Solution structure of oligonucleotides covalently linked to a psoralen derivative

O. Bornet, C. Prévost, F. Vovelle, M. Chassignol, N.T. Thuong and G. Lancelot*

Centre de Biophysique Moléculaire, CNRS, rue Charles Sadron, 45071 Orléans cédex 2, France

Received November 14, 1994; Revised and Accepted January 20, 1995

ABSTRACT

Psoralen (pso) was attached via its C-5 position to the 5'-phosphate group of an oligodeoxynucleotide d(TAAGCCG) by a hexamethylene linker (me). Complex formation between pso-m6-d(TAAGCCG) and the complementary strands d(CGCTTAT)[7-8mer] or d(CGCTTA)[7-7mer] was investigated by nuclear magnetic resonance in aqueous solution. Structural informations derived from DQF-COSY and NOESY maps, revealed that the mini double helix adopts a B-form conformation and that the deoxyriboses preferentially adopt a C2'-endo conformation. The nOe connectivities observed between the protons of the bases or the sugars in each duplex, and the protons of the psoralen and the hexamethylene chain, led us to propose a model involving an equilibrium between two conformations due to different locations of the psoralen. Upon UV-irradiation, the psoralen moiety cross-linked the two DNA strands at the level of 5'TpA3' sequences. NMR studies of the single major photo-cross-linked duplex pso-m6-dCTAAGCCG) and d(CGCTTA) were performed. The stereochemistry of the diadduct is indeed cis-syn at both cyclobutane rings. In addition, the effects of this diadduct on the helical structure are analyzed in detail.

INTRODUCTION

Psoralens are bifunctional photoreagents which form covalent bonds with the pyrimidine bases of nucleic acids upon UV irradiation (1). They are now successfully used in the phototherapy of skin diseases such as psoriasis (2), and have been used for many years as photoactive probes of nucleic acid structures and functions, and as tools to study the mechanisms of mutagenesis and repair processes (3). The binding of an oligonucleotide to its complementary sequence is a highly specific process governed by stacking interactions between base pairs and by hydrogen-bond formation between complementary bases. We have already shown that covalent attachment of an intercalating agent to an oligodeoxynucleotide strongly increases the stability of hybrid formation (4). Substitution of acridine derivatives by other agents such as an orthophenantroline-Cu complex leads to the cleavage of the complementary strand (5).

In this paper, we show that a psoralen derivative attached to an oligonucleotide via its C-5 can be targeted by photo-cross-link to a single strand selected sequence of nucleic acid. When intercalated at 5'TpA3' sites in DNA, psoralen forms cross-links between the two strands upon UV irradiation (1). It can thus be used to regulate gene expression providing that the target sequence is available for hydrogen-bond formation with the bases of the oligonucleotide. This is obviously the case if the sequence is located in a single-stranded nucleic acid, but also if it is part of a region of double-stranded DNA that transitorily opens during replication or transcription processes. We report on a nuclear magnetic resonance investigation on the complexes formed by a d(TAAGCCG) oligonucleotide covalently linked to a psoralen derivative and to the complementary sequence d(CGCTTAT) or d(CGCTTA). The interactions engaged between the psoralen ring and the nucleic acid bases are analyzed in detail.

MATERIALS AND METHODS

Oligonucleotide synthesis

Preparation of oligodeoxynucleotides d(CGCTTAT), d(CGCTTA), and pso-m6-d(TAAGCCG) were carried out in 10 μmol scale on a Pharmacia automatic synthesizer using the phosphoramidite chemistry. Incorporation of the psoralen group at the 5'-end of the heptamer was achieved via the phosphoramidite [obtained by condensation of 2-cyanoethyl-N,N-diisopropylchlorophosphite with 5-(6-hydroxyhexyloxy)-psoralen in the presence of diisopropylethylamine] (6). After an oxidation step of the phosphite intermediate using the standard procedure, the psoralen-oligonucleotide was deprotected by treatment with concentrated ammonia. The oligomers were purified and analysed by ion exchange chromatography FPLC on a Mono-P: HR 5/5 column (Pharmacia) or a Pharmacia HR 5/5 column (Pharmacia) with a linear gradient of NaCl from 0 to 1.5 M in 37.5 min in 0.01 M NaH2PO4 (pH 6.8), 20% CH3CN using a flow rate of 1 ml/min: d(CGCTTAT) Rf = 12 min, d(CGCTTA) Rf = 12.8 min, pso-m6-d(TAAGCCG) Rf = 19.6 min.

