Nucleic Acids Research, 1995, Vol. 23, No. 5 753–760

Site selective bis-intercalation of a homodimeric thiazole orange dye in DNA oligonucleotides

Jens Peter Jacobsen*, Jeanette B. Pedersen, Lene F. Hansen and David E. Wemmer[1]

Department of Chemistry, Odense University, Odense M, DK-5230, Denmark and [1]Structural Biology Division, Lawrence Berkeley Laboratory and Department of Chemistry, University of California, Berkeley, CA 94720, USA

Received November 24, 1994; Revised and Accepted January 10, 1995

ABSTRACT

We have used one and two dimensional ^{1}H NMR spectroscopy to characterize the binding of a homodimeric thiazole orange dye, 1,1′-(4,4,8,8-tetramethyl-4,8-diaza-undecamethylene)-bis-4-(3-methyl-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)-quinolinium tetraiodide (TOTO), to various double stranded DNA oligonucleotides. TOTO binds strongly to all the oligonucleotides used, but usually more than one complex is observed and exchange between different binding sites broadens the lines in the NMR spectra. Complete precipitation occurs when TOTO is bound to small oligonucleotides. Binding to larger oligonucleotides occurs by bis-intercalation. The 1:1 complex of TOTO with the oligonucleotide d(CCGACTGATGC):d(GCATCAGTCGG) gave only one complex that was shown to be a bis-intercalation in the CTGA:TCAG binding site. The binding to this site was also characterized by studying the TOTO complex with the d(CCGCTGAGC):d(GCTCAGCGG) oligonucleotide. NOE connectivities and molecular modelling were used to characterize the complex. The 1:1 complex of TOTO with the oligonucleotide d(CCGCTAGCG):d(CGCTAGCGG) containing a CTAG:CTAG binding site was similarly characterized by NMR. It was concluded that the binding of TOTO to larger oligonucleotides is site selective with CTAG:CTAG as the preferred binding site.

INTRODUCTION

A variety of natural products bind strongly to double stranded DNA (dsDNA) by intercalation. Clinical use of such compounds as antitumor agents has led to extensive search for new synthetic compounds with enhanced chemotherapeutic effect (1). Dimers with two intercalating chromophores linked with a chain of appropriate length (bis-intercalators) show high dsDNA binding affinities, greater by several orders of magnitude than those of the corresponding monomers.

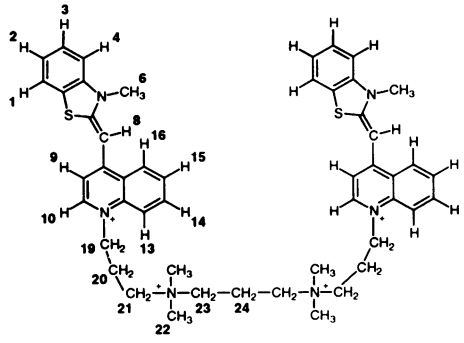
Most naturally occurring intercalators consist of rigid planar aromatic chromophores attached to bulky side groups. In some cases it has been shown that such intercalators bind preferentially to specific dsDNA sequences (site selectivity) (2). However, this site-selectivity is mainly attributed to either major or minor groove interactions of the bulky side groups with the dsDNA. In general, synthetic compounds consisting of intercalating chromophores without bulky side groups do not exhibit any selectivity in the binding to various dsDNA oligonucleotides (3).

Recently, Glazer and co-workers (4–6) synthesized and characterized a new family of high affinity fluorescent dyes for making dsDNA complexes. Among these, a homodimeric thiazole orange~dye, 1,1′-(4,4,8,8-tetramethyl-4,8-diazaundecamethylene)-bis-4-(3-methyl-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)-quinolinium tetraiodide (TOTO, Scheme 1), binds extremely strongly, but non-covalently to dsDNA. It forms complexes that are stable to a variety of conditions, in particular gel electrophoresis. The enhancement of the fluorescence quantum yield upon binding of TOTO to dsDNA is >3000-fold. Thus, this compound allows high sensitivity detection of dsDNA by fluorescence scanners (7,8).

The extremely high binding affinity of TOTO to dsDNA has prompted us to carry out structural studies to characterize the binding mode. During the course of these studies it became apparent that this compound exhibits site selectivity in the binding. The charged poly(aminoalkyl) linker of TOTO resembles spermine and spermidine and could therefore contribute to the site selectivity accordingly. Binding of spermine and spermidine to dsDNA has been investigated by several methods (9–12). Studies of the interaction using NMR spectroscopy suggest a non-selective, electrostatic interaction. Wemmer *et al.* (9) found that proton NOEs of spermine bound to d(CGCGAATTCGCG)$_2$ were positive. This indicates that the motional flexibility of the polyamine is largely independent of the overall motion of the dsDNA. Based on competitive titrations of dsDNA studied by ^{23}Na NMR spectroscopy Padmanabhan *et al.* (10,11) also concluded that the interactions were largely non-selective. Equilibrium dialysis measurements of the binding of spermidine, spermine and putrescine to dsDNA were similarly consistent with non-selective binding (12). The site selective binding of TOTO to dsDNA must therefore be caused predominantly by the thiazole orange chromophores. No previous model compound of this type has been investigated and no information about the binding was available. Consequently, the search for a preferred dsDNA binding sequence had to rely on trial and error.

* To whom correspondence should be addressed

Scheme 1. Numbering scheme in TOTO.



Scheme 2. Numbering scheme in the dsDNA duplexes used.

```
cgcg:
     5' - C1 G2 C3 G4 - 3'
     3' - G8 C7 G6 C5 - 5'

a3c2:
     5' - C1  G2  C3  A4  A5  A6  C7  C8  G9  G10 C11 - 3'
     3' - G22 C21 G20 T19 T18 T17 G16 G15 C14 C13 G12 - 5'

atat:
     5' - C1  G2  C3  G4  A5  T6  A7  T8  C9  G10 C11 C12 - 3'
     3' - G24 C23 G22 C21 T20 A19 T18 A17 G16 C15 G14 C13 - 5'

actga:
     5' - C1  G2  C3  A4  C5  T6  G7  A8  T9  G10 C11 - 3'
     3' - G22 C21 G20 T19 G18 A17 C16 T15 A14 C13 G12 - 5'

ctga:
     5' - C1  C2  G3  C4  T5  G6  A7  G8  C9 - 3'
     3' - G18 G17 C16 G15 A14 C13 T12 C11 G10- 5'

ctag:
     5' - C1  C2  G3  C4  T5  A6  G7  C8  G9 - 3'
     3' - G18 G17 C16 G15 A14 T13 C12 G11 C10- 5'

ttaa:
     5' - C1  G2  C3  G4  T5  T6  A7  A8  C9  G10 C11 C12 - 3'
     3' - G24 C23 G22 C21 A20 A19 T18 T17 G16 C15 G14 C13 - 5'

tcga:
     5' - C1  C2  G3  T4  C5  G6  A7  C8  G9 - 3'
     3' - G18 G17 C16 A15 G14 C13 T12 G11 C10- 5'
```

MATERIALS AND METHODS

Materials

The DNA oligonucleotides were either synthesized and purified as reported previously (13) or purchased from DNA Technology, Aarhus, Denmark and used without further purification. The non-selfcomplementary single stranded oligomers were added to equivalent amount of the complementary strand and the duplexes were formed by annealing from 80°C. The dsDNA oligomers are shown in Scheme 2.

