^{1}H NMR investigations on the solution structure of the oligonucleotide 5′-d(ACCT$_5$GATGT)-3′/5′-d(ACATCA$_5$GGT)-3′ and its interaction with tallimustine

Enzio Ragg,*[a] Stefania Mazzini,[a] Rita Bortolini,[a] Nicola Mongelli[b] and Roberto D'Alessio[b]

[a] *Dipartimento di Scienze Molecolari Agroalimentari, Sezione di Chimica, Università degli Studi di Milano, Via Celoria 2, 20133 Milano, Italy*
[b] *Pharmacia & Upjohn, Therapeutic Area Oncology and Immunology, Chemistry Department, Viale Pasteur 10, 20014 Nerviano, Italy*

The conformation of the DNA duplex 5′-d(ACCT$_5$GATGT)-3′/5′-(ACATCA$_5$GGT)-3′ has been studied by means of 2D ^{1}H NMR and restrained molecular mechanics and dynamics calculations. This sequence contains the tract 'T$_4$GA', which has previously been proved to be the recognition sequence of tallimustine, a distamycin-A analogue, containing a nitrogen mustard moiety and with interesting anti-leukemia properties. The tridecamer has been found to possess a B-type DNA conformation with local variations in the helical parameters which may be related to the sequence. Guanine in position 9, a residue supposed to be a key point for tallimustine recognition, is at the end of a homopyrimidinic tract where the propeller twist is found to decrease down to 9° at the flanking residue A$_{10}$. Analysis of the minor groove geometry has revealed an increase of the width up to G$_9$ and a return to average values at the level of residues A$_{10}$–C$_{18}$. The reversible interaction of this oligonucleotide with tallimustine has also been investigated by NMR and UV spectroscopy. The association constant K_a is 2×10^7 dm^3 mol^{-1}. The off-rate constant is 1.5 ± 0.5 s^{-1} as measured by NOESY-exchange experiments performed at R = [drug]/[DNA] = 0.5. A multiple mode of binding of tallimustine with the oligomer has been detected by analysis of ^{1}H NMR spectra; however, the chemical shift analysis and several inter-molecular NOE interactions observed at low drug : DNA ratios have provided sufficient structural information to define the major orientation of the drug in the minor groove. The positively charged amidinic moiety is situated at the beginning of the AT-rich region and the phenyl group resides close to the T$_8$: A$_{19}$ base pair. The specificity of interaction of tallimustine depends on a favourable steric interaction at the T$_8$G$_9$A$_{10}$ tract, as a result of the widening of the minor groove allowing accommodation of the phenyl moiety in the floor of the groove. Under the conditions applied in the NMR studies, negligible alkylation has been detected by HPLC measurements. However, after nine days of incubation at T = 37 °C several peaks corresponding to alkylation products are detected giving further indications of a multiple mode of binding of tallimustine with the oligonucleotide.

Introduction

Tallimustine **1** is a derivative of distamycin-A **2**, which exhibits a broad spectrum anti-tumor activity in a series of human tumors (such as C3H mammal tumor, colon 38 adenotumor and some melphalan-resistant leukemias[1–3] and has reached phase II of clinical trials.[4] It is constituted of a benzoic acid *para*-nitrogen mustard (BAM) conjugated with a distamycin-A moiety.[5] The former group is present also in other nitrogen mustards, like melphalan, and confers to tallimustine mild alkylating properties, while the latter group provides a high affinity towards the minor groove of AT-rich regions of DNA.[6] The mode of action of tallimustine has been hypothesized to be through alkylation of purine N-3 and it has recently been found by means of footprinting experiments on plasmids pBR322 and SV40 that the consensus sequence is 'T$_4$GPu', with a preference for the 'T$_4$GA' tract.[7] Tallimustine was found to alkylate, with much lower efficiency, other tracts, like 'T$_4$AA' or 'T$_4$GG' and negative results were obtained for similar sequences like 'T$_4$CA' also present in the plasmid.[8] These findings suggested the importance of the guanine residue in the recognition. The 'T$_4$GA' sequence is present in the junction of the breakpoint of the translocation t(9 : 22) (q34; q11) of the Philadelphia chromosome in chronic myeloid leukemia and is also present at the 3′ end of the gene encoding for DNA topoisomerase α.[9,10] In view of the biological importance of this sequence and in order to provide some experimental evidence of the mode of interaction of tallimustine towards DNA, which may explain the selectivity of the drug towards the 5′-TTTTGA-3′ sequence, we conducted an NMR study of the solution structure of the tridecamer **3**. The oligomer has been designed to correspond to the 3204-3213 coding sequence of plasmid

1 R =

2 R = H

3 5′-d(A$_1$ C$_2$ C$_3$ T$_4$ T$_5$ T$_6$ T$_7$ T$_8$ G$_9$ A$_{10}$T$_{11}$ G$_{I2}$T$_{13}$)-3′
3′-d(T$_{26}$G$_{25}$G$_{24}$A$_{23}$A$_{22}$A$_{21}$A$_{20}$A$_{19}$C$_{18}$T$_{17}$A$_{16}$C$_{15}$A$_{14}$)-5′

pBR322 (5′-ATCA$_5$GG-3′/3′-CCT$_5$GAT-5′) and to the consensus sequence of tallimustine. The designed sequence contains at the ends some minor differences compared with the pBR322 sequence. These were introduced retaining the purine–pyrimidine alternation in order to avoid the introduction of minor binding sites, which may also be detected at the high concentrations required for the NMR experiments. We also wish to report here our results on the reversible and irreversible interactions between **3** and the drug, as observed in ^{1}H NMR and HPLC studies.

