Covalent bonding of bridged pyridinium aldehyde derivatives with guanine N7 is controlled by CpG site conformation †

Laurence Marcourt,[a] Christine Cordier,*[a] Gildas Bertho,[b] Jean-Pierre Girault,[b] Odile Convert[c] and Guy Dodin[a]

[a] *ITODYS, Laboratoire associé au CNRS, UPESA 7086, Université D. Diderot (Paris 7), 1 Rue Guy de la Brosse, Paris, 75005, France. E-mail: cordier@itodys.jussieu.fr; Fax: +33 (0) 1 44 27 68 37*

[b] *Laboratoire de chimie et biochimie pharmacologiques et toxicologiques CNRS, UMR 8601, Université René Descartes (Paris 5), 45 Rue des Saint-Pères, 75270 Paris Cedex 06, France*

[c] *Laboratoire de chimie structurale organique et biologique, UMR 7613, Université Pierre et Marie Curie (Paris 6), 4 Place Jussieu, 75005 Paris, France*

Received (in Cambridge, UK) 4th May 2001, Accepted 21st June 2001
First published as an Advance Article on the web 16th August 2001

The ability of $BPAC_8$, a member of the bis-pyridinium aldehyde (BPA) family with a linear octamethylene chain linking the two charged pyridinium moieties, to covalently bond with a guanine residue at a 5′-CpG-3′ site is studied. Three oligomers including a central CpG step with different conformations are studied. By ^{1}H NMR spectroscopy and gel analysis it is established that the covalent reaction occurs for the decamer CRE, d(ATGACGTCAT), and to a lesser extent for the methylated dodecamer, d(GAAAAmeCGTTTTC), while there is no reaction for the sequence d(GAAAACGTTTTC). The ease of reaction with BPA is correlated with both the geometry of CpG and the malleability of the oligomer. When the CpG structure is stiffened by a rigid nucleotide environment, such as A–T tracts, the reaction with BPA is inhibited.

Introduction

Bridged bis-pyridinium aldehydes (BPA) belong to a class of recently developed DNA-interacting compounds endowed with biological properties.[1,2] Previous investigations based on the inhibition of restriction endonuclease scission suggest that BPA reacts at Watson–Crick C–G base pairs.[3] An NMR study of the reaction with the decamer $d(ATGACGTCAT)_2$ (referred to as CRE: cyclic AMP response element) has shown that the guanine in the central 5′-CpG-3′ dimer is the preferred binding site.[4] The BPA–DNA reaction proceeds by a two-step pathway involving the formation of a non-covalent complex followed by the covalent binding of one aldehyde group in BPA to the amino group (N7) of the guanine in the minor groove. Changing the nucleotide environment of the CpG by introducing a four-adenine 5′-flanking tract to generate the self-complementary oligomer $d(GAAAACGTTTTC)_2$ (named CG) is expected to stiffen the structure. In the dodecamer,[5] the main structural parameters of the central step are: high CpG twist; sugar phase of the guanine about 180°; pronounced negative roll; significant narrowing of the minor groove at this point. In addition, the proximity of tracts rich in A–T induces unusually high propeller twists. In the CRE decamer,[5] CpG is a malleable site. Its associated conformational parameters differ markedly from those described above. These results show that the structure of the CpG step is not entirely determined by the immediate environment (here the same ACGT tetrad for both DNA fragments) but also depends on the distal upstream and downstream sequences.

In order to relieve the rigidity imposed by the A–T tracts, the central cytosine of $d(GAAAACGTTTTC)_2$ has been methylated at the C5 position. As a consequence of steric crowding by the methyls, the conformation of the central dimer in $d(GAAAAme^5CGTTTTC)_2$ (named meCG) is reorganized.[5,6] For example, an increase of rise and roll, a decrease in the me^5CpG twist and a more pronounced narrowing of the minor groove are observed. These concerted modifications of the geometry make the helix more flexible.

In this study we address the question of determining how the reaction with BPA is affected by structural modifications of the oligomers CG and meCG brought about by varying the neighbourhood of their CpG steps with respect to CRE. In the BPA series, the member which contains a $(CH_2)_8$ chain bridging the two cationic bis-pyridinium moieties was chosen. It has been shown that this length of the aliphatic chain deeply affects the supercoiling of the pBr322 plasmid while the shorter chains perturb the DNA supercoiling to a lesser extent.[3] $BPAC_8$-oligomer interactions are investigated by means of gel analysis and ^{1}H NMR spectroscopy. A relationship between the reaction of $BPAC_8$ with DNAs and their structure–malleability is established and discussed.

Experimental

Materials

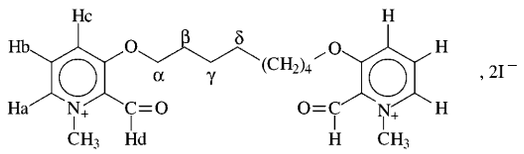
Both oligonucleotides, d(GAAAACGTTTTC) and $d(GAAAAme^5CGTTTTC)$, were provided by Dr Huynh Din (Institut Pasteur, Paris) and supplied as 11 ammoniums per strand. $BPAC_8$ was prepared and purified by published procedures.[7] Numbering of self-complementary oligomers and $BPAC_8$ is shown in Scheme 1.

† Electronic supplementary information (ESI) available: Fig. S1, binding of the CG, meCG and CRE with BPA; Fig. S2, 500 MHz 2D NOESY of $BPAC_8$–meCG mixture. See http://www.rsc.org/suppdata/p2/b1/b104022g

CG:
5' – G1 A2 A3 A4 A5 C6 G7 T8 T9 T10 T11 C12 – 3'
3' – C12 T11 T10 T9 T8 G7 C6 A5 A4 A3 A2 G1 – 5'

meCG:
5' – G1 A2 A3 A4 A5 me^{5}C6 G7 T8 T9 T10 T11 C12 – 3'
3' – C12 T11 T10 T9 T8 G7 me^{5}C6 A5 A4 A3 A2 G1 – 5'

CRE:
5' – A1 T2 G3 A4 C5 G6 T7 C8 A9 T10 – 3'
3' – T10 A9 C8 T7 G6 C5 A4 G3 T2 A1 – 5'



Scheme 1

The melting temperature (T_m) of each oligonucleotide was measured by NMR and is: 316 ± 2 K for CG, 322 ± 2 K for meCG and 319 ± 2 K for CRE. For these measurements, the DNA concentrations were: 5.87, 6.71 and 4.05 mM, respectively. The pH and the ionic strength were fixed at 7.4 and 50 mM in a phosphate buffer (KH_2PO_4 + Na_2HPO_4). Both medium and concentration were chosen to favour the duplex formation.

NMR methods

NMR samples were prepared by dissolving the lyophilized dodecamers (4 mg for CG and 4.5 mg for meCG) in 450 μL of a H_2O–D_2O (90 : 10) solvent mixture containing a phosphate buffer (pH = 7.4, I = 50 mM + 1 mM EDTA). ^{1}H NMR spectra were recorded on Bruker AM-500 and AMX-500 spectrometers. In order to observe the imino protons and to slow down the exchange processes, FIDs were collected at low temperature. Specifically, the temperature was set at 283 K and 280 K for CG and meCG, respectively, in order to minimize imino signal overlap. A jump and return experiment with low-power water presaturation was used to record 1D spectra on the AM-500 spectrometer. The delay between the two 90° pulses was fixed at 95 μs to give the principal null at the transmitter frequency placed at the water resonance frequency and the next null outside the spectral range (10.4 kHz). For 1D spectra performed with the AMX-500 spectrometer, the water line was suppressed by a WATERGATE field gradient experiment using a 3–9–19 binomial sequence.[8]

Titration. Each dodecamer was mixed with 0.5–2 equivalents of $BPAC_8$ per double strand. The 1D spectrum was recorded after each 0.5 equiv. increment of ligand.