NMR spectroscopy

All the investigated oligonucleotides were passed through a Chelex-100 column to remove paramagnetic impurities and were adjusted to pH 7.0. The NOESY data were obtained from degassed solutions contained in sealed tubes. NMR experiments

* To whom correspondence should be addressed
were carried out with a BRUKER AMX 500 spectrometer operating at 11.74 Tesla and processed on a X32 computer. Typically, spectra were recorded with 1024 complex data points in the $t_2$ domain and 512 increments in $t_1$. The data matrix was resolution-enhanced by a gaussian window function in direction 2 and by a shifted sinus bell window function in direction 1. Two-dimensional (2-D) data sets for DQF-COSY and NOESY spectra were collected in the sensitive mode with the time-proportional phase increment method (7–9). A 2-D TOCSY experiment was recorded at a 86 ms mixing time. Exchangeable protons were recorded by using the jump and return sequence (12).

Irradiation of samples

The intensity of light at the surface of the inner sample chamber in this device corresponded approximately at 820 W/m². Light was filtered through pyrex filters in a water bath to remove radiations below 310 nm, in order to avoid the formation of thymine dimer. Samples of 7-7mer and of 7-8mer were then irradiated in identical fashion and analyzed by ion exchange chromatography FPLC using the same conditions as described above.

Molecular modeling

The irradiated duplex modeling was performed using the internal coordinate program JUMNA of R. Lavery (13). Based on a combination of internal variables and helical variables, this program permits efficient and accurate structural deformations during energy minimization and allows various types of controls. The method has already proved powerful for generating 3-D structures of oligonucleotides from NMR data (14–15). In order to model the cross-linked psoralen, we first introduced a modified thymine species in the nucleic acids library associated with JUMNA. This modified thymine has its 5–6 double bond replaced by a single bond. No atom has been physically added to this bond, preparing the subsequent formation of covalent bonds with psoralen atoms C-3 and C-4 for T1 and C-4', C-5' for T2. Psoralen is treated as a ligand covalently linked to T1 and T2 through special types of bonding constraints on distances, valence angles and dihedral angles. The hexamethylene chain attached to psoralen is linked to the C-5' of the oligonucleotide backbone in the same way. The charges on the psoralen–thymine moiety were calculated using the Nchem software (16). This program uses a reparametrized Hückel Del Re method to calculate monopole charges on each atom of the system (17). It also defines internal variables for ligands. Inclusion of the NMR constraints was performed through diatomic distance constraints. We used a force constant value of 12 kcal mol⁻¹ Å⁻² for distances <3 Å and for sugar pucker constraints, and 6 kcal mol⁻¹ Å⁻² for greater distances.

RESULTS

Assignment of resonances and conformation of the 7-7mer duplex before irradiation

The mixing of pso-m₆(d(TAAGCCG)) with its complementary antiparallel sequence d(CGGCTTA) led to upfield shifts and a slight broadening of the resonance lines for both compounds (see Fig. 1 for nomenclature). The temperature dependence of a mixture containing 1.6 mM of each complementary strand in 0.1 M NaCl showed a melting temperature of 53°C. The method of sequential nOe assignment of non-exchangeable resonances in nucleic acids has been described elsewhere and will not be repeated here (18–26). Figures 2 and 3 show the connectivities between the aromatic protons H8 or H6 and the H1', H2' and H2'' protons of sugars. The aromatic–aromatic connectivities as well as the connectivities of thymines methyl group or H5 of cytosines with the aromatic proton of their 5' neighboring residue, confirm the assignment and the right-handed structure assumption. The H2 protons of adenines were assigned by considering their proximity with the H1' proton and the inter-residues H2(A1)–H2(A2), H2(A2)–H2(A3) connectivities. The H2'' resonances were distinguished from the H2' resonances by strong connectivities with proton H1' on NOESY maps obtained with short mixing times (50 or 70 ms). Hartmann–Hahn spectroscopy (data not shown) was used to confirm the assignment of the H3', H4', H5' and H5'' sugar protons made on NOESY and DQF-COSY spectra. The coupling constants $J_{1'2'}$ and $J_{1'2''}$ were measured on the DQF-COSY spectra and were found around 10.1 and 5.7 Hz, respectively. Moreover, an analysis of the DQF-COSY map shows that except for the quasi-degenerated H2' and H2'' resonances, each H2' resonance was connected to the H3' resonance while H2'' resonances did not show any connectivities. These results are in agreement with sugars in a C2'-endo conformation where $J_{2'3'} = 5.8$ Hz and $J_{2''3'} = 1.2$ Hz, rather than a C3'-endo conformation where two cross peaks should be observed ($J_{2'3'} = 7.3$ Hz and $J_{2''3'} = 9.7$ Hz).