Experimental

Tallimustine **1** and the oligonucleotides corresponding to the two single strands 5′-d(AC$_2$T$_5$GATGT) and 5′-d(ACATCA$_5$-G$_2$T) sodium salts **3** were kindly supplied by Pharmacia & Upjohn (Nerviano, Italy). About 5 mg of each strand were dissolved in 0.3 cm^3 of a water solution containing 10% D_2O, 0.1 mol dm^{-3} NaCl, 1 mmol dm^{-3} NaN_3 and 10 mmol dm^{-3} phosphate buffer (pH 6.7). Concentrations were estimated by UV spectroscopy. The two solutions were mixed at room temperature in order to produce 0.5 cm^3 of a solution containing a 1 : 1 molar ratio of the two strands. Excess of either strands were subsequently checked by ^{1}H NMR spectroscopy and corresponding aliquots of the strand in defect were added until no residual signals were observed for the strand in excess. Because of the very low solubility in water of tallimustine, a stock solution in ethanol was prepared (about 1 mg cm^{-3}) and the concentration was determined by UV spectroscopy. The DNA–tallimustine complex was prepared by adding the calculated amount of tallimustine as solid or as an ethanolic solution to an empty NMR tube; in the latter case the ethanol was evaporated with a nitrogen gas flow and the oligonucleotide solution added directly to the NMR tube.

NMR spectroscopy

Sequential assignments were performed by performing 2D NOESY[11,12] and TOCSY[13] experiments both in H_2O–D_2O (90 : 10 v/v) and D_2O solutions using a BRUKER AMX600 NMR spectrometer, equipped with a selective 5 mm ^{1}H probe. Chemical shifts (δ) were measured in ppm, using as a reference the residual water signal set at δ 4.75. Typically, 1024 × 4K complex FIDs were acquired and transformed after apodization with a 90° shifted sinebell-squared and zero filling to 2048 × 2 K real data points. Both dimensions were baseline corrected using a fifth-order polynomial. Water suppression was achieved by either a long low-power transmitter pulse or by substituting all the 90° pulses of a standard NOESY pulse sequence with 1–1 'echo pulse' sequences, including a 2 ms homospoil pulse.[14] This latter sequence was used for the study of the exchangeable imino protons and for measuring exchange rates in the presence of tallimustine. TOCSY experiments were performed with a spin-lock value of 60 ms. Quantitative NOESY experiments were performed in D_2O solution measuring at room temperature in a single run four spectra with mixing times of 50, 100, 150 and 300 ms. Saturation transfer experiments were performed at 15 °C using a '1–1 echo' read pulse preceded by a selective decoupler pulse of 200 and 500 ms for saturation of the water resonance. Saturation transfer was quantified by difference spectroscopy, using as a reference an analogous spectrum acquired with a 1 ms saturation pulse. Off-line data processing (Fourier transformation, volume integration and distance constraints calculations) was performed using IRMA and FELIX softwares included in the NMRchitect (v.2.0) module of INSIGHTII&DISCOVER (v.2.3.5 Biosym Technologies, San Diego, USA).

UV and HPLC studies

Tallimustine concentration was determined in ethanol using as molar extinction coefficent ε = 65 800 dm^3 mol^{-1} cm^{-1}. Oligonucleotide concentrations were estimated by UV absorbance, using molar extinction coefficients calculated from the single strand compositions (ε = 118 800 and 139 900 dm^3 mol^{-1} cm^{-1} for 'T$_5$GA' and 'TCA$_5$' respectively). Association constants were determined by means of the oligonucleotide melting temperature shift, measured upon binding with tallimustine. The melting of the DNA oligomer was followed through UV absorbance using a Beckman DU-50 spectrophotometer equipped with a thermostatted cell. HPLC analysis was performed using a Hewlett Packard HPLC equipped with Rheodyne injector (0.02 cm^3 loop), diode array and an RP-18 reverse phase column (Merck Lichrosorb 5 μm with 250 mm length and 4 mm inner diameter) with the following elution gradient program. Solvent A: water, KH_2PO_4 (10 mmol dm^{-3}); solvent B: CH_3CN; flow rate : 1 cm^3 min^{-1}; gradient: from 5 to 60% B in 25 min at 24 °C. Detector set at λ of 260 and 314 nm.

Computer-aided molecular modeling

All MM and MD calculations were performed using the DISCOVER program running on a Silicon Graphics 4D35GT workstation. Model geometries were generated starting from both canonical A-DNA and B-DNA structures using the fragment library. Distance restraints were generated using at first the 'two-spin approximation' using peak volumes measured in the 50 ms NOESY experiment and subsequently refined using three cycles of IRMA calculations combined with a 30 ps MD simulation at 300 or 500K. During restrained MM and MD a force constant of 41.8 kJ mol^{-1} Å^{-2} was applied for all distance restraints. This would produce a pseudo-energy contribution of 4.18 kJ mol^{-1} when the model distance differs 0.31 Å from the target. IRMA (iterative relaxation matrix analysis) calculations were performed with uniform weighting coefficients for the cross-peaks volumes measured at all mixing times, assuming an isotropic correlation time of 3.5 ns and a default T_1 leakage of 1 s.[15] For energy minimization and MD calculations both CVFF and AMBER forcefields were used; in the latter case a scaling factor of 0.5 was used for 1–4 interactions. In our case AMBER proved to generate structures with a more satisfactory agreement with the experimental restraints. However, the CVFF force field was chosen for the subsequent 'docking' calculations. In all cases a distance-dependent relative permittivity $\varepsilon = 4.0r$ was used to simulate solvent effects. In the case of tallimustine, atomic charges were calculated from a MOPAC calculation of an energy minimized structure using the CVFF force field. Helical parameters were calculated using the program NEWHEL90 included in the InsightII& Discover library. Signs of propeller twists were changed from negative to positive values in order to comply to the IUPAC rules.

