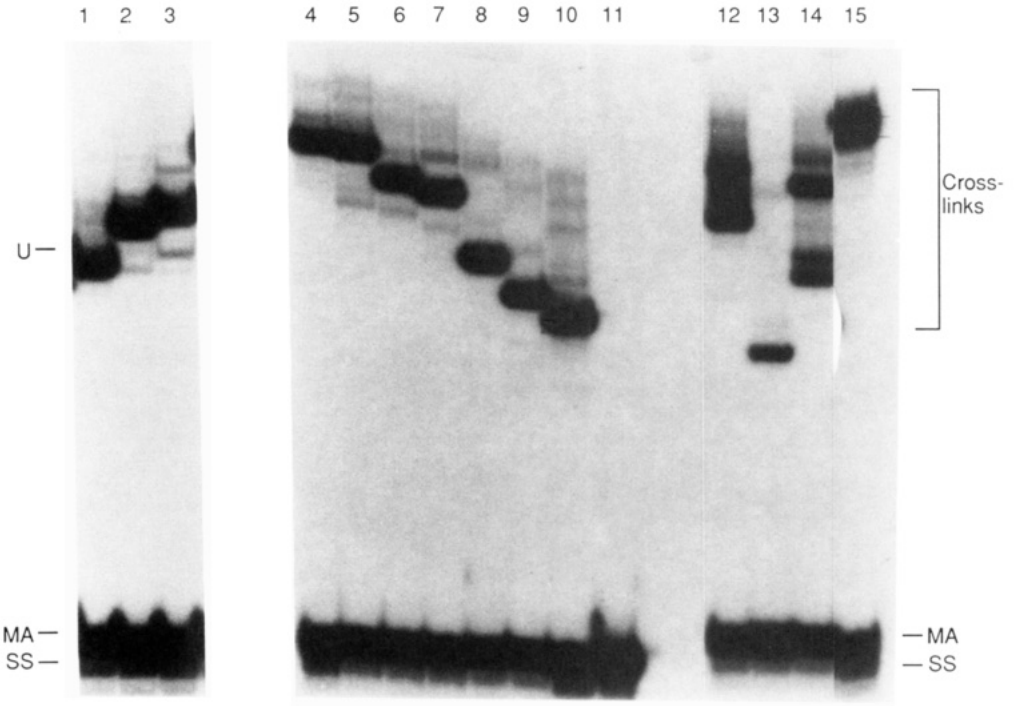


FIGURE 7: Electrophoretic analysis of TMP-cross-linked oligonucleotides, examining the base-pairing of the six base pairs nearest to the cross-linked TpA site. This is shown as an autoradiogram of a 7 M urea, 49 °C DNA sequencing gel, 20% acrylamide [acrylamide: bis-(acrylamide) 37.5:1] in 1× TBE. The bottom strand is 5′ labeled. The cross-links contain the furan ring on the bottom DNA strand and they were prepared as described in Experimental Procedures. MA, bottom-strand oligonucleotide containing furan-side monoadduct; SS, unmodified bottom-strand oligonucleotide; U, upper cross-link. Lanes 1–3 were run in a different gel than lanes 4–15 and therefore they show slight differences in electrophoretic mobility. Longer electrophoresis leads to greater separation of U from MA in lane 1 than in lane 10, although relative mobility measurements, with respect to the migration of the monoadduct bands, showed that the bands of structures 1 and 10 have similar relative mobilities. The same is true for the upper cross-link bands and lower cross-link bands in lanes 1 and 10 of Figure 9, respectively.