Location of the psoralen derivative

The DQF-COSY, NOESY and TOCSY maps show connectivities at 5.83–7.48 p.p.m. and 6.43–7.24 p.p.m. which were assigned on the basis of their chemical shifts, to the couple of protons H3–H4 and H4'–H5' (Fig. 2). Moreover one resonance at 6.22 p.p.m. was assigned to the H8 resonance. The same kind of proximity or scalar coupling was observed on isolated pso-m₆(d(TAAGCCG)) in solution. All methylene resonances could be located on DQF-COSY and TOCSY maps. For a duplex concentration of 1.6 mM, the resonances of CH$_3$(1) and CH$_2$(6) groups are separated and only the resonance at 3.94 p.p.m. showed connectivities with the H4 and H4' protons of psoralen. This resonance was assigned to the CH$_2$(6) group. In order to locate the psoralen moiety, nOe spectra were recorded at 70, 100, 150, 200 and 250 ms mixing times. The proton resonances of psoralen and of methylene groups showed many strong connecti-
Figure 2. Expansion of the aromatic to H1' proton region (left) and the H1' to H1' proton region (right) of the NOESY spectrum of pso-m6-d(TAAGCGG) + d(CGGCTTA) at 15°C in 0.1 M NaCl, pH 7.0. The sequential assignment of the two strands is indicated with a solid line. Cross peaks between protons of bases and protons of psoralen are shown by arrows: H8(A1)–H8(pso), H2(A1)–H4'(pso), and H1'(A1)–H8(pso).

Figure 3. Expansion of the aromatic to H2'/H2'' proton region (left) and the H1' to H2'/H2'' proton region (right) of the NOESY spectrum of pso-m6-d(TAAGCGG) + d(CGGCTTA) at 15°C in 0.1 M NaCl, pH 7.0. The resonance frequencies of the base aromatic protons are given as well as the assignment of all H2'/H2'' resonance lines. T1(5), T2(5), T3(5) and CH2 show the position of the methyl group of thymines and of the methylene chain: PO4-CH2(1)-CH2(2)-CH2(3)-CH2(4)-CH2(5)-CH2(6)-pso. Cross peaks between protons of bases and protons of psoralen are shown: H2''(T2)–H8(pso), H2'(A1)–H8(pso), and H2'(A1)–H8(pso).

activities with the oligonucleotide protons of residue A1, and weak connectivities with residues T2 and A2 (Table 1).

**Duplex 7-8mer**

The assignment of all non-exchangeable resonances of the duplex pso-m6-d(TAAGCGG)-d(CGGCTTAT) were obtained by using the same procedure as for the 7-7mer duplex. Thymine 0 was assigned considering its connectivities with adenine A1. Although the structure (B form) and the position of the resonance lines were found quite analog for both oligonucleotides (except for A1–T2 junction), the location of psoralen was different (Table 1). Connectivities were found between protons H8 and H4' of psoralen and the protons of thymines T0, T1, T2 and adenine A1, while the protons H5' of psoralen showed some cross peaks with the protons of adenine A1. Many connectivities were also observed between methylene protons and thymine 1.
Figure 4. Expansion of the aromatic to the H1'/H5 protons region (left) and the H4'/H5'/H5'' to H1'/H5 protons region (right) of the NOESY spectrum (mixing time = 150 ms) of the interstrand cross-linked 7-7 dimer. The spectrum was recorded in D2O, 1.4 mM, 0.1 M NaCl, pH 6.5, at 15°C. Note the large upfield shifts of the H6 protons of the thymines 1 and 2. These large upfield shifts are due to the change in hybridization upon formation of a cyclobutane between thymines and psoralen. The dotted lines indicate the weakness of the connectivities H8(A1)-H1'(T2) and H8(A2)-H1'(T1).