TOTO is almost insoluble in water. The complexes with dsDNA can therefore not be made by simple titration of the dsDNA with a TOTO solution. Consequently, the experimental procedure developed is to dissolve TOTO in DMSO-d_6 and add the proper amount of this solution to approximately 5 ml of a diluted solution of the oligonucleotide in water. The addition was done dropwise under rapid stirring and the reaction mixture was lyophilized on a Speed vac immediately following the addition. Lyophilizing twice from D_2O left no visible DMSO and no lines from DMSO-d_5 in the ^1H NMR spectrum. All the complexes referred to in this paper were made as 1:1 complexes of TOTO and dsDNA, but often the complex was made in steps, yielding first the 1:2 TOTO:dsDNA products. These products were first characterized as mixtures of free dsDNAs and complexes before further addition of TOTO to equivalent amount. In making the complexes, the dsDNA was slightly undertitrated with TOTO to reduce precipitation due to cross-linking. Occasionally, this leaves visible signals from the free oligonucleotide in the spectra. The concentrations of the dsDNA and the TOTO in the stock solutions were determined by UV spectroscopy using the extinction coefficients reported in the literature (4,14).

A major problem encountered in working with TOTO is the ability to cross-link separate oligonucleotide duplexes (4). The degree of cross-linking is dependent on the duplex and the experimental procedure used. Usually, the cross-linking results in only slight precipitation. The TOTO samples made are quite stable with slight precipitation (probably due to cross-linking) as the most severe problem during ageing of the sample and on lyophilization.

NMR samples were prepared by dissolving the lyophilized complexes in 0.5 ml of 10 mM sodium phosphate buffer (pH 7.0) and 0.05 mM sodium EDTA. The dsDNA concentrations were between 1 and 3 mM. For experiments carried out in D_2O, the solid complexes lyophilized twice from D_2O were redissolved in 99.96% D_2O (Cambridge Isotope Laboratories). A 90% H_2O/10% D_2O mixture (0.5 ml) was used as solvent for experiments in H_2O.

Methods

All NMR experiments were performed at 500 MHz on either a GE500 or a Varian Unity 500 spectrometer. NOESY spectra were acquired with various mixing times from 50 to 200 ms in D_2O using 1024 complex points in t_2 and a spectral width of 5000 Hz. 512 t_1 experiments were recorded using either the States or TPPI phase cycling scheme (15,16). Normally 64 scans were acquired for each t_1 value. The TOCSY experiments (17,18) with mixing times of 30, 90 and 130 ms were obtained in the TPPI mode (16) using 1024 complex points in t_2, 512 t_1 experiments and by acquiring 64 scans for each t_1 value. The NOESY spectra in H_2O were acquired with a spectral width of 10 000 Hz in 2048 complexes points using a pulse sequence where the last 90° pulse was replaced by a 1-1 jump and return sequence to suppress the solvent signal (19). Spectra were obtained at 25°C but commonly spectra were also obtained at lower temperature (down to 2°C) to investigate the presence of dynamic interchange. The acquired data were processed using Felix (version 2.1, Biosym Technologies, San Diego). Whenever possible the TOCSY and NOESY spectra were interpreted by conventional methods (20,21).

The model of the complex of TOTO with the sequence d(CCGCTGAGC): d(GCTCAGCGG) was constructed using the Biopolymer modul of InsightII (version 2.3.5) (Biosym Technologies, San Diego). The assigned NOESY crosspeaks were integrated and incorporated into a restrained molecular dynamics (RMD) procedure. The RMD and energy minimization calculations were performed using Discover (version 2.95) (Biosym Technologies, San Diego) with a modified AMBER force-field potential. TOTO was docked manually into a B-DNA form of the oligonucleotide. The NOE derived distance restraints were applied and the model was energy minimized. This was followed by 28 ps of restrained molecular dynamics with the following temperature profile: 600 K for 4 ps cooled to 200 K in 50 K steps of 3 ps each. The final structure was then energy minimized. The C1C2G3:C16G17G18 and the G8C9:G10C11 ends of the duplex were fixed in B-DNA conformation by artificial NOE restraints deduced from a B-DNA structure built in InsightII. Standard

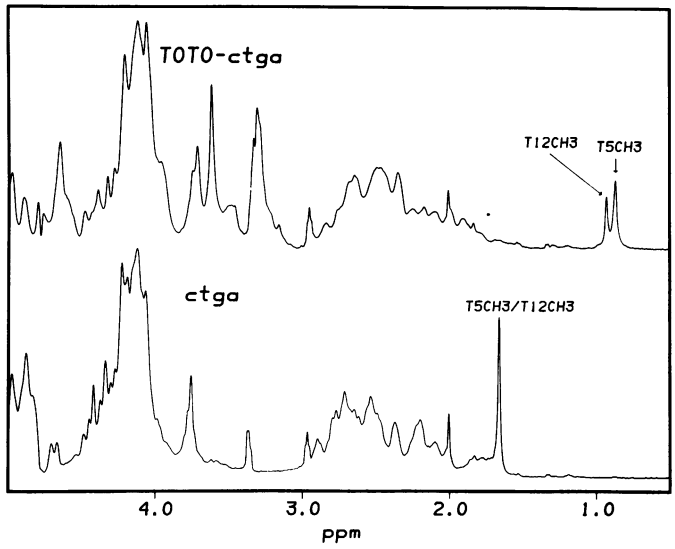


Figure 1. The aliphatic region of the 1D spectrum of the 1:1 TOTO–ctga complex (top) compared to the free ctga oligonucleotide (bottom). The large shift of the two thymidine methyl group protons on complex formation is clearly visible.

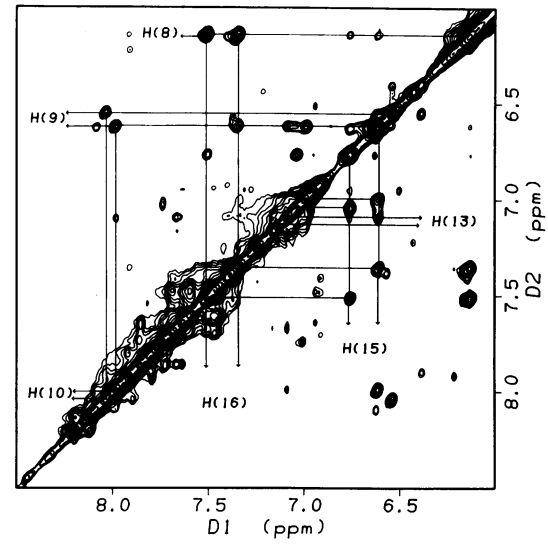


Figure 3. Aromatic to aromatic part of the 200 ms NOESY spectrum of the 1:1 TOTO–ctga complex. The two sequential paths of H(8)–H(16)–H(15)–H(14)–H(13) have been indicated together with the connectivities between the two H(9) and H(10) protons.