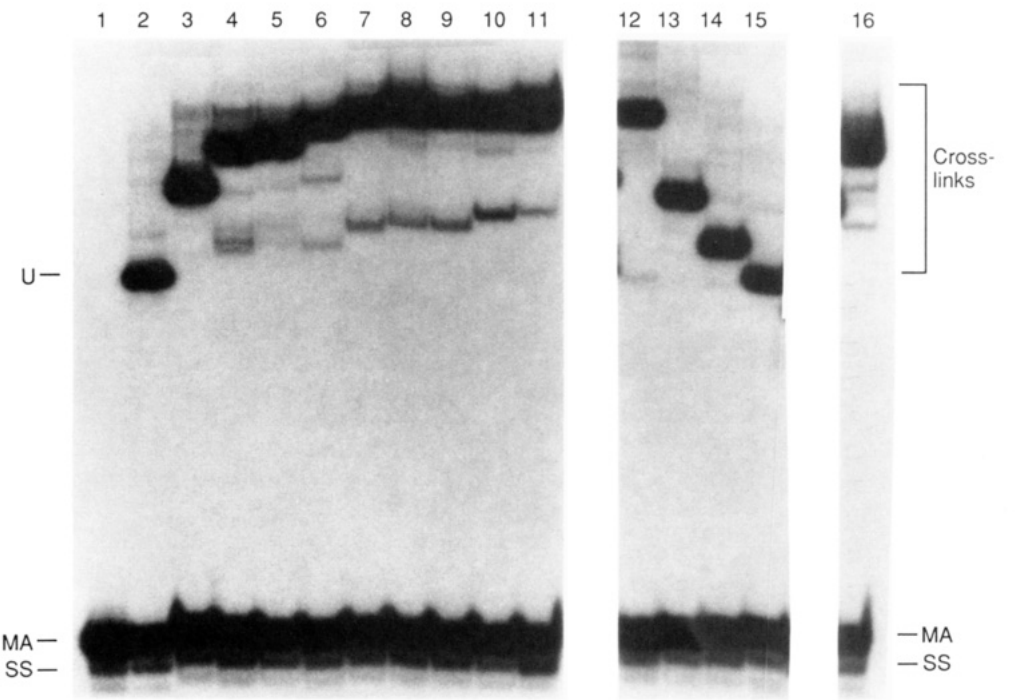


FIGURE 8: Electrophoretic analysis of TMP-cross-linked oligonucleotides, examining the base-pairing of the 10 base pairs on each side of the cross-linked TpA site. The lane numbers correspond to the structure numbers in Figure 6. The structures were prepared as for the experiments in Figure 7. The formation of the upper cross-link was assured by first purifying the TMP-reacted bottom strand containing the furan-side monoadduct.

terstrand cross-links. It is possible that, for a psoralen interstrand cross-link located in a long DNA fragment, the destabilized base pairs may be located as much as 12 bases away from the cross-linked bases. If that were true, then the location of the destabilized bases has the potential to be related to our previous observation that the base composition at these locations determines whether the repair enzyme UvrABC endonuclease will nick the TMP cross-linked DNA on the furan-side strand or on the pyrone-side strand (Jones & Yeung, 1990).