NOESY. 2D-NOESY experiments were acquired in the States-TPPI mode[9] on the AMX-500 spectrometer with a WATERGATE technique (95 μs delay).

For the $BPAC_8$–CG (2 : 1) mixture, a set of NOESYs was recorded (100, 150 and 200 ms mixing time) with 4096 points in t_2 over 20 ppm and 512 points in t_1. 96 scans per experiment and a relaxation delay of 1 s were applied. A second set of NOESYs was performed with the same acquisition parameters after the sample had been heated for 3 hours at 38 °C.

For the methylated duplex, a first 2D NOESY experiment was run for a 2 : 1 ratio of $BPAC_8$–meCG without sample heating. A second NOESY was obtained for a 4 : 1 ratio after the mixture had been heated at 38 °C for 5 hours. This latter was recorded on a Bruker DMX-500 spectrometer. Both spectra were acquired with 4096 points in t_2 and 400 points in t_1. The main acquisition parameters were: mixing time, 200 ms; relaxation delay, 1.8 s; delay associated with the 3–9–19 binomial sequence, 80 μs.

Data were processed on Silicon Graphics stations with the help of GIFA.[10,11] A Burg linear prediction method in t_1 led to a 1024 × 4096 data point matrix. A shift sinebell window function in t_2 and sine-squared function in t_1 were applied before Fourier transformation. The baseline was corrected using a polynomial function.

Gel electrophoresis

The decamer (CRE) or a dodecamer (CG or meCG) was treated with polynucleotide kinase (Roche) in an appropriate incubation buffer. T4 DNA ligase (Roche) was then introduced. The reaction mixture was subsequently treated with an equal volume of phenol–chloroform and the polymerized oligomer precipitated with ethanol in the presence of 0.3 M sodium acetate.

Plasmid pBr322 DNA and restriction endonucleases Psp1406 I (AA↓CGTT), Sfu I (TT↓CGAA), Mae II (A↓CGT) and Nde I (CA↓TATG), provided with appropriate digestion buffers, were purchased from Roche.

Agarose gel electrophoresis. Electrophoreses of plasmid DNA were run in 1% agarose gels with a tris-acetate EDTA (TAE) (pH 7.4, 40 mM tris-acetate) running buffer at 25 °C. Gels were stained with ethidium bromide.

Polyacrylamide gel electrophoresis. 13% to 18% polyacrylamide gels were used in the electrophoresis of single or polymerized oligonucleotides with a tris-borate, 1 mM EDTA (TBE) (pH 8.0, 45 mM tris-borate) running buffer at 25 °C.

Results

$BPAC_8$–CRE

The interaction between $BPAC_8$ and the self-complementary decamer d(ATGACGTCAT) (CRE) has been previously studied by gel analysis and NMR spectroscopy.[4] In the present work, the NMR data are re-analysed to facilitate comparison with the results reported here for the $BPAC_8$–dodecamer interactions (*vide infra*). Re-investigation of the earlier data confirms that $BPAC_8$ reacts with CRE *via* a two-step pathway.

A fast non-covalent interaction occurs at room temperature. 2D NOESY spectra recorded for a $BPAC_8$–CRE ratio of 2 : 1 show negative intra-$BPAC_8$ as well as inter-$BPAC_8$–CRE dipolar correlations, which supports the formation of a complex. Since the intra-ligand NOEs are positive for the free $BPAC_8$, the inversion of the NOE phase indicates that the $BPAC_8$ correlation time is modified by the presence of DNA. No new resonance is detected in NOESY spectra at 283 K (H_2O–D_2O 90 : 10) and 298 K (D_2O), which means that the duplex remains self-complementary. The invariance of the chemical shifts of free and bonded cytosine amino protons demonstrates that the C–G base pairing is not modified. Small frequency shifts of the H2A1/A4, H8G6, H6T7, H6T10 signals (and to a lesser extent H8G3 and H6C5) are observed. This information indicates that the B-type conformation of DNA is not altered substantially by $BPAC_8$.

The gradual variation of chemical shifts, upon titration to equivalence, is typical of fast exchange (1D NMR experiments). The two-fold degeneracy for both the ligand and the DNA is seen to be retained at each point of the titration. Either the ligand binds to a single centrosymmetric site or it exchanges rapidly between two symmetrically equivalent binding sites of the DNA host. $BPAC_8$ can span four base pairs (bp) without any unfavourable stretching. In the first case, the set of chemical shift perturbations suggests that the central A4–T7 region is the binding location and that there is a fast exchange process between the free and complexed forms. In the second case, the

Table 1 $BPAC_8$–CRE connectivities for the complex at 283 K (H_2O–D_2O) and 298 K (D_2O) and for the adduct at 283 K (H_2O–D_2O)

Free $BPAC_8$ protons	$BPAC_8$–CRE complex T = 283 K	$BPAC_8$–CRE complex T = 298 K	$BPAC_8$–CRE adduct T = 283 K
Ha	H4′C5	—	H4′C5
	—	—	H4′C8a
Hb	—	H4′T2 or H4′C8	—
	—	—	H1′G3
	—	H4′G3 or H4′C5	—
	—	—	H2′/H2″G6a
	NH_2C8	—	—
	—	H4′T7	—
Hc	H6C5	—	—
	H4′C5	—	H4′C5
	—	—	H1′C5
	—	H1′C8	—
	H4′C8	H4′C8	H4′C8
	—	—	NH_2C8
	—	H4′T7	—
Hd	—	—	H4′T2
	—	H1′C5	—
	—	—	H4′C5
	H5C8	—	H5C8
	H4′C8	—	H4′C8
CH_3	—	—	H8G6
	—	H6T7	—
	H8A9	H8A9	—
$CH_2\alpha$	H8G3	H8G3	—
	—	—	H8A4
	H6C5	—	—
	—	—	H4′C5
	H8G6	H8G6	—
	H6C8	—	H6C8a
	NH_2C8	—	NH_2C8
	H8A9	H8A9	—
	—	H6T10	—
$CH_2\beta$	—	H6T7	—
$CH_2\gamma,\delta$	—	H1′G6	H8G6a
	H2A9	—	—

ligand would exchange between two overlapping sites (G3–G6 and C5–C8) each being asymmetrically disposed about the dyad axis of the duplex.

Intermolecular NOE contacts are mainly located on the C–G base pairs (both C5–G6 and C8–G3), as listed in Table 1. The DNA protons concerned indicate that both grooves are binding sites for $BPAC_8$. A model with a single binding site in the major DNA groove is not consistent with the full set of NOE contacts. Consequently, we consider that exchange occurs between two symmetrically equivalent binding sites in the major groove. NOEs place one pyridinium moiety facing the C8–G3 base pair, deep in the major groove, in good agreement with dipolar correlations between the {H8G3, NH_2C8, H5C8, H6C8} set and the aromatic protons of $BPAC_8$. The other pyridinium is close to the backbone, near the CpG step, in dipolar contact with H8G6 and H6C5. This arrangement corresponds to a binding site spanning 3.5 bp (Fig. 1). For the minor groove alternative, because of strong overlap between $BPAC_8$ and some DNA signals, the number of unequivocal NOEs is too small to allow a detailed description of ligand insertion in this groove.