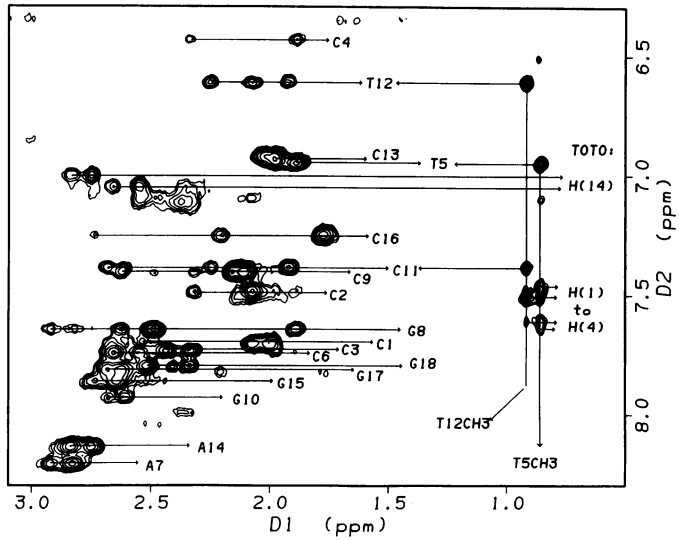


Figure 2. H(2′)/H(2″) to aromatic part of the 200 ms NOESY spectrum of the 1:1 TOTO–ctga complex. The assignments of the H(6)/H(8) of the various nucleotides are indicated together with the two thymidine methyl protons and some of the protons on TOTO. The eight TOTO H(1–4) protons cannot be assigned unambiguously. Note that the two chromophores of TOTO yield two distinguishable chemical shifts for H(14).

Watson–Crick base pairing was obtained by including hydrogen bonds as distance restraints. Twelve experimentally determined NOE restraints between TOTO and protons in the binding site were introduced together with the two intramolecular NOE derived restraints between TOTO H(8) and H(16). The NOE restraints were obtained in the two spin approximation from the NOESY spectrum obtained with a mixing time of 100 ms. Force constants of 50 kcal/mol/Å^2 were used.

RESULTS

Oligonucleotide d(CGCG):d(CGCG) (cgcg)

The obvious first choice of a binding sequence for TOTO is a 4 bp oligonucleotide since this contains only one target for the formation of a bis-intercalated TOTO–dsDNA complex. This led us to use the d(CGCG):d(CGCG) duplex in our first attempt to make a TOTO complex. As mentioned, TOTO is insoluble in water and is therefore dissolved in DMSO and added dropwise to a diluted aqueous solution of the nucleotide. Using this procedure in case of d(CGCG):d(CGCG) led to complete precipitation, probably caused by formation of insoluble complexes. The large degree of charge neutralization of the 4 bp oligonucleotide by the positively charged TOTO molecule is probably responsible for the precipitation. Thus, use of small oligonucleotides in an attempt to make soluble TOTO–dsDNA complexes is not a feasible procedure.

Oligonucleotide d(CGCAAACCGGC):d(GCCGGTTTGCG) (a3c2)

The use of larger oligonucleotides in the complex formation with TOTO represents the obvious problem of possible formation of several different complexes. TOTO may be bound to various binding sites. This will occur unless the oligonucleotide contains a single preferred binding site for TOTO. The a3c2-oligonucleotide has a restricted A–T base pair region and a possible A–T preference in the binding of TOTO should make this oligonucleotide a favorable choice.

The experimental procedure used actually yielded soluble complexes of TOTO with this oligonucleotide, but the NMR spectra clearly showed the existence of at least two different complexes. One of them with broad lines indicating the presence of interchange between different binding sites. The lines in the spectra were too broad to allow assignments.

Table 1. Chemical shifts (in p.p.m.) of the TOTO–ctga complex compared to the free oligonucleotides given in parantheses.

ctga	H(1')	H(2')	H(2")	H(6)/H(8)	H(5)/H(2) /CH_3	H(1)/H(3)	H(4)	H(4)
C1	5.95 (6.03)	2.01 (2.09)	2.48 (2.56)	7.69 (7.76)	5.90 (5.98)		-	-
C2	5.50 (5.63)	2.08 (2.17)	2.33 (2.47)	7.47 (7.57)	5.63 (5.71)		8.55 (8.62)	6.82 (6.89)
G3	5.67 (5.99)	2.44 (2.73)	2.35 (2.78)	7.73 (7.95)		12.70 (12.99)		
C4	5.98 (5.96)	1.90 (2.01)	2.51 (2.50)	6.43 (7.41)	4.60 (5.38)		7.95 (8.25)	6.57 (6.54)
T5	5.04 (5.67)	1.90 (2.01)	1.99 (2.36)	6.95 (7.29)	0.87 (1.65)	13.84 (13.85)		
G6	5.06 (5.46)	2.56 (2.71)	2.67 (2.77)	7.34 (7.90)		12.38 (12.62)		
A7	5.99 (6.07)	2.84 (2.69)	2.94 (2.88)	8.20 (8.12)	7.07 (7.57)			
G8	5.79 (5.83)	2.50 (2.46)	2.64 (2.66)	7.64 (7.63)		12.58 (12.99)		
C9	6.15 (6.17)	2.09 (2.13)	2.18 (2.18)	7.40 (7.42)	5.37 (5.41)		-	-
G10	5.93 (6.04)	2.62 (2.71)	2.70 (2.81)	7.92 (8.00)		-		
C11	5.90 (6.13)	1.92 (2.21)	2.25 (2.60)	7.37 (7.58)	5.25 (5.43)		8.17 (8.30)	6.57 (6.71)
T12	6.09 (6.10)	2.09 (2.18)	2.68 (2.54)	6.61 (7.46)	0.93 (1.65)	13.97 (13.85)		
C13	4.28 (5.43)	1.99 (2.09)	2.05 (2.38)	6.91 (7.55)	4.68 (5.70)		7.68 (8.58)	6.72 (6.84)
A14	5.50 (6.00)	2.85 (2.76)	2.75 (2.91)	8.12 (8.22)	7.12 (7.67)			
G15	5.87 (5.73)	2.65 (2.52)	2.75 (2.62)	7.86 (7.70)		11.98 (12.76)		
C16	5.57 (5.67)	1.79 (1.77)	2.11 (2.25)	7.24 (7.24)	5.29 (5.30)		8.19 (8.30)	6.32 (6.34)
G17	5.59 (5.65)	2.64 (2.64)	2.70 (2.71)	7.81 (7.84)		13.10 (13.17)		
G18	6.11 (6.18)	2.53 (2.53)	2.35 (2.37)	7.79 (7.81)		-		

TOTO	ring 1	ring 2	TOTO	ring 1	ring 2
H(13)	7.10	7.09	H(8)	6.14	6.14
H(14)	7.03	6.99	H(9)	6.63	6.55
H(15)	6.77	6.62	H(10)	7.98	8.03
H(16)	7.51	7.35	CH_3(6)	3.62	3.62

The values are given at 25°C relative to HOD at 4.78 p.p.m. Hydrogen-bonded amide protons are underlined. H(1–4) overlap in the region 7.4–7.6 p.p.m.