Table 1. Observed connectivities between the psoralen moiety and the 7-7mer or the 8-7mer for a mixing time of 150 ms

<table>
<thead>
<tr>
<th>7-7 mer</th>
<th>7-8 mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4'(ps0)-H2(A1) s</td>
<td>H4'(ps0)-CH3(T1) w</td>
</tr>
<tr>
<td>H8(ps0)-H8(A1) s</td>
<td>H8(ps0)-H8(A1) s</td>
</tr>
<tr>
<td>H8(ps0)-H2'(A1) s</td>
<td>H8(ps0)-H2'(A1) s</td>
</tr>
<tr>
<td>CH3(1), CH3(2), CH3(4)-H6(T1) s</td>
<td>CH3(1), CH3(4)-H6(T1) s</td>
</tr>
<tr>
<td>CH3(2), CH3(4), CH3(5)-H1'(T1) w</td>
<td>CH3(2), CH3(5)-H1'(T1) w</td>
</tr>
</tbody>
</table>

s = strong, m = medium, w = weak. Small connectivities were observed at a mixing time of 200 ms between H8(ps0)-H2(A2), H8(ps0)-H1'(T2), H8(ps0)-H3'(A1) and H3(ps0)-H2(A1) for the 7-7mer.

Assignment of resonances and conformation of the photo-cross-linked 7-7mer duplex after irradiation

As already reported (27), the covalent photo-addition of psoralen across the 5-6 double bonds of thymines 1 and 2 shifts the H6 resonances into the same region as H4', H5' and H5'' resonances of sugars (Fig. 4 and Table 2). This was also observed for thymine photodimer (28). It is worth noting that sequential inter-residue H8(A1)-H1'(T2) connectivity was very weak and H8(A2)-H1'(T1) connectivity was not observed.

Another evidence for the formation of an interstrand cross-link is shown on Figure 5, which exhibits large upfield shifts for protons H5'(4.50 p.p.m.), H4'(4.05 p.p.m.), H4 (3.60 p.p.m.) and H3 (particular chemical shift at 1.85 p.p.m.) of sugars (Fig. 4 and Table 2). This was also observed for thymine photo-dimer (28). It is worth noting that sequential inter-residue H8(A1)-H1'(T2) connectivity was very weak and H8(A2)-H1'(T1) connectivity was not observed.

Irradiation of psoralen

Melting temperatures of 53 and 55°C were measured respectively for the 7-7mer and the 7-8mer. Samples of 7-7mer and of 7-8mer were irradiated at wavelength above 310 nm using the same experimental conditions (see Materials and Methods). After irradiation, up to 95% of the absorption band at 320 nm disappeared. FPLC on Mono-P column showed that up to 80% of the 7-7mer and the 7-8mer were converted into a single major species (photoproducts: Rt 7-7mer = 17.70 min, Rt 7-8mer = 18.30 min).

As already reported (27), the covalent photo-addition of psoralen across the 5-6 double bonds of thymines 1 and 2 shifts the H6 resonances into the same region as H4', H5' and H5'' resonances of sugars (Fig. 4 and Table 2). This was also observed for thymine photodimer (28). It is worth noting that sequential inter-residue H8(A1)-H1'(T2) connectivity was very weak and H8(A2)-H1'(T1) connectivity was not observed.

Assignment of resonances and conformation of the photo-cross-linked 7-7mer duplex after irradiation

As already reported (27), the covalent photo-addition of psoralen across the 5-6 double bonds of thymines 1 and 2 shifts the H6 resonances into the same region as H4', H5' and H5'' resonances of sugars (Fig. 4 and Table 2). This was also observed for thymine photodimer (28). It is worth noting that sequential inter-residue H8(A1)-H1'(T2) connectivity was very weak and H8(A2)-H1'(T1) connectivity was not observed.
with a cis stereochemistry. Only a cis–syn stereochemistry for both cyclobutane junctions can account for these results.