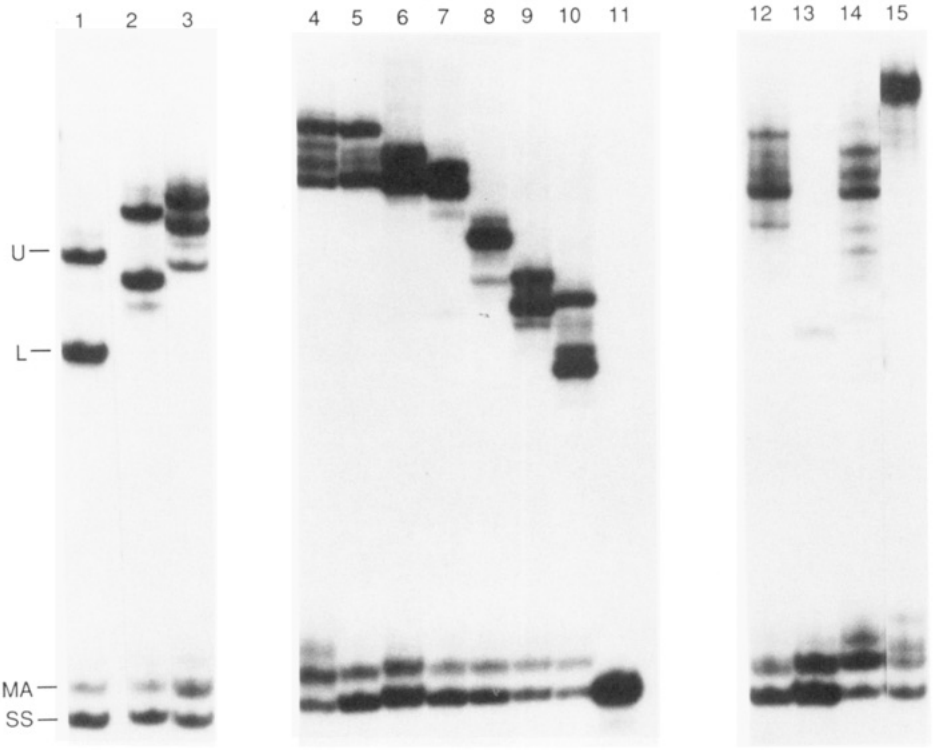


FIGURE 9: Electrophoretic analysis of both cross-link isomers of TMP-cross-linked oligonucleotides, examining the base-pairing of the six base pairs nearest to the cross-linked TpA site. To contrast the effects of base-pair mismatches on the electrophoretic mobilities of the two cross-link isomers, the structures in Figure 5 were formed by hybridization of the 5′ ^{32}P-labeled bottom strand with the respective top strands and photoreacted with TMP in 365-nm UV as described for 10 min. The reacted samples were dried, dissolved in tracking dye, and analyzed on a 20% acrylamide DNA sequencing gel. The lane numbers correspond to the structure numbers in Figure 5. By comparing Figure 7 with Figure 9, one can see that the topmost major band in Figure 9 is the same as corresponding bands in Figure 7. Thus the positions of the lower cross-link bands in each structure can be assigned.

We do not have the information necessary to determine why the arms of the cross-linked duplex are least stable in the middle. Perhaps the psoralen cross-link distorts the helix in each arm, but the stress destabilizing the helix is partly overcome on one side by the physical linkage of the two strands at the psoralen molecule and at the other side, at each duplex terminus, by the increase in hybridization of the two DNA termini because of their proximity in a cross-linked duplex. The net result is that the stress on the helix may be manifested as somewhat destabilized base pairs near the middle of each arm of the duplex, and the relative position depends on the length of each arm and possibly the base composition.

Differences between Upper Cross-Link and Lower Cross-Link. A key feature in our electrophoretic analysis of psoralen cross-links has been the observation that each DNA sequence produces two orientational isomers of different electrophoretic mobilities. To assess whether there is any significance to this difference between cross-link isomers, in Figures 9 and 10, we show base pair by base pair comparisons of the stabilities of the two cross-link isomers. Because there are two cross-link isomers in each lane of Figures 9 and 10, it is necessary to assign the bands in each lane. This is done by comparing Figures 7 and 8, in which only the upper cross-links are visualized, with Figures 9 and 10, in which both the upper cross-link and the lower cross-link bands are visualized. It is apparent that the top major band is always the upper cross-link isomer and the lower major band is the lower cross-link isomer. By following the melting behavior of each arm as mismatches are introduced, one can see that the two cross-link isomers have significantly different local stabilities. Other minor bands are visible, but it is unclear whether they represent alternative conformations of these two major forms. Instead, these minor bands may have originated from side reactions of TMP with the oligonucleotides or from the possibility that the oligonucleotides may not be completely homogeneous.

The differences between the upper cross-link and the lower cross-link can be demonstrated both in cross-link duplex arms six bases long and in those 10 bases long. One can analyze the six base pairs on the left side of the cross-linked duplex by starting at lane 1 of Figure 9 where the duplex is correctly base-paired. This duplex shows the typical resolution of the two cross-link isomers (U and L). As the base pairs at the left termini are melted by mismatches, the relative mobility difference of the two isomers is maintained in the duplexes in lane 2, until in lane 3, when the third and fourth base pair farthest from the psoralen molecule are melted, the difference in the mobilities between the two cross-link isomers disappeared. Furthermore, one can examine the six base pairs on the right side of the cross-linked duplex and see that the melting of the fifth and sixth base pairs from the TpA site (compare lanes 8 and 9) is enough to account for the observed difference between the two cross-link isomers of structure 10. Similarly, for the 10-base-pair long arms of a cross-linked duplex, in Figure 10, the difference between upper cross-link and lower cross-link can be attributed to the base pairs of the fifth and sixth bases from the left of the TpA site (compare lanes 5, 6, and 7) and the seventh and eighth bases from the right side of the TpA site (compare lanes 13 and 14). This comparison supports our interpretation above that the base pair of the least stability moves away from the TpA site as the length of the arms of the cross-linked duplex increases. The location of the base pair of the lowest stability in each arm of the cross-linked duplex coincides with the location of the differences between the upper cross-link and the lower cross-link for cross-link arms of either six or 10 base pairs.