Oligonucleotide d(CGCGATATCGCG):d(CGCGATATCGCG) (atat)

In order to investigate the possibility of a pure ApT:ApT or a pure CpG:CpG site selectivity of TOTO the atat oligonucleotide was used in the complex formation. However, several complexes between TOTO and this oligonucleotide were formed and the lines in the NMR spectra were broad. Sequential assignments were not possible.

Oligonucleotide d(CCGACTGATGC):d(GCATCAGTCGG) (actga)

The actga-oligonucleotide has a pyrimidine–pyrimidine–purine–purine step in the CTGA:TCAG sequence of the oligonucleotide. Complex formation with TOTO showed that this is in fact the preferred binding site in this oligonucleotide. Only one major complex between this oligonucleotide and TOTO was formed. The complex yielded 2-D NMR spectra of sufficiently high quality to allow sequential assignment in the H(1')–H(6)/H(8) and H(2')/H(2")–H(6)/H(8) regions of the NOESY spectra. Furthermore, a few NOE connectivities between protons on TOTO and protons on dsDNA were observed. The sequential NOE connectivities unequivocally showed that TOTO bis-intercalates in the CTGA:TCAG binding site with one of the thiazole orange dye ring systems located between the CpT:ApG base pairs and the other between the GpA:TpC base pairs.

The quality of the NMR spectra of the TOTO–actga complex was more than sufficient to verify the binding site, but the lines in the spectra were slightly too broad to allow the observation of more than a few intermolecular NOE connectivities. Formation of a complex with narrower NMR lines was desirable since such connectivities between TOTO and dsDNA are important for a precise structure determination. Consequently, a 9 bp oligonucleotide version of the actga-oligonucleotide was synthesized.

Oligonucleotide d(CCGCTGAGC):d(GCTCAGCGG) (ctga)

The ctga-oligonucleotide contains the CTGA:TCAG binding site similar to the actga oligonucleotide, but due to the smaller size a better behaved complex is to be expected. The ^{1}H NMR spectrum of this complex is shown in Figure 1. One of the characteristic features of this spectrum is the up-field shift of the thymidine methyl group of about 0.8 p.p.m. An equivalent shift was also observed in the TOTO–actga complex. However in contrast to the actga case, the two methyl signals at $\delta = 0.8$ p.p.m. have unequal intensity. With several other features in the spectrum this proves that more than one complex was formed in substantial amount.

NOESY spectra obtained in both D_2O and H_2O and at different temperatures showed that ~80% of the binding of TOTO to the oligonucleotide occurred in the CTGA:TCGA binding site while <20% occurred in the TGAG:CTCA site. Parts of the NOESY spectrum are shown in Figures 2 and 3. From these spectra it was possible to follow the sequential NOE connectivities from aromatic H(6)/H(8) protons to H(1') and H(2')/H(2") of the major form. Similar to the TOTO–actga complex it was found that these connectivities were interrupted at the 5'-C4pT5–3' and 5'-G6pA7-3' base pair steps. This is a clear evidence of bis-intercalation. The interruption of the sequential connectivities is also demonstrated in Figure 3 since there is no connectivity between the methyl group of T5 and the H(6) proton on C4. The chemical shifts of all the aromatic protons [H(6), H(5), H(8) and H(2)] and the deoxyribose protons [H(1'), H(2') and H(2")] of the dsDNA in the complex were assigned in the conventional way (20,21) and listed in Table 1.

Internal NOE connectivities in each chromophore of TOTO show a distinct pattern in the NOESY spectra. An extremely strong cross peak between H(8) and H(16) is observed. Combined with connectivities in the path [H(16)–H(15)–H(14)–H(13)], [H(8)–CH_3(6)–H(4)– H(3)–H(2)–H(1)] and [H(9)–H(10)] this establishes the assignments and structure of the TOTO ligand. Each of the two chromophores of TOTO yielded an independent set of connectivities (Fig. 3). The chemical shift values of the protons in each aromatic chromophore of TOTO were assigned and are given in Table 1. The overlap of H(1–4) resonances of the chromophores was too severe to allow individual assignments of these protons. Assignments of the protons in the linker chain were ambiguous due to both strong couplings and overlapping multiplets. However, some CH_2 protons were identified, e.g. the H(19) protons at 4.15 and 4.65 p.p.m. and the CH_3(22) at 3.30 p.p.m.

Table 2. The dye-dsDNA NOE connectivities observed in the 200 ms NOESY spectrum of the TOTO–ctga complex

TOTO-1	Nucleotide protons	TOTO-2	Nucleotide protons
H(1–4)	T5CH_3, T5H(6)	H(1–4)	T12CH_3,C13H(6)
		H(13)	T12H(2'), T12H(2")
H(14)	A14H(2'), A14H(2"), A14H(1')	H(14)	G6H(2'),G6H(2"),G6H(3')
CH_3(22)	A14H(1')	CH_3(6)	T12CH_3

The chemical shift values of observed labile protons in the TOTO–ctga complex are included in Table 1. The NOESY spectra of the complex obtained in H_2O contain the normal cross peaks between the thymidine imino protons (T5/T12 H3) and the H2 protons of the base paired adenosines (A14/A7 H2) as well as the normal cross peak between the guanine imino protons and the

cytosine amino protons. This shows that the normal Watson–Crick base pairing is retained on complex formation.

It was not possible to follow the sequential connectivities of the minor form of the complex of TOTO to ctga due to the low abundance. However, parts of this connectivity path are interrupted between 5′-T5pG6-3′ and 5′-A7pG8-3′ in accordance with the suggested bis-intercalation in the TGAG:CTCA site. Thus, the preference of TOTO for binding between a cytosine and thymidine base pair is dominating.

Several connectivities between TOTO and the ctga oligonucleotide were observed and identified in the NOESY spectrum of the major form (Table 2). The methyl group of T5 has a very strong cross peaks to H(1–4) of the TOTO chromophore. This combined with observed crosspeaks of the same TOTO chromophore to H(2′)/H(2″) of A14 determines the position of the benzothiazole ring. Similar cross peaks determine the position of the other chromophore. Additional connectivities in the NOESY spectra show that TOTO binds to the dsDNA with the linker chain in the minor groove.

The NOE constraints given in Table 2 have been used to develop a model of the complex. The connectivities in the 100 ms mixing time NOESY spectrum were converted to distance restraints in the two-spin approximation, ignoring spin diffusion. Molecular mechanics minimization and molecular dynamics calculations resulted in the model shown in Figure 4. Bis-intercalation of TOTO in dsDNA is clearly proved.