Portion of a NOESY spectrum in 0.1 M NaCl, pH 6.5, at 2°C, in 90% H2O/10% D2O is shown in Figure 6. The imino resonance of T2 in the cross-linked DNA is shifted upfield to ~11.3 p.p.m. from its position at 13.65 p.p.m. before irradiation, suggesting that it is only weakly hydrogen-bonded. No sequential connectivity is observed between imino protons of T1 and T2. This suggests that the imino proton of T1 is accessible to solvent and not hydrogen bonded.

**Solution structure determination**

A global correlation time of 3.46 ns was obtained by fitting the experimental NOE values for H6–H5 of cytosines for different mixing times. The 97 NOE values were converted into distances by using the relaxation matrix calculation (33). After the first constrained energy minimization, the resulting structure allowed us to calculate the new target distances using the complete 168 × 168 relaxation matrix (34) and the relation:

\[ d(\text{new}) = d(\text{old}) \times \text{[NOE(simulated)/NOE(experimental)]}^{16} \]

These new distances were determined by averaging the calculated d(new) values for six mixing times (30, 50, 70, 100, 150 and 250 ms) and were used in the constrained molecular mechanics program for a new iteration. After three iterations, this process converged (35) and reasonable agreement was found between the calculated and all the experimental interproton distances. The atomic root mean square deviation (RMSD) after each cycle of refinement (RMSD between the previous and the new refined structures) were 2.01, 0.58, 0.15 and 0.02 Å successively.

Simultaneously the calculated and experimental interproton distance RMSD between two successive steps was 0.25, 0.15, 0.13 and 0.13 Å.

**DISCUSSION**

Our data unambiguously shows that in aqueous solution, psor-m6-d(TAAGCCG) binds to d(CGGCTTA) or to d(CGGCTTA) and H6(T2)–CH2(T2). The protons of the hexamethylene linker were found not equivalent, indicating a defined conformation of the linker. The CH2(1) and CH2(6) proton resonances occur at 3.95, 4.05, 4.15 and 4.20 p.p.m.
Figure 6. Expansion of the imino to aromatic/amine protons region (upper) and the imino to imino protons region (lower) of the 2-D NOESY (mixing time = 150 ms) of the interstrand cross-linked 7-7 dimer. The spectrum was recorded in 90% H2O/10% D2O, in 0.1 M NaCl, pH 6.5, at 2°C. The thymine 2 imino proton (11.28 p.p.m.) exhibited a large upfield shift. f = free, b = bound.

methylene groups of the linker and thymine 1 (Table 1). A longer mixing time (200 ms) leads to appearance of weak cross peaks between the H8 or H3 protons of psoralen and protons of adenine 1 and 2 and of thymine 2 [H3(pso)−H2(A1), H8(pso)−H3′(A1), H8(pso)−H1′(T2), H8(pso)−H2(A2)], as well as between the methylene groups of the linker and the sugar protons of thymine 1. These data suggest the existence of an equilibrium between two conformational families (Fig. 7) where the psoralen moiety is stacked on the A1-T1 base pair or intercalated between the A1-T1 and A2-T2 terminal base pairs. It is worth noting that the intercalation of the psoralen moiety must increase the distance between the two base pairs A1-T1 and A2-T2. Conversely, despite the expected spacing of the base pairs, the 70 ms mixing time nOe values deduced from the volume measurement of H1′(T2)−H8(A1) and H1′(T1)−H8(A2) cross peaks are only slightly weaker than other inter-residue nOe(s) H1′−H6 or H8. This indicates an almost identical average distance between all the base pairs. In addition, the weakness of nOe connectivities between the aromatic protons of psoralen and the protons of base pair T2−A2 suggests that psoralen is more favorably stacked on the A1−T1 base pair rather than intercalated between base pairs A1−T1 and A2−T2.