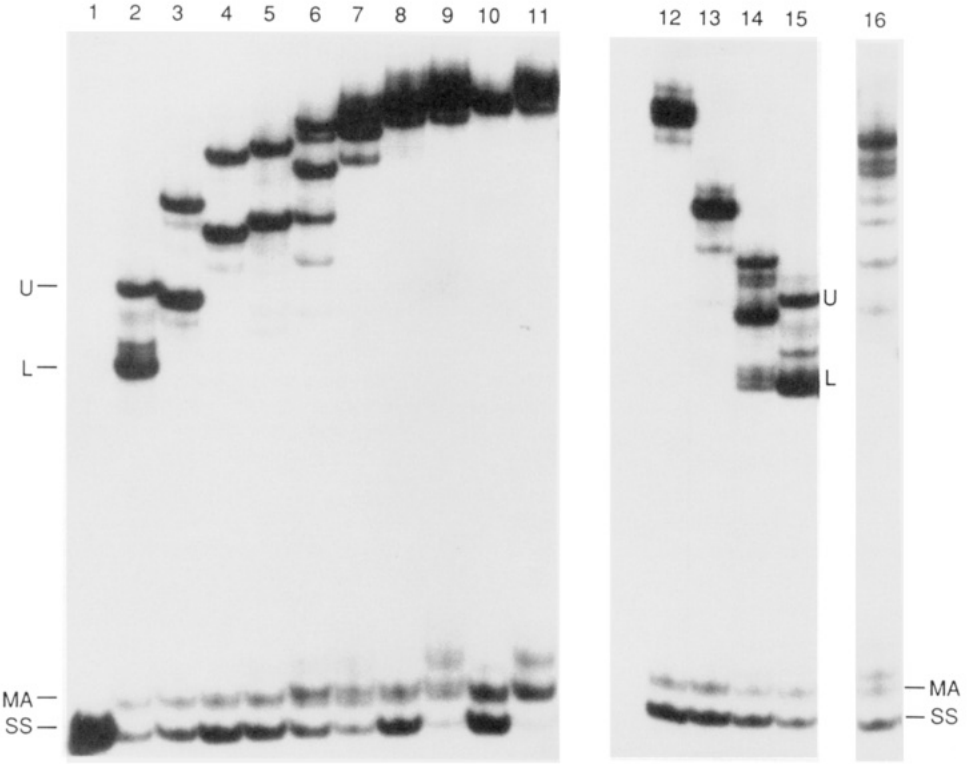


FIGURE 10: Electrophoretic analysis of both cross-link isomers of TMP-cross-linked oligonucleotides, examining the base-pairing of the 10 base pairs on each side of the cross-linked TpA site. This figure is the same as Figure 9 except the structures of Figure 6 were used in random photoreaction with TMP. The lane numbers correspond to the structure numbers in Figure 6. By comparing Figure 8 with Figure 10, one can see that the topmost major band in each lane of Figure 10 is the corresponding band in Figure 8. Thus the positions of the lower cross-link bands in each structure can be assigned.