There are small but significant differences between spectra recorded at 283 and 298 K. At 298 K is seen: i) broadening of specific signals (H5/H6 protons of cytosines, H6/H1′/CH_3T7 and G6 sugar protons) accompanied by a decrease in the intranucleotide NOE crosspeaks between aromatic and sugar protons for C5 and G6, ii) loss of intermolecular crosspeaks between $BPAC_8$ and H5/H6C8 or H6C5, probably due to broadening, and iii) appearance of new inter-$BPAC_8$–CRE correlations concerning mainly the protons of the minor groove (Table 1). This line broadening may result from a

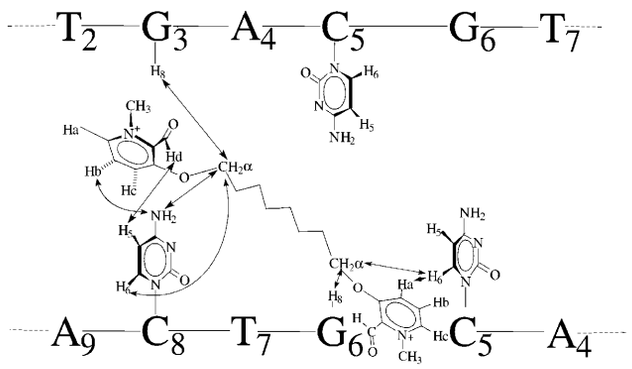


Fig. 1 Schematic view of one of the two possible symmetric binding sites of $BPAC_8$ in the major groove of CRE. Only the relevant part of the CRE nucleotidic chain is shown. The intermolecular dipolar contacts are shown as arrows.

heterogeneous chemical environment caused by the $BPAC_8$ interacting with different regions of the minor groove when the temperature is raised.

We reject the idea that $BPAC_8$ is intercalated, because the NOE intensity is low and there are no breaks or irregularities in the aromatic-H1′DNA walk, which would be expected for such a binding mode.[12] Further, the complex formed is easily dissociated when the mixture is treated with ethanol to precipitate the DNA.

In short, there is no preference for either the minor or the major groove of the DNA when the complex is formed, since intermolecular NOE contacts are observed with both grooves. However, orientation of $BPAC_8$ towards the minor groove is slightly favoured when the temperature increases.

A covalent $BPAC_8$-CRE adduct occurs at 38 °C. As previously described,[4] the adduct is only formed when the $BPAC_8$–CRE mixture is heated for at least 20 minutes at 38 °C. Such slow binding suggests that a covalent bond is formed. This is confirmed by the significant electrophoretic retardation which is observed for both the monomeric (Fig. S1 as electronic supplementary information) and polymeric CRE. The molecular association appears to be stable, insofar as it resists two precipitation–redissolution cycles without immediate dissociation. It was previously shown[4] that gel retardation of the adduct suggests a 1 : 1 $BPAC_8$–CRE stoichiometry. No other available experimental data invalidate this stoichiometry.

BPA alters gel migration of polyCRE and inhibits the restriction by Mae II (at the $A_4C_5G_6T_7$ site) and Nde I (at the neoformed $C_8A_9T_{10}A_1T_2G_3$ site) to apparently a similar extent for both enzymes (Fig. 2). The fact that the treated polyCRE is not cleaved by either Nde I or Mae II indicates that both G3 and G6 are covalent bonding sites. In another set of experiments, treatment of plasmid pBr322 with $BPAC_8$ over increasing periods of time gives rise to a progressive retardation of the electrophoretic migration which is accompanied by a decrease in the extent of cleavage at the ACGT (Mae II) and CATATG (Nde I) sites, the latter being however slightly less sensitive to inhibition by $BPAC_8$. This is consistent with previous results showing protection against endonuclease action at C–G base pairs (Fig. 3).[3]

This adduct formation is revealed in NMR by the following spectral changes. a) Appearance in the NOESY spectrum of many new correlation lines of protons that resonate around 10 ppm and in the imino and aromatic regions. The intensities of the new lines at 13.5, 12.7 and around 10 ppm are only slightly reduced when the NMR experiment temperature rises, whereas the imino lines of the free DNA disappear under the same conditions. b) Dipolar correlations between four new imino groups at 13.5 ppm and four new peaks at 10 ppm (Fig. 4).

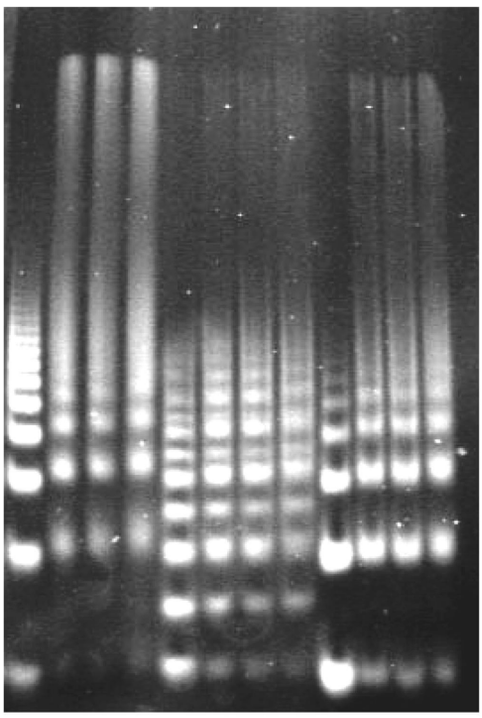


Fig. 2 Binding of $BPAC_8$ to polyCRE and its inhibitory effect on restriction by Mae II and Nde I. 13% polyacrylamide; TBE running buffer. 4 nmol polyCRE in tris-EDTA, 10 mM; pH 7.4, [BPA]/[polyCRE] = 4 and 1 unit Mae II or Nde I in appropriate buffer. Lane 1: without BPA; lanes 2–4: with BPA; t_i = 1, 2, 5 h at 38 °C; lane 5: without BPA, +Mae II; lanes 6–8: with BPA, t_i = 1, 2, 5 h, +Mae II; lane 9: without BPA, +Nde I; lanes 10–12: with BPA, t_i = 1, 2, 5 h, + Nde I.