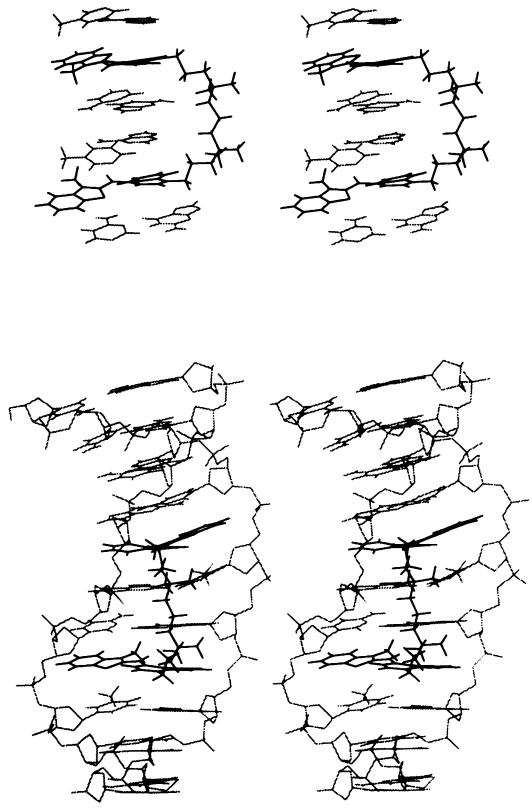


Figure 4. Stereo view of a stick plot of the structure of the TOTO–ctga complex. The dsDNA is shown in dotted lines while the TOTO molecules is drawn in solid lines. **Bottom**: The total complex looking into the minor groove. Deoxyribose protons have been omitted for clarity. **Top**: The CTGA:TCAG binding site looking from the side. Only the nucleobases of the oligonucleotides are shown.

Oligonucleotide d(CCGCTAGCG):d(CGCTAGCGG) (ctag)

The two complexes formed between TOTO and the ctga-oligonucleotide showed that TOTO has a preference for binding in a CpT:GpA site. Consequently, an obvious choice for a preferred binding sequence for bis-intercalation is the CTAG:CTAG sequence. The one-dimensional ^{1}H NMR spectrum of the complex compared to the free dsDNA duplex is shown in Figure 5. Part of the corresponding NOESY spectrum is given in Figure 6.

NOESY as well as the TOCSY spectra of the TOTO–ctag complex show that it is almost symmetric. The NOESY spectra of the complex exhibit the characteristic features of dsDNA connectivities but the sequential NOE connectivities from aromatic H(6)/H(8) protons to H(1′) and H(2′),H(2″) are interrupted at the 5′-C4pT5-3′ and 5′-A6pG7-3′ base pair steps. This is clear evidence of bis-intercalation. The interruption of the sequential connectivities can also be observed in the cross peak pattern of the methyl group of T5 where the peak to H(6) of C4 is missing. The aromatic [H(6), H(5), H(8), and H(2)] and deoxyribose proton [H(1′), H(2′), H(2″), H(3′) and some of H(4′)] resonances of dsDNA in the complex were assigned in the conventional way (20,21) and are given in Table 3.

The internal NOE connectivities in the chromophore of TOTO are again distinct features in the NOESY spectra of the ctag complex, equivalent to the ctga case. However, due to the near symmetry of the complex, the spectra of the two chromophore systems are superimposed. An extremely strong cross peak between H(8) and H(16) is observed. Combined with the connectivities [H16–H(15)–H(14)–H(13)], [H(8)–CH_3(6)–H(4)–H(3)–H(2)–H(1)] and [H(9)–H(10)] this establishes the assignment and structure of the TOTO ligand. The linker chain in TOTO can be assigned to a certain degree based on combined use of TOCSY and NOESY cross-peaks. The assigned chemical shift values of TOTO in this complex are given in Table 3.

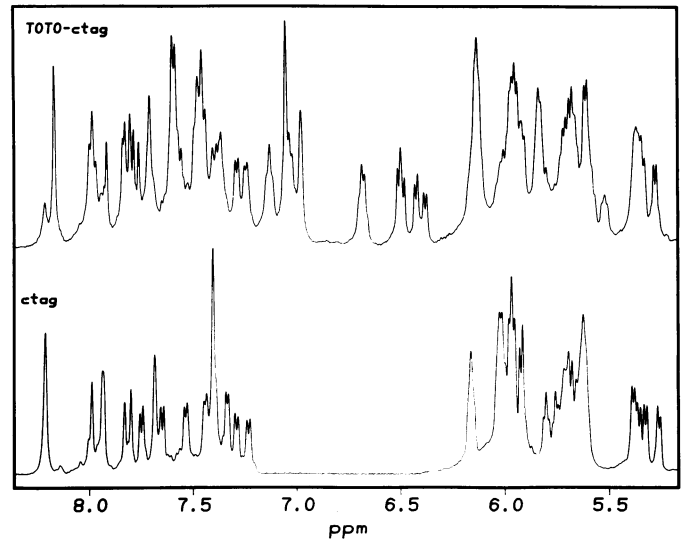


Figure 5. The aromatic part of the 1D spectrum of the 1:1 TOTO–ctag complex (**top**) compared to the free ctag oligonucleotide (**bottom**). Lines from protons on TOTO are easily identified in the region 6.3–7.2 p.p.m. (cf. Table 3).

Table 3. Chemical shifts (in p.p.m.) of the TOTO–ctag complex compared to the free oligonucleotide given in parantheses