Results and discussion

^{1}H NMR assignments of the oligomer 3

Sequential assignments were accomplished by applying well established procedures for the analysis of double stranded oligomers in the B-DNA form.[16] TOCSY experiments provided the easy detection of cytidine 5-H/6-H and of thymine 5-Me/6-H spin systems as well as all intra-ribose connectivities. The analysis of aromatic/1′-H [see Fig. 1(*a*) and (*b*)] and 1′-H/5-Me of thymine regions of the NOESY spectra provided the sequential connectivities. In fact, each aromatic proton displayed two NOESY cross-peaks, of which the stronger one provided the intra-residue connectivity and the weaker one, mostly due to spin-diffusion, provided the inter-residue connectivity. The T$_4$–T$_8$ protons resonated in an overlapped region. However, aromatic, methyl and 1′-H protons of the T$_4$ and T$_8$ residues, at the beginning and at the end of the 'T' tract, were identified by their sequential interactions with C$_3$ and G$_9$ protons. Although the 6-H aromatic protons of the T$_4$–T$_6$ tract were found coincident, it was nevertheless possible to make a complete sequential

assignment for this tract making use of the slight differences in chemical shift of the 1′-H protons. Similar sequential assignments were obtained by analysing the aromatic/2′, 2″-H region. All the 3′-H and 4′-H ribose protons were assigned analysing the 1′-H/3′-H, 1′-H/4′-H and aromatic/3′-H regions. 5′ and 5″-H protons were assigned, where possible, on the basis of the connectivities observed in the TOCSY spectrum with the 1′-H protons and confirmed with the interactions found in the NOESY spectra.

The δ 11–13 region of the NMR spectra performed in H_2O–D_2O (90:10 v/v) solution displayed well resolved resonances corresponding to the imino protons of the individual hydrogen bonded base pairs. These resonances were better observed using '1–1 echo' based experiments, because this pulse sequence produces a negligible transfer of saturation between water and the exchangeable protons. Nevertheless, only 11 out of the 13 expected signals were observed even at low temperature. The two missing resonances were therefore attributed to imino protons of the first and last base pairs, namely $A_1:T_{26}$ and $A_{13}:T_{14}$, and the lack of their resonances was interpreted as having been caused by an extensive 'fraying' of the terminal parts of the double helix. Two other resonances, observed in the spectral region typical for C:G base pairs, were found to possess larger linewidths and reduced intensities and were better observed at lower temperature. They have subsequently been identified as the imino protons of the second terminal $C_2:G_{25}$ and $G_{12}:C_{15}$ base pairs. The sequential assignment of the exchangeable imino protons was accomplished by the through-space interactions found with neighbour imino protons; in particular, the interaction between C_3, resonating at fields higher than thymine imino protons, was used as the starting point for the sequential assignment. The assignments for 2-H protons were performed by means of the NOESY interactions observed with the imino protons in the same AT base pairs.[17] In addition, we found intra-strand sequential interactions between the aromatic 2-H protons of adenines and the 1′-H of the 5′ neighbouring residue in the A_{19}–G_{24} tract. Moreover, other inter-strand interactions between 1′-H in the T_5–T_8 tract and 2-H of the 3′-neighbouring adenines on the complementary strand were detected. All these inter-strand interactions gave valuable information about the groove width at the level of the 'AT' rich region of the oligonucleotide. NOESY experiments performed in D_2O solution at longer mixing times (300–500 ms) also provided other aromatic–aromatic sequential interactions, which proved to be useful for an internal check of the sequential assignment. The ^{1}H chemical shift values of aromatic, ribose, amino and imino protons are reported in Tables 1–3.

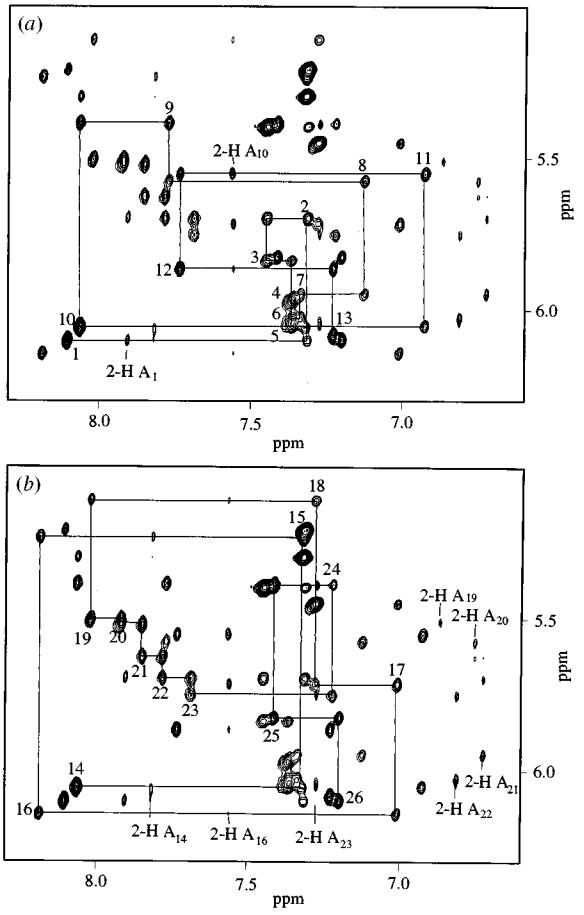


Fig. 1 Sequential assignments of oligomer **3**. Expanded region of a 2D NOESY experiment performed with a 300 ms mixing time in D_2O buffer solution (pH 6.7), 0.1 mol dm^{-3} NaCl, 1 mmol dm^{-3} NaN_3 at 20 °C. The NOE connectivities between aromatic protons and their own sugar and 5′-flanking sugar 1′-H protons are traced (*a*) A_1 to T_{13} for 'T_5GA' strand; (*b*) A_{14} to T_{26} for 'TCA_5' strand. In the pictures are indicated the 2-H protons of adenines.