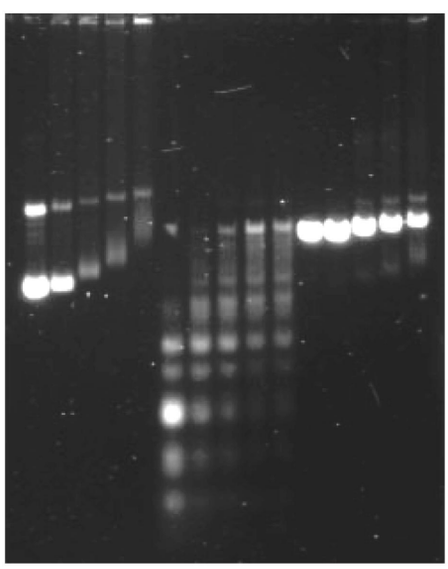


Fig. 3 Binding of $BPAC_8$ to pBr322 and its inhibitory effect on restriction by Mae II and Nde I. 1% agarose gel; TAE running buffer. Lane 1: control, 4 nmol DNA in tris-EDTA, 10 mM, pH 7.4; lanes 2–5: [BPA]/[DNA] = 4; incubation time t_i = 0, 1, 2, 5 h at 38 °C; lane 6: 4 nmol DNA + 1 unit Mae II in appropriate buffer, 1 h at 40 °C (= conditions 1); lanes 7–10: [BPA]/[DNa] = 4, t_i = 0, 1, 2, 5 h, Mae II digestion under conditions 1; lane 11: 4 nmol DNA + 1 unit Nde I in appropriate buffer, 1 h at 40 °C (= conditions 2); lanes 12–15: [BPA]/[DNA] = 4, t_i = 0, 1, 2, 5 h; Nde I restriction under conditions 2.

c) Duplication of many resonances in the aromatic proton region. d) Additional inter-$BPAC_8$–CRE NOE contacts which involve protons mostly located in the minor groove of CRE (Table 1).

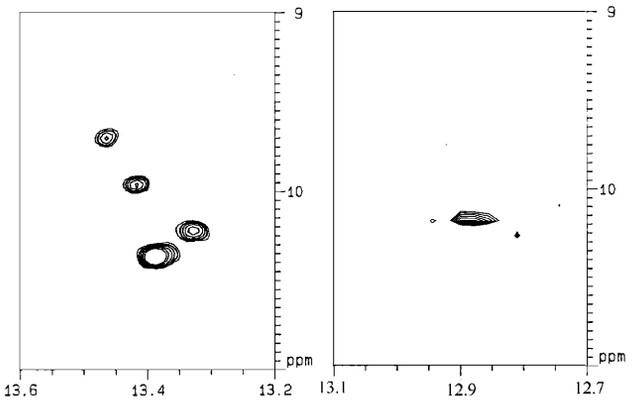


Fig. 4 Dipolar correlations between imino protons and NH hemi-aminal peaks at ~10 ppm. 2D NOESY spectra were recorded in H_2O–D_2O (90 : 10) for 2 : 1 $BPAC_8$–CRE (left) and 4 : 1 $BPAC_8$–meCG mixtures (right).

These results confirm our earlier conclusion[4] that the adduct is formed by nucleophilic attack of the exocyclic guanine amino group (located in the minor groove of the helix) on the aldehyde moiety of $BPAC_8$ to give a hemi-aminal function. From their dipolar correlations with the new imino at 13.5 ppm, the lines at 10 ppm are assigned to the NH protons of the hemi-aminal groups. The appearance of four NH hemi-aminal resonances is evidence for multiple binding sites or/and epimers, since there are only two possible sites of linkage (NH_2G3 and NH_2G6). The relative intensities of the NOE crosspeaks in this region show that CpG is slightly preferred to G3.

Despite the fact that the $BPAC_8$–CRE mixture contains an excess of ligand, the yield of the covalent reaction is low. An attempt to perform a 2D NOESY experiment on the isolated adduct failed because the covalent reaction is reversible, the adduct disappears when it is left in solution for many hours and the 1D NMR proton spectrum of the free decamer is restored. Consequently, the NMR spectra are too complex to be analysed accurately; this makes it impossible to describe the position of the ligand within the DNA unambiguously.

$BPAC_8$–CG

A non-covalent $BPAC_8$–CG complex is formed immediately. At room temperature, $BPAC_8$ (in H_2O–D_2O 90 : 10) was added in steps of 0.5 equivalent to the CG dodecamer solution (H_2O–D_2O 90 : 10) up to a ligand–DNA duplex ratio of 2 : 1. The titration was monitored by means of 1D NMR spectra recorded at 283 K, the imino and aromatic proton regions being particularly examined. The addition of 2 equivalents of $BPAC_8$ causes only small frequency shifts in the 1D spectra; there are no other spectral changes (line disappearance or appearance and line broadening). The lack of additional new peaks which could be assigned to either the ligand or the DNA indicates fast exchange between the free and complexed forms and the retention of two-fold symmetry for both molecules. The most significant frequency shifts are observed for the protons of the adenine and thymine tracts: H2A3, H2A4, H2A5 and IT11, located in the minor groove of the helix (Fig. 5). Nevertheless, the chemical shift variation is small and does not exceed 0.05 ppm.

No disruption of the H1′(*n*)–H6/H8(*n*)–H1′(*n* + 1) sequential connectivity is seen in the NOESY spectrum recorded on the 2 : 1 mixture of $BPAC_8$–CG. The relative intensities of the intramolecular-DNA crosspeaks are identical to those observed for the free dodecamer; this establishes that the conformation of the DNA is not significantly different from what it was in the absence of BPA. As noted for the $BPAC_8$–CRE complex, the negative intramolecular-$BPAC_8$ NOEs are the result of an increase in the correlation time of the ligand when it binds to