ctag	H(1')	H(2')	H(2")	H(3')	H(4')	H(6)/H(8)	H(5)/H(2) /CH$_3$	H(1)/H(3)	H(4)/H(6)	H(4)/H(6)
C1	5.92 (6.00)	2.07 (2.07)	2.47 (2.53)	4.65 (4.68)	4.12 (4.13)	7.71 (7.75)	5.93 (5.96)		-	-
C2	5.53 (5.61)	2.01 (2.15)	2.33 (2.45)	4.82 (4.87)	4.10 (4.14)	7.46 (7.53)	5.63 (5.69)		8.47 (8.60)	6.83 (6.92)
G3	5.68 (5.96)	2.33 (2.70)	2.44 (2.77)	4.86 (5.00)	4.26 (4.39)	7.71 (7.93)		12.73 (13.00)		
C4	5.94 (5.91)	1.90 (2.03)	2.52 (2.49)	4.65 (4.72)	- (4.22)	6.37 (7.39)	4.51 (5.36)		7.91 (8.24)	6.55 (6.57)
T5	4.63 (5.61)	1.92 (2.12)	1.92 (2.47)	- (4.87)	- (4.14)	6.97 (7.40)	0.90 (1.68)	13.62 (13.73)		
A6	5.37 (6.02)	2.73 (2.73)	2.84 (2.88)	5.11 (5.03)	- (4.40)	8.17 (8.20)	7.06 (7.34)		7.29 (-)	6.68 (-)
G7	5.84 (5.71)	2.66 (2.50)	2.72 (2.60)	4.89 (4.95)	4.15 (4.36)	7.82 (7.67)		12.04 (12.77)		
C8	5.73 (5.75)	1.86 (1.87)	2.30 (2.32)	- (4.78)	- (4.12)	7.28 (7.29)	5.35 (5.33)		8.23 (8.35)	6.47 (6.48)
G9	6.14 (6.16)	2.61 (2.60)	2.35 (2.35)	4.68 (4.65)	4.19 (4.17)	7.91 (7.92)		-		
C10	5.69 (5.79)	1.88 (2.00)	2.30 (2.45)	4.65 (4.72)	4.05 (4.08)	7.56 (7.64)	5.85 (5.92)		-	-
G11	5.68 (5.96)	2.34 (2.70)	2.45 (2.77)	4.86 (5.00)	4.26(4.39)	7.76 (7.98)		12.79 (13.04)		
C12	5.98 (5.97)	1.92 (2.07)	2.53 (2.51)	4.65 (4.76)	- (4.25)	6.42 (7.44)	4.54 (5.39)		7.94 (8.27)	6.59 (6.59)
T13	4.63 (5.61)	1.92 (2.12)	1.92 (2.47)	- (4.87)	- (4.14)	6.97 (7.40)	0.90 (1.68)	13.62 (13.75)		
A14	5.37 (6.02)	2.73 (2.73)	2.84 (2.88)	5.11 (5.03)	- (4.40)	8.17 (8.20)	7.06 (7.33)		7.29 (-)	6.68 (-)
G15	5.84 (5.71)	2.66 (2.50)	2.72 (2.60)	4.89 (4.95)	4.15 (4.36)	7.84 (7.67)		12.00 (12.76)		
C16	5.58 (5.65)	1.78 (1.75)	2.21(2.24)	- (4.78)	- (4.12)	7.24 (7.23)	5.29 (5.26)		8.18 (8.31)	6.32 (6.34)
G17	5.60 (5.63)	2.63 (2.61)	2.69 (2.71)	4.96 (4.95)	4.38 (4.36)	7.80 (7.82)		13.01 (13.19)		
G18	6.13 (6.16)	2.52 (2.52)	2.35 (2.35)	4.68 (4.64)	4.19 (4.21)	7.78 (7.79)		-		

TOTO		TOTO		TOTO	
H(13)	7.13	H(8)	6.13	H(19)	4.15, 4.71
H(14)	7.03	H(9)	6.49	H(20),H(24)	~2.50
H(15)	6.67	H(10)	7.99	H(21)	3.50, 3.64
H(16)	7.37	CH$_3$(6)	3.60	CH$_3$(22)	3.28

The values are given at 25°C relative to HOD at 4.78 p.p.m. Hydrogen-bonded amide protons are underlined. H(1–4) overlap in the region 7.4–7.6 p.p.m.

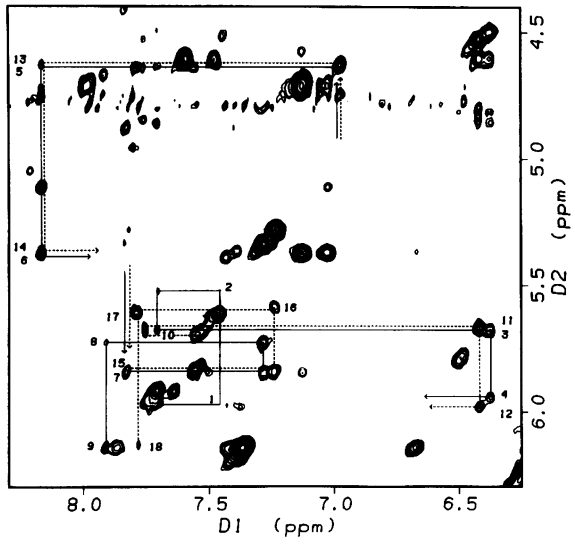


Figure 6. The H(1') to aromatic part of the 200 ms NOESY spectrum of the 1:1 TOTO–ctag complex. The sequential H(1')–H(6)/H(8)–H(1') connectivity pathways for each dsDNA strand are indicated with a full and dotted line respectively. The interrupted connectivities at the two intercalation sites are indicated with arrows.

The chemical shift values of the observed labile protons in the TOTO–ctag complex are included in Table 3. The NOESY spectra of the complex obtained in H_2O contain the normal cross peaks between the thymidine imino protons (T5/T13 H3) and the H2 protons of the base paired adenosines (A14/A6 H2) as well as the normal cross peak between the guanine imino protons and the cytosine amino protons. This shows that the normal Watson–Crick base pairing has been retained on complex formation. Furthermore, the NOESY spectra at 25°C of the complex dissolved in H_2O contain distinct cross peaks between the thymidine imino proton (T5/T13 H3) and the base paired adenosine amino protons (A14/A6 H6). These peaks are absent in normal dsDNA at room temperature due to exchange. The presence in the spectra of the complex therefore shows that the exchange of these protons is slower in the complex than in the free dsDNA. This lowering of the exchange rates of the adenosine amino protons in the TOTO complex is a clear consequence of the bis-intercalation.

The spectra of the TOTO–ctag complex have line widths that are smaller than those observed in any other complex we have made. Furthermore, there is no sign whatsoever of any minor form. This means that there is a strong preference for the CTAG:CTAG site. There is no sign of any dynamic interchange between different complexes.

Several connectivities between TOTO and dsDNA were observed and identified in the NOESY spectra. The methyl groups of T5 and T13 have strong cross peaks to H(1–4) of each of the two TOTO chromophores. This determines the position of the benzthiazole ring above the cytosine ring. The quinolinium ring is similarly positioned on top of an adenine base by cross peaks between the adenine H(2'),H(2") and the TOTO H(14). Several cross peaks between the aromatic protons of dsDNA in the binding sites and the protons in the TOTO chromophore support these positions. Additional cross peaks in the NOESY spectra show that TOTO binds to the dsDNA with the linker chain positioned in the minor groove.

Oligonucleotide d(CCGTCGACG):d(CGTCGACGG) (tcga)

The well characterized TOTO complexes formed by both the ctga and the ctag oligonucleotides leave the question of whether CpT:ApG or TpC:GpA is the preferred binding site for a thiaxole orange dye chromophore. To clarify this point we have used the tcga oligonucleotide in an attempt to make a TOTO complex following the prescription used in the formation of the other complexes. However, TOTO does not form well defined complexes with this oligonucleotide. More than one complex is formed and the NMR spectrum consists of broad lines. There is definitely no sign of any major product with bis-intercalation in the TCGA:TCGA site.

Oligonucleotide d(CGCGTTAACGCG)$_2$ (ttaa)

The strong preference of TOTO for especially a CTAG:CTAG step made it feasible also to investigate whether other pyrimidine–pyrimidine–purine–purine steps are favored binding sites as well. The ttaa is an obvious choice in this context.

At least two complexes with a dynamic interchange are formed between TOTO and the ttaa oligonucleotide. It is obvious from the spectra that intercalation in the ttaa duplex to some extent also involves the G4/G16 nucleotides. The interchange between the complexes is slow on the NMR time-scale but fast enough to yield various exchange cross peaks in the NOESY spectra. The existence of a symmetric and an asymmetric complex is a straightforward explanation of the observed exchange cross peaks pattern. No detailed characterization of the complexes with the ttaa oligonucleotide can be given. However, it is beyond any doubt that this sequence does not contain a single preferred binding site for TOTO.