Structures A and D of Figure 4 May Describe the Conformation of TMP Cross-Linked DNA Duplex under Our Electrophoretic Conditions. Six pairs of hypothetical conformations of psoralen cross-links during denaturing gel electrophoresis are shown in Figure 4. Results described above eliminated structures B and C in this figure as explanations of the differences between upper cross-link and lower Cross-link. Psoralen intercalation and cross-linking unwinds and extends a helix by about 1 bp (Wiesehahn & Hearst, 1978; Haran & Crothers, 1988). Structure F illustrates the possibility that helix length may lead to differential mobilities. This is not a good explanation for our observation because our short oligonucleotide duplexes each contain only one intercalated cross-linked psoralen molecule. The amount of helix unwinding and extension of the two psoralen orientational isomers may be similar and their mobility difference must be smaller than that expected for the addition of a single base to a DNA duplex. This limitation is not consistent with our observation that the mobilities of the two cross-link isomers differ by approximately a spacing of six bases. An alternate hypothetical explanation of DNA gel mobility is that one helix may have the left-handed supercoiled conformation and the other the right-handed supercoiled conformation (Drak & Crothers, 1991) (structure E, Figure 4). This interesting hypothesis is unlikely to apply to our case because the oligonucleotide duplexes are probably too short to possess the properties of supercoiled DNA.

Structures A and D in Figure 4 remain as possible explanations for the differences between upper cross-link and lower Cross-link. Structure A describes a cross-linked duplex in which each arm is more stably base-paired at the termini than at the base pairs between the termini and the interstrand cross-link. These regions of lower stability may be more unstable in the upper cross-link than in the lower cross-link. In a DNA sequencing gel, an unstable base pair may lead to slower electrophoretic mobilities similar to the ability of a single base mismatch to form a bubble and retard the mobility of a 200-base-pair fragment in denaturing gradient gel electrophoresis (Frischer & Lerman, 1979, 1983; Myers et al., 1987), a method that is used for the identification of sequence polymorphism. The regions of lower stability may cause our upper cross-link and lower cross-link to contain bubbles of different sizes under our electrophoresis conditions. The data in Figures 9 and 10 would support this hypothesis by illustrating that the lower cross-link is of greater helix stability than the upper cross-link in all the duplexes we examined.

In structure D, upper cross-link and lower cross-link may differ in the amount of kink at the cross-linking site. Polyacrylamide gel electrophoresis is normally not expected to resolve small oligonucleotide duplexes differing only in the amounts of kinking. In our analysis system, structure D would play a role only if there is a relationship between the kinking at a cross-link and the stability and melting of the cross-linked base pairs. Whether a psoralen interstrand cross-link creates a kink in the DNA helix is still a matter of controversy (Pearlman et al., 1985; Siden & Hagerman, 1984; Peckler et al., 1982; Haren & Crothers, 1988). The experiments described in this paper do not directly address this issue.

One Base Pair on Each Side of the Cross-Linked TpA Site Is Much More Stable Than Other Base Pairs in the Cross-Link. We investigated the effects of proximity of the base-pair mismatch to the psoralen cross-link and whether a base-pair mismatch nearer the psoralen molecule might have greater effects on duplex stability than mismatches at the termini. The structures we tested are shown in Figure 5 (structures 12–15) and Figure 6 (structure 16). The gel mobility of structure 16, Figure 6, is shown in Figures 8 and 10, lanes 16. It appears that a single C·C mismatch one base 3′ of the TpA site on the furan-side strand is sufficient to produce the same amount of destabilization as any six mismatched base pairs

in structures 7–11 of Figure 6 (lanes 7–11 of Figures 8 and 10)! Similarly, but to different extents, one can see other situations where a large amount of helix destabilization results from mismatches near the psoralen cross-link (lanes 12, 14, and 15 in Figure 7; lane 13 is an exception that will be discussed below). Each mismatched base pair immediately next to a cross-linked TpA site has destabilized the helix by the equivalent of 4–6 mismatched base pairs located farther from the cross-link. Lanes 15, Figures 7 and 9, show that the combined effects of two mismatches, one 5′ and one 3′, in the same molecule can add up to the largest amount of electrophoretic retardation seen in this work.