Structure determination of the oligomer 3

After the sequential assignment, a set of 640 cross-peak volumes was measured for each of the NOESY experiments performed with mixing times ranging from 50 to 300 ms. The 'T'-rich tract of the oligomer was the most difficult to measure because of the extensive overlapping of the cross-peaks. Since T_5 and T_6 protons were found to be coincident, it was decided not to include any measured cross-peak volume, except those arising from the neighbouring 2-H adenine protons, because the corresponding measured distances would be meaningless. It may therefore be anticipated that the conformation of the T_5 and T_6 residues is mostly determined by the applied force-field; however, as will be seen later, some inter-strand distances were nevertheless measured even for these two residues leading to an accurate measurement of the minor-groove width. From the cross-peak volumes an initial set of 340 interatomic distances was derived using the 'two-spin approximation', by

Table 1 ^{1}H NMR chemical shift assignments (δ) for the 'T_5GA' strand[a]

Residue	8-H, 2-H	6-H, 5-H	NH_2[b]	1′-H	2′,2″-H	3′-H	4′-H	5′,5″-H
A_1	8.11, 7.91		n.d.	6.10	2.71, 2.56	4.70	4.13	3.62, 3.62
C_2		7.32, 5.21	7.97, 6.57	5.70	2.29, 2.06	4.68	4.10	4.05, 3.95
C_3		7.45, 5.40	8.11, 6.78	5.84	2.41, 2.05	4.61	4.10	4.02, 3.99
T_4		7.37, 1.45		5.97	2.50, 2.12	4.76	4.11	3.97[c]
T_5		7.37, 1.47		6.04	2.47, 2.12	4.76	4.19	3.97[c]
T_6		7.37, 1.47		6.03	2.48, 2.08	4.76	4.20	3.97[c]
T_7		7.34, 1.47		5.95	2.46, 2.01	4.77	4.05	*d*
T_8		7.13, 1.49		5.58	2.26, 1.87	4.74	3.96	*d*
G_9	7.78			5.39	2.26, 2.57	4.86	4.22	3.96, 3.96
A_{10}	8.07, 7.57		7.44, 5.79	6.05	2.75, 2.47	4.86	4.30	4.05, 3.97
T_{11}		6.93, 1.21		5.56	2.13, 1.74	4.68	4.07	3.96, 3.96
G_{12}	7.74			5.87	2.55, 2.48	4.80	4.22	4.08, 4.03
T_{13}		7.23, 1.37		6.09	2.09, 2.09	4.76	4.39	3.92, 3.92

[a] Measured at 20 °C. [b] Measured at 15 °C. [c] Estimated accuracy δ 0.04. [d] Not assigned because of extensive overlapping

Table 2 ^{1}H NMR chemical shift assignments (δ) for the 'TCA$_5$' strand[a]

Residue	8-H, 2-H	6-H, 5-H	NH_2[b]	1′-H	2′,2″-H	3′-H	4′-H	5′,5″-H
A_{14}	8.07, 7.82		n.d.	6.06	2.67, 2.50	4.71	4.11	3.58, 3.58
C_{15}		7.32, 5.30	8.06, 6.58	5.24	2.27, 2.04	4.70	4.03	4.05, 3.95
A_{16}	8.19, 7.56		7.47, 6.18	6.14	2.80, 2.56	4.88	4.30	4.04, 3.96
T_{17}		7.01, 1.23		5.72	2.22, 1.81	4.68	4.14	3.98, 3.98
C_{18}		7.28, 5.45	8.33, 6.55	5.11	2.04, 1.73	4.63	4.25	3.88, 3.88
A_{19}	8.02, 6.87		7.44, 6.01	5.51	2.62, 2.53	4.85	4.18	3.91, 3.80
A_{20}	7.93, 6.76		7.21, 5.65	5.52	2.62, 2.42	4.85	4.24	4.01, 3.99
A_{21}	7.85, 6.73		6.99, 5.61	5.63	2.70, 2.37	4.85	4.25	4.05, 4.02
A_{22}	7.78, 6.81		6.99, 5.72	5.70	2.74, 2.33	4.85	4.25	4.08, 4.04
A_{23}	7.69, 7.28		7.18, 5.78	5.76	2.67, 2.28	4.84	4.25	4.08, 4.04
G_{24}	7.23			5.39	2.43, 2.23	4.80	4.21	4.03, 4.03
G_{25}	7.42			5.83	2.52, 2.33	4.71	4.20	4.07, 4.02
T_{26}		7.20, 1.24		6.10	2.10, 2.10	4.71	4.40	3.92, 3.92

[a] Measured at 20 °C. Estimated accuracy δ 0.01. [b] Measured at 15 °C.