DISCUSSION

The conformation of free TOTO in a DMSO solution has been studied by Rye *et al.*(4) Based on the NOESY spectrum, e.g. a strong cross peak between H(8) and H(16), they found that the relative conformation of the two ring systems in each chromophore is the one indicated in Scheme 1. For TOTO bound to dsDNA we found the same cross peak pattern as the one observed for free TOTO. The relative conformation in the chromophores is therefore unaffected by intercalation.

The various complexes formed between TOTO and the oligonucleotides as described in this paper prove that the molecule predominantly bis-intercalates in dsDNA. TOTO binds to all the sequences used. Usually a dynamic mixture of different binding sites exists. For the oligonucleotides containing a symmetric cytosine–thymidine binding site TOTO binds preferentially to that site. In the presence of a CTAG:CTAG binding site TOTO binds exclusively to this sequence without showing any interchange to other binding sites. Although care must be taken when interpreting dynamic properties in terms of stability, it seems reasonable to conclude that TOTO has a high degree of site selectivity in binding to dsDNA.

For the ctag oligonucleotide we can estimate that there is at least 100-fold preference for binding to the CTAG:CTAG site than to any other site present in that oligonucleotide. The spectra of the complex formed with the actga oligonucleotide demonstrate that the binding to the CTGA:TCAG dominates the binding to any other site in this oligonucleotide to at least 50-fold.

The complex formed between TOTO and the ctga oligonucleotide shows that there is a competition between binding TOTO to the CTGA:TCAG and the CTCA:TGAG sites with preference to the former. The degree of preferential binding of the thiazole orange chromophore of TOTO to a CpT:ApG site compared to a TpC:GpA is further clarified through the unsuccessful attempt to make a single complex of TOTO with the tcga oligonucleotide. The formation of more than one complex yielding an NMR spectrum with broad lines shows that the TCGA:TCGA is not a preferred binding site, since the broad lines show that a dynamic interchange between different binding sites takes place.

It may seem obvious from the results presented in this paper to conclude that TOTO selectively binds to the CTAG:CTAG site, but it has to be emphasized that there are a large number of possible 4 bp binding sites of the bis-intercalating TOTO molecule. We have not systematically examined all these, but adding up all the sequences done we have covered a significant fraction. More importantly, we have only to a small degree examined the degree of preference between different binding sites by use of competition experiments with more than one binding site in a strand.

With respect to the line widths, the NMR spectra of the TOTO–ctag complex are distinctly different from the spectra obtained of all the other complexes. The lines are much narrower for this complex than for any of the other complexes. No dynamical interchange on the NMR time-scale between different complexes exists. This strongly supports the conclusion that TOTO selectively binds to CTAG:CTAG if this site is present in the oligonucleotide. A trial and error strategy was used to obtain this result, and to quantify the degree of preferential binding to this sequence we must rely on studies of this site in oligonucleotide with competitive sites. Thus, quantitative estimation of the preference in binding to this sequence can only to a minor degree be given based on the results in this paper. The molecular basis for this site selectivity has been clarified by the determination of the three dimensional structure of a similar complex (22).

The model of the TOTO–ctga shown in Figure 4 demonstrates that TOTO bis-intercalates with the benzothiazole ring above the cytosine ring and the quinolinium ring system on top of an adenine base. The linker chain connecting the two chromophores is positioned in the minor groove. It should be emphasized that we have only obtained a model of the complex and shown it in Figure 4 to illustrate that the bis-intercalating binding mode is in agreement with the NMR data. This should not be confused with 3-D structure where more work has to be done to convert the NOESY cross peaks to distance restraints. A similar model was also obtained for the TOTO–ctag complex while a complete structure determination of a symmetric complex is in progress (22).

The chemical shift values of the protons in TOTO–ctag complex exhibit characteristic features that can be attributed to the intercalation of TOTO. Compared to the free dsDNA, the T5CH$_3$(6) protons are shifted 0.8 p.p.m. up-field. This is in agreement with a model of the complex where this methyl group is positioned exactly above the benzothiazole ring and thus exposed to a large ring current effect. Similar strong effect are observed for T5H(1′), T5H(2″), A6H(1′), A6H(2″), C4H(6) and T5H(6) consistent with the positions of these protons above or below a ring system of the intercalating chromophore. It is commonly accepted that intercalation imposes a large shift (~0.5–1 p.p.m.) on imino protons (23,24). However, this is only observed in case G7H(1) while T5H(3) remains almost unaffected by the intercalation of TOTO. A model indicates that the T5H(3) proton is positioned above the methylidene group right next to the aromatic ring systems and therefore is not influenced by additional ring current effect from the drug molecule.

There are a number of dsDNA complexes with other bis-intercalators that have been structurally characterized. The quinomycin antitumor antibiotics are a family of cyclic depsipeptides that bis-intercalate in dsDNA. Echinomycin is probably the best known among these compounds. The interaction of echinomycin with various dsDNA oligonucleotides has been described (23,24). Feigon and co-workers have studied the bis-intercalation complexes of echinomycin with two different octamers and a

decamer (25–27). They found that echinomycin bis-intercalates selectively on each side of a 5′-CpG-3′ site. Similar results have been found by Gao and Patel (28). TANDEM, a fully synthetic compound, belongs to the triostin family of naturally occurring antibiotics that are closely related to quinomycins. While triostin and echinomycin show high preference for 5′-CpG-3′ rich dsDNA a methylated TANDEM analog preferentially binds to alternating 5′-TpA-3′ and 5′-CpI-3′ sequences (29–31). Luzopeptin is also a cyclic depsipeptide antibiotic that bis-intercalates in dsDNA. It has been shown that it binds by bracketing two AT base pairs (32,33).

Common to the bis-intercalator of the quinomycin, triostin and luzopeptin type is the presence of two intercalating chromophores linked with cyclic peptide groups. The site selectivity of these compounds is predominantly ascribed to interaction of the peptide with the dsDNA in the grooves. However, recent molecular mechanics calculations on echinomycin binding to several dsDNA hexamers seem to indicate that stacking interaction may contribute to the selectivity of this drug (34,35).

Ditercalinium is a synthetic bis-intercalator with two rigid 7H-pyrido[4,3-c]carbazole chromophores linked with an aliphatic piperidine-containing chain. An analog of ditercalinium has a more flexible spermine type of linker chain. In a series of papers, Delepierre *et al.* (36–40) have investigated the binding of these compounds to dsDNA of the size from 4 to 8 bp. They concluded that the molecules bis-intercalate in the oligonucleotides with the chromophores preferentially at 5′-CpG-3′ sites and the linker chain in the major groove. While the conclusion with respect to the position of the linker chain is supported by an X-ray structural study (41) it is unclear to what extent the binding is selective. Most of the oligomers used were fairly short with only a few (or no) competing binding sites. Line broadening upon complex formation is probably due to dynamic exchange between different binding sites and makes spectral assignments difficult.