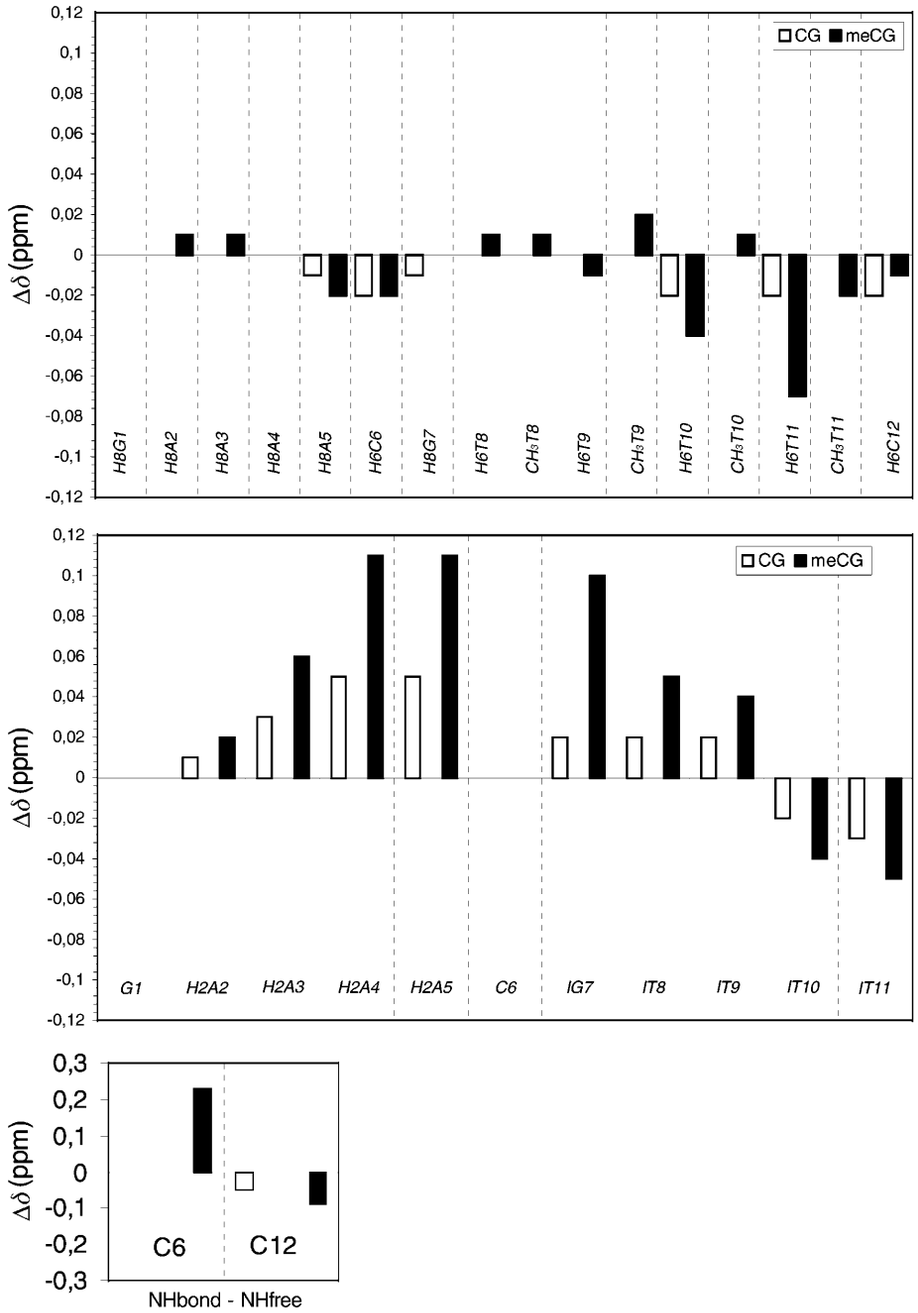


Fig. 5 The difference, δ(BPAC$_8$–DNA complex) − δ(free DNA), is given for protons located in the major (top) and minor grooves (middle). White columns represent the CG protons and black columns the meCG protons. For the amino protons (lower), the change in δNHb − δNH on going from free DNA to complexed is considered.

CG. Moreover, weak NOE crosspeaks between the two molecules confirm that a complex is formed. These NOE contacts concern mainly H2A4 and H2A5 with the full BPAC$_8$ proton set (Table 2). The asymmetry of the A2–A5 binding site relative to the dyad axis and the retention of C_2-symmetry imply that BPAC$_8$ exchanges rapidly between the two homologous A–T tracts of the helix. The intermolecular contacts position the ligand deep in the minor groove, as suggested by NOEs with the H2 and H1′ protons of the DNA host. This suggests that one BPA pyridinium faces the A2–T11 and A3–T10 bp and that the other is close to the A4–T9 and A5–T8 bp, this second cationic group being more deeply docked in the groove (more NOEs).

These experimental results exclude covalent linkage of the BPA with DNA as well as its intercalation through the DNA, since its presence causes only small chemical shift variations and no irregularity in the sequential connectivity. At room temperature, BPAC$_8$ interacts with CG as it does with CRE. The minor groove is only concerned when BPAC$_8$ docks to CG, whereas neither groove is favoured for the location of BPAC$_8$ in CRE.

The covalent reaction is inhibited. In order to form a covalent BPAC$_8$–CG linkage, the 2 : 1 BPAC$_8$–CG mixture was incubated at 38 °C for 48 hours. Advancement of the reaction was monitored at various time intervals (2 h, 3 h, 24 h and 48 h) by 2D NMR NOESY experiments recorded in H_2O–D_2O (90 : 10). Even after the longest incubation time, no spectral changes are detected, *i.e.* i) no frequency shift and ii) no new resonance line, neither in the 9–11 ppm and imino regions (12.5–14.5 ppm) nor in the non-exchangeable proton regions. No covalent reaction between BPAC$_8$ and CG occurs when the temperature increases. This is confirmed by study of polyCG scission at the AACGTT and TTCGAA sites by restriction endonucleases.

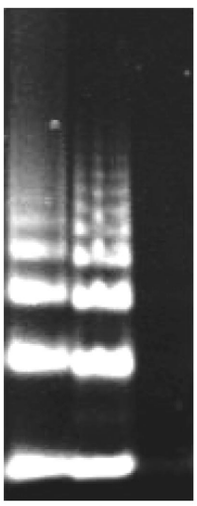


Fig. 6 Scission of polyCG by Psp1406 I and Sfu I restriction endonucleases. 13% polyacrylamide; TBE running buffer. 4 nmol polyCG in tris-EDTA, 10 mM; pH 7.4 and 1 unit Psp1406 I or Sfu I in appropriate buffer. Lane 1: polyCG; lane 2: polyCG + Psp1406 I; lane 3: polyCG + Sfu I.

The neo-formed TTCGAA site created in the polymerization of the CG dodecamer is readily recognized by the restriction enzyme Sfu I (Fig. 6). However, neither the monomeric nor the polymeric CG dodecamers are cleaved by Psp1406 I although, under similar conditions, plasmid pBr322 undergoes scission at its four Psp1406 I sites. Since none of the AACGTT sites of pBr322 is embedded in such the large AAAACGTTTT sequence as in CG and polyCG, we postulate that the two $(A–T)_4$ tracts prevent the enzyme cleaving the target sequence. This postulate is in agreement with the fact that $BPAC_8$ does not react to a detectable extent with the dodecamer CG, whereas it binds to CRE (Fig. S1). This lack of reactivity is tentatively assigned to the marked stiffness conferred to the CpG step by its nucleotide environment in CG.

$BPAC_8$–meCG

A fast non-covalent interaction occurs at room temperature. 1D spectra ($H_2O–D_2O$ 90 : 10 at 280 K) were recorded at each step of the titration of meCG by $BPAC_8$ (0.5 $BPAC_8$–meCG equivalent per addition). For the final 2 : 1 mixture, some peaks are particularly shifted (>0.1 ppm downfield). The protons concerned are located in the minor groove of the DNA, both within the adenine tracts (H2A4 and H2A5) and at the CpG step (IG7). The thymine tract proton signals are less shifted. The chemical shifts of these protons depend on their location relative to the central CpG step: proton signals of thymines close to CpG move downfield (IT8 and IT9) whereas IT10, H6T10, IT11 and H6T11 move upfield (Fig. 5). No new lines are detected. As described for the $BPAC_8$–CG complex, a rapid exchange of the ligand between two symmetrical binding sites in the minor groove explains why the C_2-symmetry of the DNA and $BPAC_8$ is retained.

The NOESY spectrum of the same sample (2 : 1 mixture of $BPAC_8$ and meCG) exhibits: i) negative intra-ligand NOEs, ii) intermolecular dipolar contacts involving mainly the adenine and thymine tract protons (Table 2), and iii) full sequential connectivity (Fig. 7). The relative intensities of the intermolecular NOE crosspeaks are higher for the $BPAC_8$–meCG complex than for the non-methylated complex. The binding site is similar to that described for CG, it is extended to three bp (A3–T10 to A5–T8) as shown on Fig. 8. The chemical shift difference between the free and bonded amino protons of C6 ($\delta_{NHbC6} - \delta_{NHC6}$) decreases upon $BPAC_8$ addition, indicating slight C6–G7 base pair opening. This observation, taken with the downfield shift of H2A5 and IG7, suggests that the conformation of the C6–G7 bp is altered. These results agree with the fast formation of a non-covalent complex, as reported for the oligonucleotides studied above. NOE contacts indicate that the $BPAC_8$ is in the minor groove opposite the adenine tracts, similar to the location of the ligand within CG.