Table 3 ^{1}H NMR chemical shift assignments (δ) of the imino protons measured in the absence and presence of tallimustine and solvent exchange rate constants (k_w/s^{-1}) for oligomer **3**[a]

	Free oligomer		Bound oligomer	
Base pair	NH	k_w/s^{-1}	NH	ΔNH ($\delta_{bound} - \delta_{free}$)
A_1:T_{26}	n.d.	n.d.	n.d.	n.d.
C_2:G_{25}	12.77	3.3	12.77	0.00
C_3:G_{24}	12.81	1.7	12.81	0.00
T_4:A_{23}	14.15	1.1	14.72	+0.57
T_5:A_{22}	13.93	0.9	14.38	+0.45
T_6:A_{21}	13.78	0.6	14.32	+0.54
T_7:A_{20}	13.69	0.8	13.53	−0.16
T_8:A_{19}	13.47	1.9[b]	12.93	−0.54
G_9:C_{18}	12.25	0.8	12.25	0.00
A_{10}:T_{17}	13.40	2.3	13.60	+0.20
T_{11}:A_{16}	13.45	1.9[b]	n.d.	*c*
G_{12}:C_{15}	12.46	3.9	12.46	0.00
T_{13}:A_{14}	n.d.	n.d.	n.d.	n.d.

[a] Measured at 15 °C, estimated accuracy ±0.1 s^{-1}. [b] Nearly isochronous; their average value is reported. [c] Estimated value δ <0.2.

calculating the ratios with cross-peak volumes relative to protons at known distances, *e.g.* 6-H/5-H of cytidines at positions 2, 3 were used as a reference (2.45 Å). This set was used as a starting point for the IRMA calculation,[18] which has been proved to provide reliable distance restraints since, based on a full relaxation matrix approach, it takes into account spin-diffusion and thus it is able to set reliable upper and lower limits for the restraints. An isotropic correlation time of 3.5 ns was assumed, although for such long oligonucleotides anisotropic motions may in principle become important. However, it has been proved[15] that distance errors induced by taking into consideration isotropic motions are in practise less than 0.3 Å for protons oriented parallel to the long axis of the double helix. The distance restraints were used in restrained MM energy minimization followed by a 30 ps MD simulation performed at 500 K, sampling the trajectory every 2 ps. In order to avoid breakage of the hydrogen bonds between the complementary base pairs during the MD calculations, a set of loose restraints for hydrogen bonds (upper and lower limits set respectively at 3.0 and 1.8 Å) was introduced. A new averaged structure was generated from the last five sampled structures and employed for the next two IRMA cycles. After the second run of IRMA calculations some distances were still found to give restraint violations greater then 0.8 Å. They corresponded to poorly resolved cross-peaks and unreliable volume measurements and were therefore deleted from the restraints set. The final set thus contained a total of 330 values. Two sets of 12 conformations were then generated by two sequential MD runs performed at 500 and 400 K, sampling a geometry every 2 ps and energy minimizing the sampled structures. Motional averaging was taken into account in this way during IRMA calculations by considering the last generated set of 12 minimized structures. The same procedure was applied starting on models based both on the canonical A-type and B-type DNA geometries and converged to very similar structures. In fact, a superposition of the A-type and B-type starting structures produced an rms deviation on heavy atoms of 7 Å, whereas the final structures, generated after three iterations of IRMA, gave an rms deviation equal to 1 Å. Fig. 2(*a*) and (*b*) show the distribution of the experimentally derived NOE restraints for individual residues and the distribution of distance restraint violations exceeding 0.4 Å found for the refined structure. Only 12 violations over 330 restraints were found between 0.5 Å and 0.65 Å. Fig. 2(*c*) shows the superposition of the 12 conformations derived from the last MD and MM run.

Tables 4 and 5 list the mean and variance values for the backbone torsion angles derived from the final 12 structures. There is general agreement between the calculated mean values of either strands and the torsional angles defining the canonical B-DNA structure. In particular torsion angles δ, exocyclic with respect to the ribose rings, are a measure of the ring puckering and are correlated with the glycosidic angles χ (anti-correlation principle).[19] The 'T$_5$GA' strand, rich in pyrimidine nucleosides, displays δ values, 114° on average for C$_2$–T$_8$ tract (C1′-*exo* conformation of ribose) coupled with angles χ between 187 and 260° ('anti' orientation of the base). The opposite 'TCA$_5$' strand, rich in purine nucleosides, shows higher δ values, 129° on average for the A$_{19}$–G$_{25}$ tract (C2′-*endo* ribose conformation), coupled to χ angles between 226 and 285° ('high anti' orientation of the base). This feature has already been observed in the crystallographic structure of d(CGCGAATTCGCG),[20] where the δ torsion angles tend to cluster in two gaussian distributions centered at 110 and 135°, respectively, for pyrimidines and purines. At the same time, glycosyl torsion angles χ were found, in agreement with quantum chemical calculations,[19] to strongly correlate with sugar puckering. The phosphate geometry at the ε,ζ torsion angle pairs does not show significant differences from standard B-DNA. These torsion angles are correlated and their difference ($\varepsilon - \zeta$) can assume two values, −90 and +90°, corresponding respectively to the B$_I$ and B$_{II}$ conformations.[21] The 'TCA$_5$' tract assumes entirely the B$_I$ conformation, which is in general energetically favoured, whereas the complementary tract 'T$_5$GA' shows two exceptions at the level of T$_8$ and G$_9$ bases, where both B$_I$ and B$_{II}$ conformations were found with equal probability in the minimized structures. The other phosphate torsion angles do not show any appreciable difference with standard values. β angles display values slightly higher than the canonical ones, but are compatible with those found in other oligonucleotide X-ray structure, like the Dickerson's dodecamer d(CGCGAATTCGCG).[20]