CONCLUSIONS

In this paper we have reported an investigation of the site selectivity of the bis-intercalating TOTO molecule. Using various oligonucleotides we found that TOTO prefers to bind in CTAG:CTAG sites. Although, we have not included all the possible combinations of a 4 bp binding site we do believe that the binding selectivity to this site is high. We also emphasize that the observed binding selectivity upon intercalation is caused mainly by the interaction of the two thiazole orange chromophores with the oligonucleotide.

ACKNOWLEDGEMENTS

The authors are grateful to Professor A. Glazer, University of California, Berkeley, USA for providing TOTO and to Dr B. Geierstanger, University of California, Berkeley, USA for providing some of the purified oligonucleotides. Dr H. P. Spielmann, University of California, Berkeley, USA is acknowledged for providing the modified AMBER force field. Dr B. Geierstanger and Dr H. P. Spielmann are further acknowledged for many helpful discussions. One of us (JPJ) was supported by the Danish Natural Science Research Council (SNF, grant no. 11-9701-1)

REFERENCES

1 Wakelin, L. P. G. (1986) *Med. Res. Rev.*, **6**, 275–340.
2 Searle, M. S. (1993) *Prog. in NMR Spectroscopy*, **25**, 403–480.
3 Wilson, A. D. and Jones, R. L. (1981) In Whittingham, M. S. and Jacobsen, A. J., (Eds.), *Intercalation Chemistry*. Academic Press, pp. 446–501.
4 Rye, H. S., Yue, S., Wemmer, D. E., Quesada, M. A., Haugland, R. P., Mathies, R. A. and Glazer, A. N. (1992) *Nucleic Acid Res.*, **20**, 2803–2812.
5 Benzon, S. C., Singh, P. S. and Glazer, A. N. (1993) *Nucleic Acid Res.*, **21**, 5727–5735.
6 Benzon, S. C., Mathies, R. A. and Glazer, A. N. (1993) *Nucleic Acid Res.*, **21**, 5720–5726.
7 Rye, H. S., Yue, S., Quesada, M. A., Haugland, R.P., Mathies, R. A. and Glazer, A. N. (1993) *Methods Enzymol.*, **217**, 414–431.
8 Rye, H. S., Dabora, J. M., Quesada, M. A., Mathies, R. A. and Glazer, A. N. (1993) *Anal. Biochem.*, **208**, 144–150.
9 Wemmer, D. E., Scrivenugopal, K. S., Reid, B. R. and Morris, D. R. (1985) *J. Mol. Biol.*, **185**, 457–459.
10 Padmanabhan, S., Richey, B., Anderson, C. F. and Record, M. T., Jr. (1988) *Biochemistry*, **27**, 4367–4376.
11 Padmanabhan, S., Brushaber, V. M., Anderson, C. F. and Record, M. T., Jr. (1991) *Biochemistry*, **30**, 7550–7559.
12 Braunlin, W. H., Strick, T. J. and Record, M. T., Jr. (1986) *Biopolymers*, **21**, 1301–1314.
13 Pelton, J. G. and Wemmer, D. E. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5723–5727.
14 Brown, T. and Brown, D. J. S. (1991) In Eckstein, F. (Ed.), *Oligonucleotides and Analogues. A Practical Approach*, Oxford University Press, Oxford pp 1–23.
15 States, D. J., Haberkorn and R. A. and Reuben, F. J. (1982) *J. Magn. Reson.*, **48**, 286–292.
16 Bodenhausen, G., Kogler, H. and Ernst, R. R. (1984) *J. Magn. Reson.*, **58**, 2161–2180.
17 Levitt, M., Freeman, R. and Frenkiel, T. (1982) *J. Magn. Reson.*, **47**, 328–330.
18 Bax, A. and Davis, D. (1985) *J. Magn. Reson.* **65**, 355–360.
19 Skelenár, V. and Bax, A. (1987) *J. Magn. Reson.*, **74**, 469–479.
20 Hare, D. R. and Wemmer, D. E., Chou, S.-H., Drobny, G. and Reid, B. R. (1983) *J. Mol. Biol.*, **171**, 319–336.
21 Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.
22 Spielmann, H. P., Wemmer, D. E. and Jacobsen, J. P. (1995) *Biochemistry* (submitted).
23 Assa-Munt, N., Leupin, W., Denny, W.A. and Kearns, D. R.(1985) *Biochemistry*, **24**, 1441–1449.
24 Assa-Munt, N., Leupin, W., Denny, W.A. and Kearns, D. R. (1985) *Biochemistry*, **24**, 1449–1460.
25 Gilbert, D. E., Van der Marel, G. A., Van Boom, J. H. and Feigon, J. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 3006–3010.
26 Gilbert, D. E. and Feigon, J. (1991) *Biochemistry*, **30**, 2483–2494.
27 Gilbert, D. E. and Feigon, J. (1992) *Nucleic Acid Res.*, **20**, 2411–2420.
28 Gao, X. and Patel, D. J. (1989) *Q. Rev. Biophys.*, **22**, 93–138.
29 Addess, K. J., Gilbert, D. E., Olsen, R. K. and Feigon J. (1992) *Biochemistry*, **31**, 339–350.
30 Addess, J. and Feigon, J. (1994) *Biochemistry*, **33**, 12386–12396.
31 Addess, J. and Feigon, J. (1994) *Biochemistry*, **33**, 12397–12404.
32 Searle, M. S., Hall, J. G., Denny, W. A. and Wakelin, P. G. (1989) *Biochem. J.*, **259**, 433–441.
33 Zhang, X. and Patel, D. J. (1991) *Biochemistry*, **30**, 4026–4041.
34 Gallego, J., Luque, F. J., Orozo, M., Burgos, C., Alvarez-Builla, J., Rodrigo, M.M. and Gago, F. (1994) *J. Med. Chem.* **37**, 1602–1609.
35 Gallego, J., Luque, Ortiz, A. R. and Gago, F. (1993) *J. Med. Chem.* **36**, 1548–1561.
36 Delepierre, M., Maroun, R., Garbay-Jaureguiberry, C., Igolen, J. and Roques, B. P. (1989) *J. Mol. Biol.*, **210**, 211–228.
37 Maroun, R., Delepierre and M., Roques, B. P. (1989) *J. Biomol. Struct.*, **7**, 607–621.
38 Delepierre, M., Dinh, T. H. and Roques, B. P. (1989) *Biopolymers*, **28**, 2115–2142.
39 Delepierre, M., Mihle, C., Namane, A., Dinh, T. H. and Roques, B. P. *Biopolymers*, **31**, 331–353.
40 Pothier, J., Delepierre, M., Barsi, M.-C., Garbay-Jaureguiberry, C., Igolen, J., Le Bret, M. and Roques, B. P. (1991) *Biopolymers*, **31**, 1309–1323.
41 Williams, L. D and Gao, Q. (1992) *Biochemistry*, **31**, 4315–4324.