Table 2 Connectivities for the $BPAC_8$–CG complex (283 K), the $BPAC_8$–meCG complex and adduct (280 K). These dipolar contacts are observed in $H_2O–D_2O$

Free $BPAC_8$ protons	$BPAC_8$–CG complex T = 283 K	$BPAC_8$–meCG complex T = 280 K	$BPAC_8$–meCG adduct T = 283 K
Ha	H2A2	—	—
	H2A3	H2A3	H2A3
	H2A4	H2A4	H2A4
	H2A5	H2A5	—
Hb	—	H2A3	—
	H2A4	H2A4	—
	H2A5	H2A5	—
		H1′T9 or/and T10	H1′T8/T9/T10/T11
Hc	—	H2A3	—
	H2A4	H2A4	—
	H2A5	H2A5	—
Hd	—	H1′A4 or G7	—
	H2′/H2″G7		—
	—	H1′T9 or/and T10	—
CH_3	—	—	H2A2
	—	H2A3	H2A3
	H2A4	H2A4	H2A4/A4′
	—	H2A5	H2A5/A5′
$Ch_2\alpha$	—	—	H2A2
	H2A3	H2A3	H2A3
	H2A4	H2A4	H2A4/A4′
	H2A5	H2A5	H2A5
$CH_2\beta$	H2A3	—	—
	H2A4	—	—
$CH_2\gamma,\delta$	—	—	—

CG methylation enables covalent reaction with $BPAC_8$. Electrophoretic migration of meCG (monomeric species) was studied after the $BPAC_8$–meCG mixture had been incubated from 0 to 8 hours. The slight retardation observed suggests that a covalent bond links the two molecules (Fig. S1). Cleavage experiments by a selective restriction enzyme cannot be performed on polymerized meCG because no endonuclease is able to recognize a methylated cytosine residue in the nucleotide sequence AmeCGT or AAmeCGTT.

After the 2 : 1 $BPAC_8$–meCG mixture had been incubated for 2 hours at 38 °C, a 1D NMR control experiment was performed ($H_2O–D_2O$ 90 : 10 at 280 K). New small peaks are detected in the 9–10.5 and 12.5–14 ppm regions. Their intensities did not increase significantly when the mixture was heated for 24 hours and/or 2 further equivalents of $BPAC_8$ were added.

A NOESY spectrum was recorded ($H_2O–D_2O$ 90 : 10 at 280 K) for the latter mixture after standing at 38 °C for 5 hours. Many new lines of connectivities are observed over the entire spectrum. In particular, the same type of correlation as for the $BPAC_8$–CRE adduct is detected between one new imino close to 13 ppm and one new peak around 10 ppm (Fig. 4). This suggests the formation of a covalent $BPAC_8$–meCG adduct which leads to the loss of the DNA C_2-symmetry. The location of the covalent bond at the G7 residue is consistent with the partial assignment of the new system (Fig. S2 as electronic supplementary information). A full assignment is ambiguous because the set of connectivities is incomplete and the new resonances overlap with those from the non-covalent complex. An attempt to isolate the covalent adduct by elimination of the excess $BPAC_8$ failed: the 1D NMR spectrum, run immediately after redissolution of the material

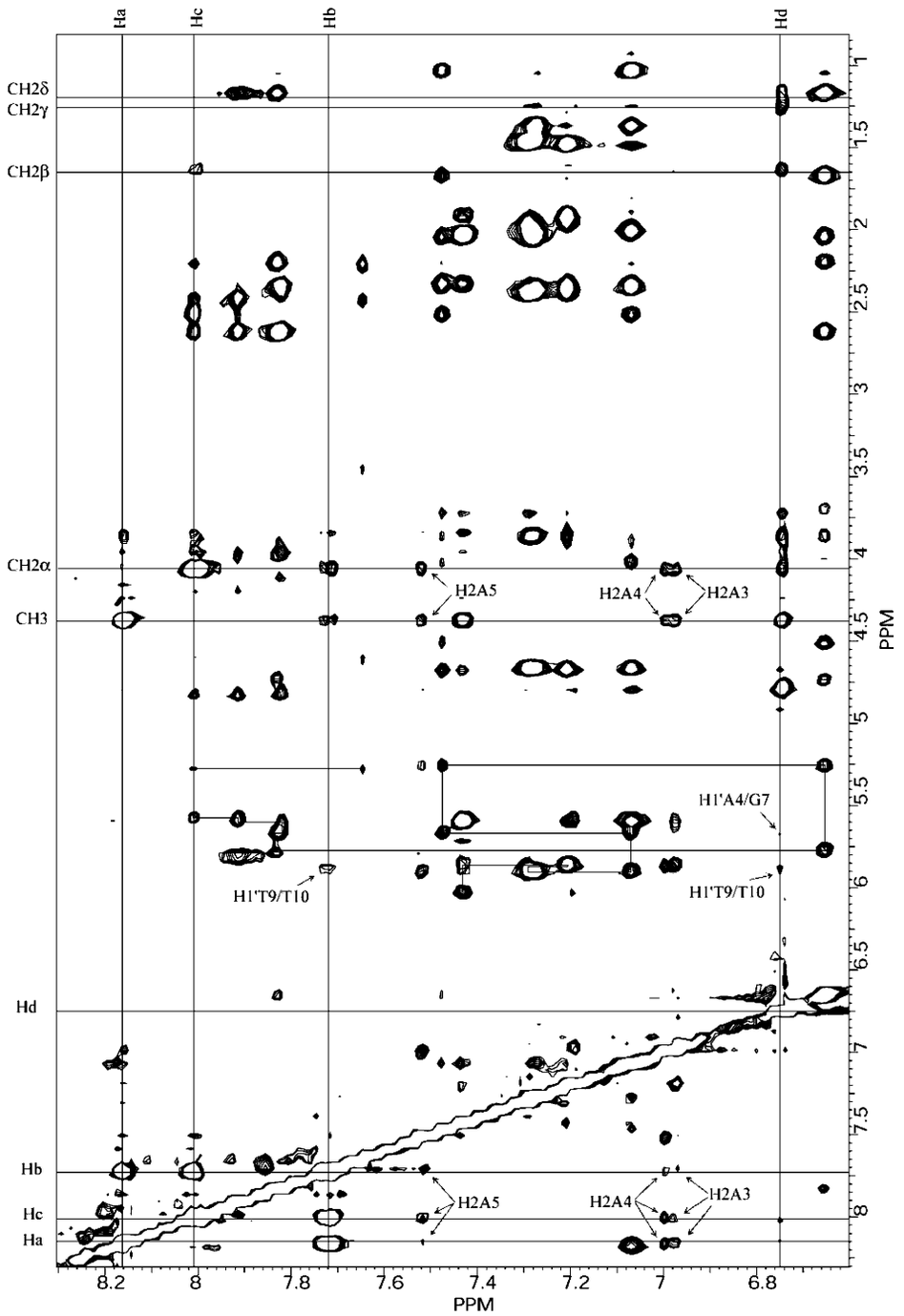


Fig. 7 500 MHz 2D NOESY of $BPAC_8$–meCG mixture (4 : 1) before heating, in H_2O–D_2O (90 : 10) at 280 K. F2 region: 6.6–8.3 ppm, F1 region: 0.9–8.3 ppm. The connectivities between $BPAC_8$ and non-exchangeable meCG protons are indicated with arrows and the resonance frequencies of $BPAC_8$ protons by solid lines. The aromatic-H1′DNA walk is also displayed.

precipitated from the $BPAC_8$–meCG mixture by ethanol, shows only free DNA.