Thus, the overall structural feature of the oligomer is that of a typical B-type DNA. The average helical twist angle is 34°, corresponding to 10.6 residues per helix. However, more significant than the average helical parameters are the local

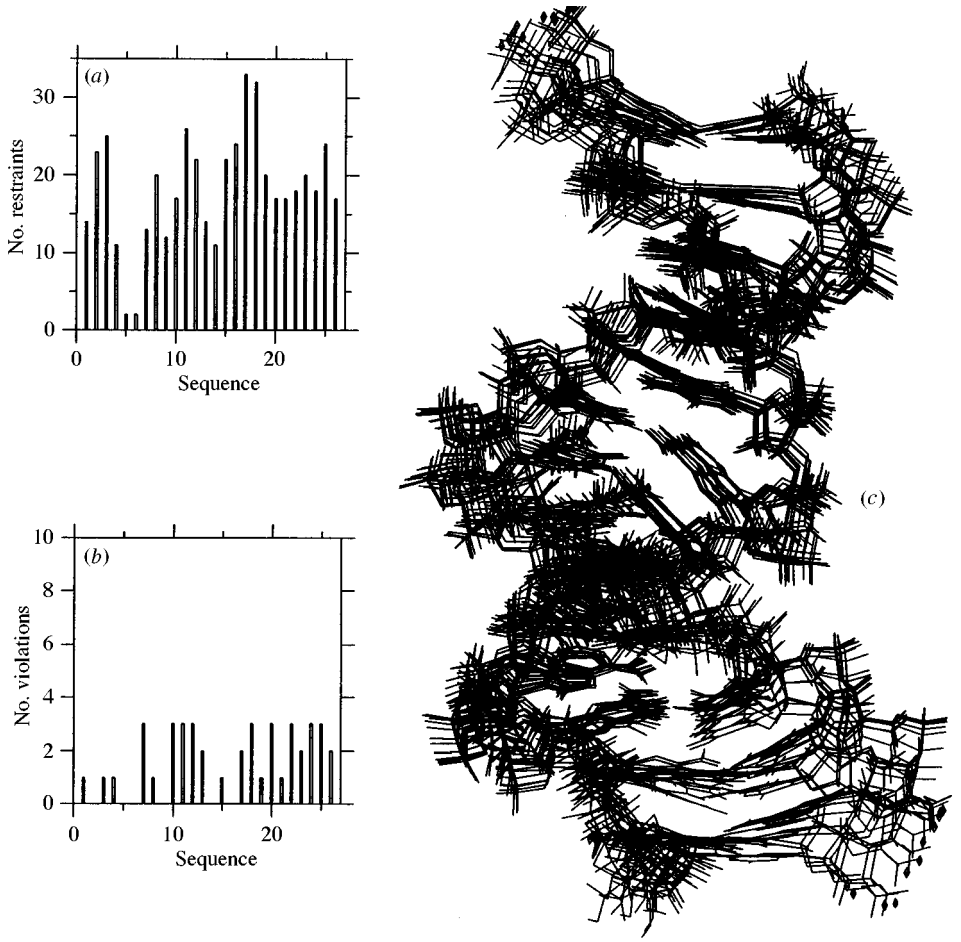


Fig. 2 (*a*) Distribution of the experimental NOE restraints derived for individual residues in the structure calculation of oligomer **3**. (*b*) Distribution of distance restraint violations exceeding 0.4 Å found in the refined structure. (*c*) Superposition of the 12 structures derived from MD calculations.

Table 4 Backbone torsional angles (°) obtained for the 'T_5GA' strand of oligomer **3**[a]

Residue	χ	α	β	γ	δ	ε	ζ
A_1	240 ± 16			60 ± 2	123 ± 17	196 ± 5	286 ± 7
C_2	249 ± 4	268 ± 6	172 ± 8	50 ± 5	94 ± 15	180 ± 2	261 ± 12
C_3	218 ± 2	270 ± 4	178 ± 2	60 ± 1	106 ± 1	212 ± 3	307 ± 2
T_4	187 ± 2	278 ± 2	168 ± 1	67 ± 2	77 ± 1	184 ± 3	282 ± 3
T_5	244 ± 2	296 ± 1	179 ± 1	67 ± 1	136 ± 2	179 ± 1	260 ± 1
T_6	250 ± 12	292 ± 1	181 ± 1	58 ± 1	131 ± 2	183 ± 1	260 ± 3
T_7	242 ± 4	288 ± 3	179 ± 1	58 ± 2	115 ± 8	183 ± 3	270 ± 1
T_8	260 ± 7	290 ± 6	176 ± 2	53 ± 1	140 ± 5	279 ± 2 186 ± 6	190 ± 5 251 ± 5
G_9	270 ± 7	286 ± 5	150 ± 32	46 ± 8	137 ± 9	274 ± 1 185 ± 6	166 ± 12 246 ± 2
A_{10}	239 ± 1	291 ± 3	144 ± 27	54 ± 7	113 ± 12	177 ± 3	268 ± 1
T_{11}	246 ± 1	297 ± 2	178 ± 1	56 ± 1	130 ± 3	176 ± 1	264 ± 1
G_{12}	246 ± 1	291 ± 2	187 ± 1	49 ± 1	126 ± 1	168 ± 1	271 ± 1
T_{13}	242 ± 1	290 ± 1	184 ± 1	58 ± 1	135 ± 1		
Mean	241 ± 20	286 ± 10	173 ± 13	57 ± 6	120 ± 19	184 ± 12[b]	273 ± 15[b]
B-DNA[c]	258	319	136	38	139	227	203