The presence of a base pair of unusual stability on either side of the cross-linked TpA site may explain some of the minor bands seen in Figures 7–10. In Figures 9 and 10, the two main bands in each lane are the upper cross-link and the lower cross-link. We know these are distinct species because they retain their respective gel mobilities after repeated isolation and electrophoresis in these denaturing DNA sequencing gels. However, in these figures, there are minor bands which have not been characterized as to whether they are conformers of the major bands. In a normal DNA helix, one may expect that small DNA duplexes may not maintain bands of alternative conformations during electrophoresis. However, in the case of cross-linked duplexes, the presence of these two base pairs of unusual stability may form sufficient barriers for the transition of conformations during electrophoresis. For example, the conformers may differ in whether these base pairs are melted.

The T·A Base Pair 5′ of the Furan Side of the TMP-Cross-Linked TpA Site May Not Be Hydrogen Bonded. In striking contrast to the expectation that a mismatched base pair will destabilize a helix, a G·T mismatch within the TpA cross-linking site produced no destabilization of the upper cross-link band in the DNA sequencing gel (structure 13, Figure 5; lane 13, Figure 7). This result illustrates that there is normally no need for hydrogen bonding, at the T·A base pair on the 5′ side of the furan side of the TMP-cross-linked TpA site, under our gel analysis conditions. Moreover, the mobility of this cross-link slightly increased, reflecting greater helix stability than the original cross-link containing no mismatch. It is possible that a new hydrogen bond may have formed between the G·T mismatch where there was none in the original T·A base pair. Tomic et al. (1987) detected a similar lack of hydrogen bonding in their cross-link with AMT but at the other T·A base pair within the TpA site. They showed that the T·A base pair on the 3′ side of the furan-side strand of the TpA site is not hydrogen bonded while the T·A base pair on the 5′ side of the furan side of the cross-linked TpA site is very stable. In our case, it remains to be directly determined whether the 3′ T·A base pair can accommodate a mismatch for the TMP cross-link. It appears that further gel mobility studies similar to those described here can quickly shed light on the effects of various psoralen derivatives on cross-link structure.

Conclusions on the Structure of the Psoralen Interstrand Cross-Link. It can be concluded from this work that a spectrum of cross-link conformations are possible. A summary of our findings on the structure of a TMP cross-link of lower stability (the upper cross-link) is shown in Figure 11. Whether this structure which is determined under our electrophoresis conditions is applicable to the physiological conditions remain to be established. The primary determinants of the structure of the cross-link are the steric and electronic factors imposed by the cross-linking agent, and one may find significantly

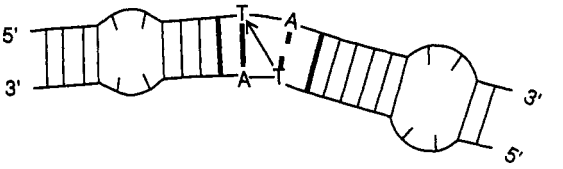


FIGURE 11: Structure of a cross-link of TMP of lower helix stability during electrophoresis in denaturing polyacrylamide gels. This structure consists of an oligonucleotide duplex of 22 base pairs containing a 5′ TpA site in the middle. The directions of the DNA strands (horizontal lines) are indicated. The orientation of the psoralen molecule is indicated by an arrow in the TpA site, with the furan ring to pyrone ring orientation indicated by the direction of the arrow. The vertical lines joining the two DNA strands represent hydrogen-bonded base pairs and the disrupted vertical lines represent non-hydrogen-bonded base pairs. The vertical lines of double thickness represent base pairs of unusually high stability. This structure illustrates that, under our analysis conditions, the TMP cross-link in a DNA sequence of lower stability shows disruption of base-pairing on the 5′ side of the furan-side strand within the cross-linked TpA site, but the base pairs on either side of the TpA site are very stable. A region of decreased helix stability occurs in the mid-section of each arm of this cross-linked sequence. The two duplex termini are properly base-paired. The amount of bend in the duplex at the TpA site is hypothetical, although it is reasonable to assume that the helix should be distorted in some manner when the normal base-pairing is disrupted.

different cross-link conformations with 8-MOP, TMP, or AMT. Agreeing with the observation of Tomic et al. (1987), we find that the most prominent helix distortion is within the cross-linked TpA site. However, different psoralen derivatives may differ in whether the thymine moiety bonded to the furan ring, or the one bonded to the pyrone ring, is hydrogen bonded to the adenosine moiety in the respective opposite strand. Moreover, our current analysis of this unpaired base pair at the cross-linked site is limited to the upper cross-link, the cross-link with the lower helix stability. A base-pair mismatch at this site has not been tested for the lower cross-link, the cross-link of higher stability.