In the 7-8mer, the existence of strong nOe connectivities between psoralen and adenine 1 as well as thymine 2, shows that the location of psoralen is different than in 7-7mer. Moreover, the presence of medium or weak connectivities with thymine T0 suggests the presence of an equilibrium between two conformations where psoralen is intercalated, with a higher probability between base pairs A1-T1 and A2-T2 than between base pair A1-T1 and thymine T0.

The nOe effects observed between protons of the methylene spacer arm and the protons H6, H5′, H5″ of T1 involve the location of the hexamethylene arm attached to the C-5 position of the psoralen moiety in the major groove of the mini double helix. On the basis of the evidence of a high yield (>80%) of photo-cross-linking between the two complementary sequences upon irradiation, we conclude that in the case of the 7-7mer, formation of the photoproduct between the intercalated psoralen and thymine T2 displaces the equilibrium from the stacked conformation (II) to the intercalated one (I). The FPLC analysis of the photoproducts only shows 8% of the free heptamer d(CGGCTTA). We conclude that the cross-link between the stacked psoralen and thymine T1 in conformation II is negligible.

For the 7-7mer, we obtain a single major diadduct where cycloadditions have taken place between the 5–6 double bond of T1 and the 3–4 (pyrone) double bond of psoralen, and between the 5–6 double bond of T2 and the 4′–5′ (furan) double bond of psoralen. The stereochemistry of the diadduct is cis-syn at both cyclobutane junctions. Figure 8 shows the structure of this interstrand photo-cross-linked DNA. The stacking and base pairing pattern of base pairs 3–7 is unmodified with respect to a free DNA with the same sequence. Deformations are located exclusively in the last two base pairs, that present a rise of 8.8 Å due to the insertion of psoralen. The base pairing of T2–A2 is deformed and weakened, but conserved, while base pairing of A1–T1 only conserves one hydrogen bond. The adenines stay...
Figure 8. Stereoview of the interstrand cross-linked 7-7 dimer. Cycloadditions have taken place between the 5-6 double bond of the thymine 1 and the 3-4 (pyrone) double bond of psoralen, and between the 5-6 double bond of the thymine 2 and the 4'-5' (furan) double bond of psoralen. The stereochemistry of the diadduct is cis-syn at both cyclobutane junctions.

roughly parallel to each other and to the plane of psoralen, while the modified thymines undergo important angular deformations, especially T1 which is situated at the extremity. No unwinding perturbation is observed, as opposed to the results obtained by Tomic et al., also from NMR data (27). This is probably due to the location of psoralen at the extremity of the oligonucleotide in our case. All the sugars in the structure are in south conformation, as already noticed.

The surprising chemical shift of the H3 proton of psoralen at 1.85 p.p.m. can be explained by its spatial position in the structure. Indeed, the vertical distance between the A2 purine base plane and the H3 proton of psoralen is 3.1 Å. The horizontal distances (the radial distance from the center of the ring) are 2.2 and 1.6 Å for the six- and five-membered rings respectively. These results are consistent with a possible shift around 1 p.p.m. found by using the spatial dependence of the ring-current magnetic anisotropy of nucleic acid bases calculated by Giessner-Prettre et al. (36). Taking into account a chemical shift at 2.86 p.p.m. for the H3 proton of the bergamottin photoproduct reported by Martin et al. (37), we conclude that the low value of this chemical shift at 1.85 p.p.m. is due to ring-current effects.

NMR analysis of diadducts formed by several psoralen molecules with DNA has revealed that the cross-link between the two strands of DNA involves two thymines, one on each strand of DNA at 5'TpA3' sequences (27,29-31). Psoralens have previously been attached via their C-8 (38) or their C-4' (39,40) position to oligonucleotides targeted to single stranded nucleic acid sequences. Here, we have shown that an oligonucleotide covalently linked to a psoralen derivative via a six methylene arm at its C-5 position induces photo-cross-link with a targeted DNA sequence. Such a system could be used to selectively control gene expression or to induce site directed mutations.

ACKNOWLEDGEMENTS

We wish to thank Dr Hélène (Paris) for helpful discussion. We would like to thank R. Lavery for providing JUMNA software. OB acknowledges financial support from the Ministère de l'Enseignement Supérieur et de la Recherche (MESR).

REFERENCES