[a] Torsional angles are defined as follows: C5′-γ-C4′-δ-C3′-ε-O3′-ζ-P-α-O5′-β-C5′. Average and variance values calculated on the basis of the 12 structures derived from MD calculations. [b] The mean was calculated excluding T_8 and G_9 residues. For these residues two values for ε and ζ angles were found with equal probability. [c] Taken from ref. 27.

parameters, which displayed values that could be explained by the Dickerson–Calladine rules for sequence-dependent variations.[22,23] It is now well established that the geometry of the minor groove modulates the mode of interaction of minor groove binders. The minor groove width is defined by the distances, $d_{C\text{-}1',C\text{-}1'}$, between C-1′ of each ribose and the C-1′ on the 3′-neighbouring residue of the complementary strand. In the final structure these distances were found to progressively increase in the 5′ to 3′ direction of the 'T_5GA' tract (Table 6). The experimental evidence for such variation in the minor groove comes from the NOE derived distances, $d_{2\text{-}H,1'\text{-}H}$, between the 2-H of an adenine and the 1′-H of its 3′-neighbouring residue on the complementary strand. Table 6 lists the inter-proton distances derived from the final 12 structures of the oligomer, which differ from the experimental values by 0.1 Å. A progressive increase of such distances was found in the T_6–G_9 segment. The minor groove reaches its maximum width at G_9. No interaction was found between 1′-H of G_9 and 2-H of A_{19} and their distance was therefore estimated to be greater than 5.0 Å. However, no distance restraint was defined for these residues. In the final structure this distance is 5.5 Å, a value which is significantly greater than the one found

Table 5 Backbone torsional angles (°) obtained for 'TCA$_5$' strand of oligomer **3**[a]

Residue	χ	α	β	γ	δ	ε	ζ
A_{14}	255 ± 4			60 ± 1	140 ± 1	183 ± 1	243 ± 1
C_{15}	254 ± 4	292 ± 1	173 ± 4	63 ± 1	148 ± 3	185 ± 1	230 ± 1
A_{16}	261 ± 1	276 ± 18	180 ± 11	48 ± 6	143 ± 1	175 ± 1	267 ± 1
T_{17}	253 ± 1	287 ± 1	182 ± 2	55 ± 1	117 ± 2	185 ± 1	268 ± 1
C_{18}	225 ± 1	271 ± 1	168 ± 1	61 ± 2	85 ± 1	175 ± 2	269 ± 1
A_{19}	285 ± 2	301 ± 2	182 ± 2	67 ± 1	151 ± 1	174 ± 1	265 ± 1
A_{20}	265 ± 2	294 ± 4	177 ± 1	52 ± 1	131 ± 7	174 ± 1	272 ± 4
A_{21}	264 ± 2	299 ± 2	187 ± 1	51 ± 1	138 ± 1	185 ± 1	276 ± 1
A_{22}	246 ± 1	279 ± 1	173 ± 1	42 ± 2	94 ± 1	167 ± 1	280 ± 1
A_{23}	232 ± 1	306 ± 1	168 ± 1	54 ± 1	118 ± 2	186 ± 1	263 ± 2
G_{24}	226 ± 4	273 ± 2	183 ± 2	58 ± 2	121 ± 3	204 ± 2	280 ± 3
G_{25}	257 ± 5	298 ± 1	190 ± 4	39 ± 5	149 ± 2	179 ± 1	276 ± 1
T_{26}	224 ± 1	295 ± 1	172 ± 1	39 ± 3	92 ± 1		
Mean	250 ± 18	289 ± 12	178 ± 7	53 ± 19	125 ± 23	181 ± 9	266 ± 15
B-DNA[b]	258	319	136	38	139	227	203

[a] Torsional angles are defined in Table 4. Average and variance values calculated on the basis of 12 structures derived from MD calculations. [b] Taken from ref. 27.

Table 6 Helical parameters and minor groove width obtained for the 'T$_5$GAT' tract of oligomer **3**[a]

Base pairs	Propeller twist/°	Helical twist/°	Roll/°	$d_{2\text{-H},1'\text{-H}}$[b]/Å (exp.)	$d_{2\text{-H},1'\text{-H}}$[b]/Å (model)	$d_{C\text{-}1',C\text{-}1'}$[b]/Å (model)
$T_4:A_{23}$	15 ± 2	28 ± 2	−1 ± 2			7.2
$T_5:A_{22}$	15 ± 3	29 ± 1	2 ± 4	4.8–5.7	4.8	9.3
$T_6:A_{21}$	16 ± 4	37 ± 1	16 ± 1	4.0–4.1	4.1	9.1
$T_7:A_{20}$	36 ± 3	36 ± 1	−7 ± 2	4.2–4.5	4.2	8.6
$T_8:A_{19}$	17 ± 1	35 ± 1	−4 ± 1	4.8–5.1	4.7	9.7
$G_9:C_{18}$	18 ± 1	38 ± 1	−9 ± 3	*c*	5.5	11.0
$A_{10}:T_{17}$	9.0 ± 2	32 ± 1	−5 ± 1	4.9–5.1	4.9	8.9
$T_{11}:A_{16}$	9.0 ± 2	40 ± 1	6 ± 3			9.0
Mean	17 ± 8	34 ± 4	−0.25 ± 8	4.5 ± 0.4 4.9 ± 0.7	4.7 ± 0.5	9.1 ± 1.1
B-DNA[d]	0	36	0		5.0	9.2