Discussion

The formation of the $BPAC_8$–DNA complex is controlled by electrostatic interactions

In a previous study, the influence of many factors such as the length of the chain linking the two pyridinium groups, the time of reaction, the drug–DNA ratio, the temperature, pH and ionic strength has been established by gel electrophoresis analysis.[3] We have also checked that the neutral and mono-cationic ligand species do not interact with either the calf thymus or the plasmid pBr322. In line with this we extrapolate this lack of interaction to the oligonucleotides. Again, the migration retardation of the plasmid increases as a mark of a BPA–DNA interaction, when the ionic strength of the electrophoresis running buffer decreases.

For the three oligomers, a non-covalent complex is formed rapidly at room temperature. Although the central tetramer (ACGT) is the same, the ligand is not located at the same point within the helix for each oligomer: either both grooves are concerned or many sites are involved in a single groove.

It is well known that a high negative electrostatic potential and a narrowing of the minor groove are characteristic of A–T tracts.[5,13,14] Beveridge *et al.* have reported that mobile

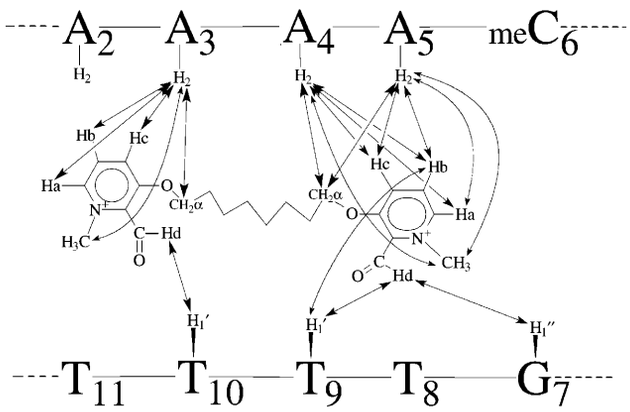


Fig. 8 Schematic view of one of the two possible symmetric binding sites of $BPAC_8$ in the minor groove of meCG. Only the relevant part of the meCG nucleotidic chain is shown. The intermolecular dipolar contacts are shown as arrows.

cations (Na^+ in the study mentioned) "may intrude on the minor groove spine of hydration" preferentially at the electronegative "AT pocket" and may induce conformational distortions of the DNA structure.[15] This $(A)_n$–$(T)_n$ chain behaves like a cation-trap on the minor groove. The dicationic $BPAC_8$ ligand would then be attracted towards these pockets. Indeed, the observed dipolar correlations indicate that $BPAC_8$ is in the minor groove at the $(A)_4$–$(T)_4$ tracts of CG and meCG.

For CRE, the schematic representation of the negative pockets along the double helix, as proposed by Beveridge *et al.*, does not show that there are strong electronegative sites in either the minor or the major groove (Fig. 9). This is consistent with the observation, based on the full NOE contacts, that $BPAC_8$ is located non-specifically.

The equilibrium constant (K) of the complex formation (Scheme 2) has been determined from the chemical shift vari-

$$\mathrm{DNA} + \mathrm{BPAC_8} \overset{K}{\rightleftharpoons} \mathrm{Complex}$$

$\delta = (1 - \alpha)\delta_0 + \alpha\delta_\infty$ where δ_0 is the chemical shift of the H2A proton (within ACGTstep) for the free DNA δ_∞ the chemical shift of the same proton for the complex (obtained after extrapolation) when the limit of the ratio $n\mathrm{BPAC_8}_0/n\mathrm{DNA}_0$ tends to $+\infty$ and α defines the fraction of the DNA that has reacted.

$$\alpha = \frac{\delta - \delta_0}{\delta_\infty - \delta_0} \text{ and } K = \frac{\alpha}{n\mathrm{DNA}_0 \times (1 - \alpha) \times \left(\frac{n\mathrm{BPAC_8}_0}{n\mathrm{DNA}_0} - \alpha\right)}$$

Scheme 2

ation of the H2A proton of the central ACGT upon titration of the DNAs by $BPAC_8$. This proton has been selected because i) it is located in the minor groove, ii) it belongs to the central tetramer ACGT present within the three oligonucleotides of interest, and iii) it is remote from the helix end-chains. This proton is a good probe to estimate the affinity of the ligand for the minor groove of each oligomer. The calculated values are: 1.1×10^6; 1.6×10^6 and 3.7×10^6 mol^{-1} (±15%) when the DNA is CRE, CG and meCG, respectively. The magnitude of the K values (~10^6 mol^{-1}) shows a marked affinity of this ligand for the minor groove of the double helix DNA. Now, a slight preference is observed for the minor grooves generated by the A–T tracts (K_{CG} or $K_{meCG} > K_{CRE}$). The presence of the methyls, *via* their "electron-donor" inductive character, seems to enhance the formation of a "minor groove complex". In short, the location of the BPA in the DNA to form the complex is guided by coulombic forces; this conclusion is in agreement with the high rate of this reaction.

Adduct formation depends on the malleability of the DNA fragment

We previously established that the CpG step conformation, in the same ACGT tetrad, depends on its distal nucleotide environment. Hence, the structure of the central CpG within CRE exhibits unusual features whereas the conformation of CpG in the CG dodecamer is normal. In addition, we proposed that the nucleotide chain of CRE favours the intrinsic malleability of this step.[5] In contrast, the presence of the A–T tracts severely reduces the degree of freedom of the central CpG.

In this work, the formation of a $BPAC_8$–DNA adduct and the recognition of DNA by appropriate endonucleases are observed when the DNA fragment is CRE, whereas neither is observed for the CG oligonucleotide. This difference in reactivity is interpreted as proposed above. The fact that the AACGTT hexamer is not cleaved by the Psp1406 I enzyme is evidence that CpG loses its malleability when it is flanked by four adenines and four thymines. In the same way, $BPAC_8$ does not link CG covalently. This suggests that adduct formation requires a malleable CpG site.

Methylation enhances malleability of CpG within a rigid environment

The effects of methylation on the C5 position of the cytosine ring occur at two different but, nevertheless, complementary levels:[6] a) a local variation of the structural parameters is observed at the ACGT tetrad; b) the grooves of the helix are reorganized. The induced structural modifications are a response to the steric requirements of the two methyl groups. The conformation of the CpG site, which in the non-methylated sequence is locked by its neighbouring A–T tracts, is perturbed by the alkyl groups. In this particular case, methylation induces significant structural changes. More specifically, the helicoidal conformation is considerably modified, the minor groove being severely pinched at the central C–G subunit. Not only do the cytosine methyl groups, in the major groove, not affect the location of electronegative pockets within the minor groove but also they enhance electrostatic interactions with cations because of the further narrowing of the minor groove. This proposal is experimentally supported by the observation of intermolecular NOE contacts stronger than those of the non-methylated helix. These NOEs still concern the protons of the A–T tracts. By NMR and gel analysis it has been established that when meCG is mixed and heated with $BPAC_8$ a covalent adduct is formed. Methylation *via* the induced structural variations makes the covalent reaction possible. We propose that the two methyls make the central CpG sufficiently malleable for the ligand to approach the exocyclic NH_2 of the guanine specifically. This preliminary step is essential to the ligand–DNA linkage.