The two arms of the cross-linked helix may vary in helix stability. The AMT data of Tomic et al. (1987) showed that the 3′ side of the furan-side strand of AMT is less stable than the 5′ side by as much as 20 °C in melting temperature. We do not seem to see significant differences of 5′ side versus 3′ side in our experiments with these TMP cross-links. It is unclear at present whether the differences originated from the use of different DNA sequences or from the structures of different psoralen derivatives. The individual base pairs near a cross-link can vary greatly in stability (Figure 11). We showed that hydrogen bonding of the base pair on each side immediately next to the cross-linked TpA site is critical for maintaining the helical structure of the cross-link. This follows from our observation that a single mismatch in those two positions leads to melting of at least 4–6 base pairs. Farther away from the cross-linked base, next to this stable region on each arm of the cross-link, there are one or two base pairs of weaker hydrogen bonding. The locations of these unstable base pairs on each arm of the cross-link moves farther away from the cross-linked site as the length of the DNA increases. It is mainly the location and helix destabilization of these two unstably base-paired regions on the two arms of the cross-link that seem to produce differential mobilities in the two cross-link isomers of a given DNA sequence. It is likely that this helix destability is sequence dependent. Farther away from the cross-link, at the two ends of the DNA duplex, the bases appeared properly hydrogen bonded.

In this study, we have not yet addressed the question of why upper cross-link and lower cross-link differed in the stabilities

of the two regions of the lowest stabilities. The explanation may lie in the interactions of the TMP molecule with the DNA sequences near the cross-link. Sequences of higher stacking energy and higher number of hydrogen bonds on both arms may allow the least amount of destabilization of the helix and therefore the greatest gel mobility. Sequences with lower stacking energy and lower number of hydrogen bonds may have the least stability and therefore the lowest gel mobility. When one arm is high in stability and the other low, then steric hindrance due to the relative orientations of the psoralen molecule will add or subtract from the stability of that arm and result in helixes of intermediate stability and gel mobility. Therefore, if one should consider a collection of DNA sequences in which each DNA duplex contains a psoralen cross-link in the middle of the helix, there should be a spectrum of helix stabilities for the cross-linked duplexes of different sequences and therefore a spectrum of gel mobilities. The explanation for the usual observation that two psoralen cross-link orientational isomers possess two different gel mobilities is consistent with this reasoning. A model oligonucleotide sequence forms two cross-link isomers. These two isomers are equivalent to having a fixed orientation of the psoralen molecule within two cross-linked sequences out of this spectrum of possible cross-links. Thus it is fortuitous that psoralen cross-link orientational isomers often show different gel mobilities. Had the psoralen molecule been perfectly symmetrical in structure, the two cross-link isomers would be identical. It follows that some lower cross-links may be more stable than others and possess faster gel mobilities. These cross-links may be more helical in conformation, less likely to melt, and possibly less kinked.

Is the local destability of the arm of a cross-link important in biology? The amount of energy involved in the variations of helix stabilities we investigated in this work are of the order of one or two base pairs, a magnitude that can be used in protein–DNA interactions. A potential system that may respond to these structural differences may be the nucleotide excision repair enzymes, for example, the UvrABC endonuclease of *E. coli* which incises some cross-links on the furan-side strand and other cross-links on the pyrone-side strand (Jones & Yeung, 1990). The efficiencies of UvrABC endonuclease incision at TMP cross-links also seem to be influenced by DNA sequence (Jones & Yeung, 1989). It remains to be determined whether the activity of other enzymes of DNA repair, such as DNA helicase II, DNA polymerase I, and RecA in *E. coli*, or other nucleotide excision repair systems of higher organisms, as well as psoralen cross-link mediated mutagenesis, may be modulated by the DNA sequence and the DNA helix stability variations near the psoralen cross-link.