Conclusion

This study confirms that the C–G base pair is the target of the bis-pyridinium aldehyde derivatives to give a covalent reaction. While the CpG dinucleotide is the preferred DNA motif for this reaction, if the conformation of this site is not right there is no covalent adduct. Our results clearly show that there is a close correlation between the CpG site geometry and the ease of reaction of $BPAC_8$ with this dimer or its recognition by specific endonucleases. CpG methylation is a biological event which is capable of changing the structure at this step. This work shows that $BPAC_8$ reacts covalently with the meCpG motif, whereas no similar reaction occurs with the non-methylated dimer within the 5′-A_4CGT_4-3′ chain.

The covalent reaction between $BPAC_8$ and the guanine residue is conditioned by: 1) access of the ligand to the minor groove *via* electrostatic forces (a dicationic ligand attracted by

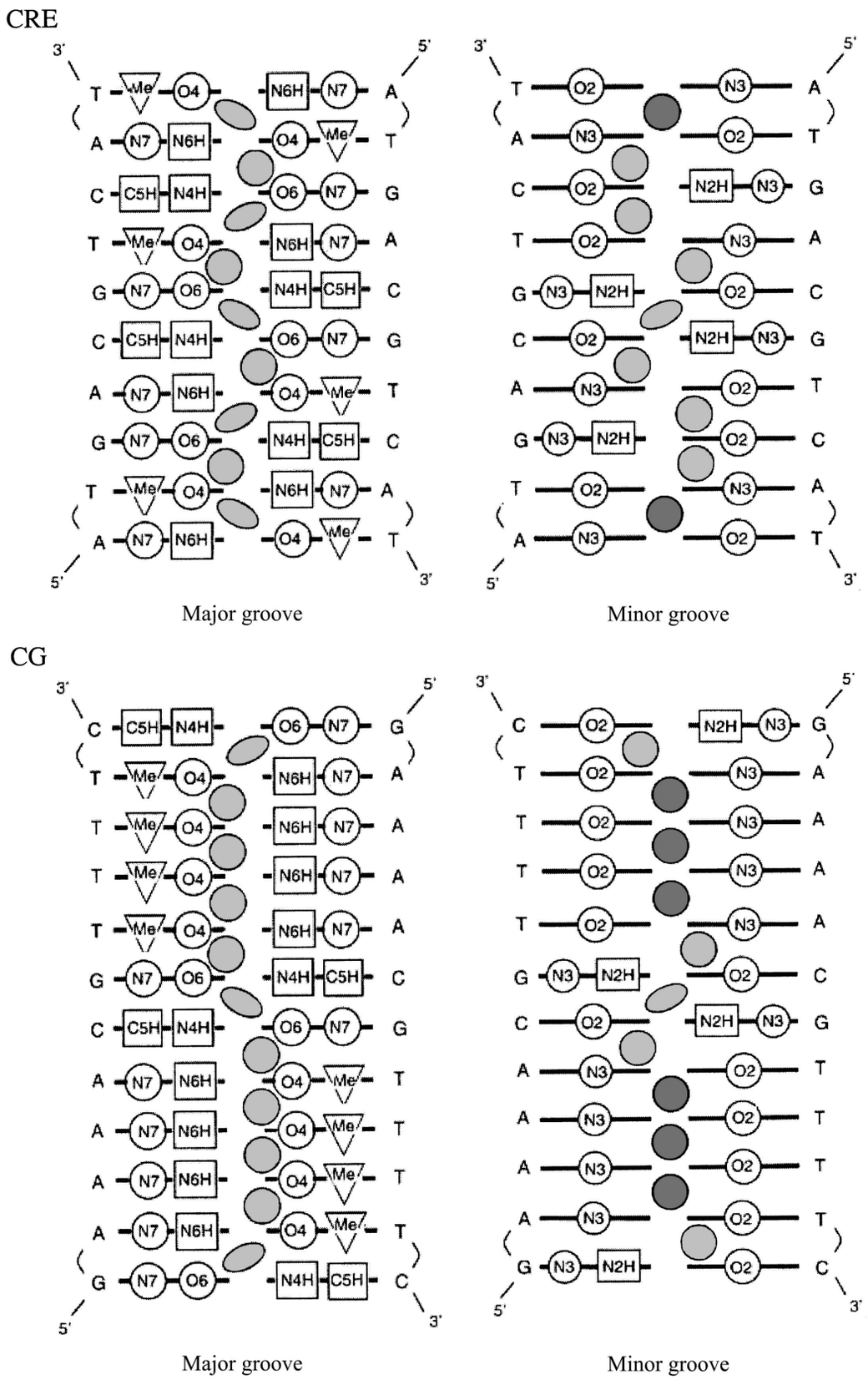


Fig. 9 Schematic views of CRE (top) and CG (bottom) base pairs along the helix, indicating the location of electrostatic pockets, as proposed by Beveridge *et al.*,[15] for major (left) and minor grooves (right). The shading is proportional to the electronegativity of the site. Convention: on the base pair edges, electronegative atoms are indicated by circles, electropositive hydrogen atoms by squares and hydrophobic methyl groups by triangles.

strong electronegative "pockets" such as A–T base pairs), and 2) the malleability of CpG which adopts a conformation adapted to the shape of the ligand.

$BPAC_8$ is interesting in that it forms a reversible covalent adduct with the oligonucleotides studied and is released without any damage to the DNA.

References

1 S. Nocentini, G. Moreno, P. X. Petit, M. Guggiari, C. Salet, P. Demerseman and G. Dodin, *Biochem. Pharmacol.*, 1997, **53**, 1543.
2 G. Dodin, in *Book of Abstracts, 207th ACS National Meeting*, (San Diego, 1994), ACS, Washington DC.
3 G. Dodin, J. M. Kühnel, P. Demerseman and J. Kotzyba, *Anti-Cancer Drug Des.*, 1993, **8**, 361.
4 C. Cordier, O. Convert, J. C. Blais, T. Couesnon, K. Zakrzewska, O. Mauffret, S. Fermandjian and G. Dodin, *J. Chem. Soc., Perkin Trans. 2*, 1998, 115.
5 C. Cordier, L. Marcourt, M. Petitjean and G. Dodin, *Eur. J. Biochem.*, 1999, **261**, 722.
6 L. Marcourt, C. Cordier, T. Couesnon and G. Dodin, *Eur. J. Biochem.*, 1999, **265**, 1032.
7 P. Demerseman, D. Kiffer, L. Debussche, C. Lion, R. Royer and H. Sentenac-Roumanou, *Eur. J. Med. Chem.*, 1988, **23**, 63.
8 M. Piotto, V. Saudek and V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661.
9 D. J. States, R. A. Haberkorn and D. J. Ruben, *J. Magn. Reson.*, 1982, **48**, 286.
10 M. A. Delsuc, in *Maximum Entropy and Bayesian Methods*, J. Skilling, ed., Kluwer Academic, Dordrecht, 1988, p. 285.
11 J. L. Pons, T. E. Malliavin and M. A. Delsuc, *J. Biomol. NMR*, 1996, **8**, 445.
12 M. Hansen and L. Hurley, *J. Am. Chem. Soc.*, 1995, **117**, 2421.
13 K. Yanagi, C. G. Privé and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 201.
14 C. G. Privé, K. Yanagi and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 177.
15 M. A. Young, B. Jayaram and D. L. Beveridge, *J. Am. Chem. Soc.*, 1997, **119**, 59.