Our work here shows that the cross-link greatly stabilizes the DNA duplex state. This great stability may have the potential for a psoralen cross-link to serve as a point of nucleation whereupon the length of the helical duplex region may become exaggerated. Moreover, we have proved here that perfect sequence homology is not a stringent requirement for psoralen cross-linking of DNA duplexes. Experiments that use psoralens to mediate hybridization or to show double-strandedness or triple-strandedness may therefore need to be evaluated with caution.

REFERENCES

Ashwood-Smith, M. J., Poulton, G. A., Barker, M., & Mildenburger, M. (1980) *Nature (London) 285*, 407–409.

Bridges, B. A., Mottershead, R. P., & Knowles, A. (1979) *Chem.-Biol. Interact. 27*, 221–233.

Cimino, G. D., Gramper, H. B., Isaacs, S. T., & Hearst, J. E. (1985) *Annu. Rev. Biochem. 54*, 1151–1193.

De Gennes, P. G. (1971) *J. Chem. Phys. 55*, 572–579.

Doi, M., & Edwards, S. F. (1978) *J. Chem. Soc., Faraday Trans. 274*, 1789–1801.

Drak, J., & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3074–3078.

Fischer, S. G., & Lerman, L. S. (1979) *Cell 16*, 191–200.

Fischer, S. G., & Lerman, L. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1579–1583.

Grossman, L., & Yeung, A. T. (1990) *Mutat. Res. 236*, 213–221.

Haren, T. E., & Crothers, D. M. (1988) *Biochemistry 27*, 6967–6971.

Jones, B. K., & Yeung, A. T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 8410–8414.

Jones, B. K., & Yeung, A. T. (1990) *J. Biol. Chem. 265*, 3489–3496.

Kanne, D., Straub, K., Hearst, J. E., & Rapoport, H. (1982a) *J. Am. Chem. Soc. 104*, 6754–6764.

Kanne, D., Straub, K., Rapoport, H., & Hearst, J. E. (1982b) *Biochemistry 21*, 861–871.

Kirkland, D. J., Creed, K. L., & Mannisto, P. (1983) *Mutat. Res. 116*, 73–82.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 498–560.

Myers, R. M., Maniatis, T., & Lerman, L. S. (1987) *Methods Enzymol. 155*, 501–527.

Parsons, B. J. (1980) *Photochem. Photobiol. 32*, 813–821.

Pearlman, D. A., Holbrook, S. R., Pirkle, D. H., & Kim, S.-H. (1985) *Science 227*, 1304–1308.

Peckler, S., Graves, B., Kanne, D., Rapoport, H., Hearst, J. E., & Kim, S. H. (1982) *J. Mol. Biol. 162*, 157–172.

Sancar, A., & Rupp, W. D. (1983) *Cell 33*, 249–260.

Sancar, A., & Sancar, G. B. (1988) *Annu. Rev. Biochem. 57*, 29–67.

Shi, Y. B., & Hearst, J. E. (1986) *Biochemistry 25*, 5895–5902.

Sinden, R. R., & Hagerman, P. J. (1984) *Biochemistry 23*, 6299–6303.

Song, P. S., & Tapley, K. J. (1979) *Photochem. Photobiol. 29*, 1177–1197.

Tessman, J. W., Isaacs, S., & Hearst, J. E. (1985) *Biochemistry 24*, 1669–1676.

Tomic, M. T., Wemmer, D. E., & Kim, S. H. (1987) *Science 238*, 1722–1725.

Van Houten, B. (1990) *Microbiol. Rev. 54*, 18–51.

Wiesehahn, G., & Hearst, J. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2703–2707.

Yeung, A. T., & Miller, C. G. (1990) *Anal. Biochem. 187*, 66–75.

Yeung, A. T., Mattes, W. B., Oh, E. Y., & Grossman, L. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 6157–6161.

Yeung, A. T., Jones, B. K., Capraro, M., & Chu, T. (1987) *Nucleic Acids Res. 15*, 4957–4971.

Yeung, A. T., Jones, B. K., & Chu, T. (1988) *Biochemistry 27*, 3204–3210.