[a] Average and variance values calculated on the basis of 12 structures derived from MD calculations. [b] Distances between 2-H (or C-1′) of an adenine and the 1′-H (or C-1′) of its 3′-neighbouring residue on the complementary strand (upper and lower limits were calculated by IRMA). [c] Not observed. [d] Torsion angles values are taken from ref. 27. $d_{2\text{-H},1'\text{-H}}$, $d_{C\text{-}1',C\text{-}1'}$ were meausured from a canonical B-DNA structure.

at the initial T_6–T_8 sequence (4.1–4.7 Å). Similar behaviour was found in the oligo(dT) tract of d(CGCAAAAAAGCG)–d(CGCTTTTTTGCG), where this opening of the minor groove causes a gradual compression of the double helix from 3′ to 5′ residues, with the result that the DNA is bent.[24,25] The increase in the minor groove width is also reflected by a slight increase in the helical twist starting from 28° at the level of $T_4:A_{23}$ and reaching a value of 38° at $G_9:C_{18}$ (Table 6). The solvent exchange rate constants (k_w) of the imino protons, measured with the saturation transfer experiments (Table 3), provided another experimental evidence for a distortion in the minor groove geometry at the end of the 'T-rich' tract. The imino protons in the internal tract $T_4:A_{23}$–$T_7:A_{20}$ and in $G_9:C_{18}$ displayed progressively lower k_w values, between 1.1 and 0.6 s^{-1}, in comparison with the external ones. An exception is represented by the $T_8:A_{19}$ base pair. Unfortunately, the chemical shift of its NH was found to be almost coincident with the NH of $T_{11}:A_{16}$. As a consequence the value of k_w measured for $T_8:A_{19}$ might be affected from the value measured for the external $T_{11}:A_{16}$ base pair. The value of 1.9 s^{-1} is in fact comparable to the value of 1.7 s^{-1} found for $C_3:G_{24}$ and is indicative of the external position, near the fraying end of these base pairs. On the other hand, the internal $A_{10}:T_{17}$ base pair is found to have a faster exchange with the solvent (2.3 s^{-1}) in contrast with 1.1 s^{-1} found for $T_4:A_{23}$. The high value found for $A_{10}:T_{17}$ might therefore be due to the increase in minor groove width and to the concomitant increase in helical twist value observed between $G_9:C_{18}$ and $A_{10}:T_{17}$ base pairs. Interestingly a trend in chemical shift was observed for the A:T base pairs of the oligomer **3**. There is an up field shift going from $T_4:A_{23}$ (δ 14.15) to $T_{11}:A_{16}$ (δ 13.45) which may be associated with the opening of the minor groove.

The distances between the 2-H protons of adjacent base pairs, $d_{2\text{-H},2\text{-H}}$, provided an additional index for the minor groove compression, because such a compression induces an approach of the 2-H protons of consecutive adenines and at the same time a spacing out of the 8-H protons residing in the opposite major groove. The corresponding helical parameter is the roll angle, defined as the angle between adjacent base pairs along the helix. The roll is positive, if the roll angle between the base pair planes opens towards the minor groove, and negative if it opens towards the major groove. The observed decrease of $d_{2\text{-H},2\text{-H}}$ from A_{23} (4.2 Å) to A_{19} (3.5 Å) may be related to a trend observed for the roll angle towards negative values (Table 6). The roll angle was measured equal to −9 ± 3° at the level of $G_9:C_{18}$. Interestingly, the phosphate of G_9 in the final structure is found to exist also in a B_{II} conformation. Although B_{II} conformations are less energetically favoured than B_I, they are in general associated with a negative roll angle,[21] which was found in the final structure of the oligomer. While a roll angle is a measure of the relative orientation of planes between successive base pairs, propeller twist is a measure of the angle formed between the hydrogen-bonded base planes. Propeller is positive for an anti-clockwise rotation of the nearer base in a view down the long axis. In the canonical B-DNA structure propeller twist is zero, whereas the tridecamer is characterized by an average propeller twist in the homopyrimidine tract equal to 16°. The value found for $T_7:A_{20}$ (36 ± 3°) is unusually high and may be due to the smaller number of interatomic distances measured for the T_6 and T_7 bases. Although the propeller twist allows a more efficient base-stacking in the homopyrimidinic tract, it causes in the alternating T_8–G_9 tract the slight overlapping of purines G_9 and A_{19}, residing in opposite strands. This purine–purine steric clash can be relieved by flattening the propeller

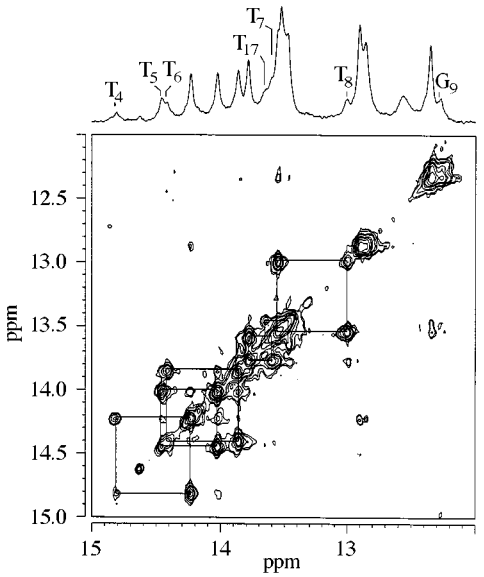


Fig. 3 2D NOESY spectrum in H_2O buffer solution (pH 6.7), 0.1 mol dm^{-3} NaCl, 1 mmol dm^{-3} NaN_3 at 25 °C of oligomer **3** in the presence of tallimustine at a drug : DNA ratio $R = 0.5$. The expansion shows cross-peaks arising from chemical exchange between imino protons in the free and bound states.

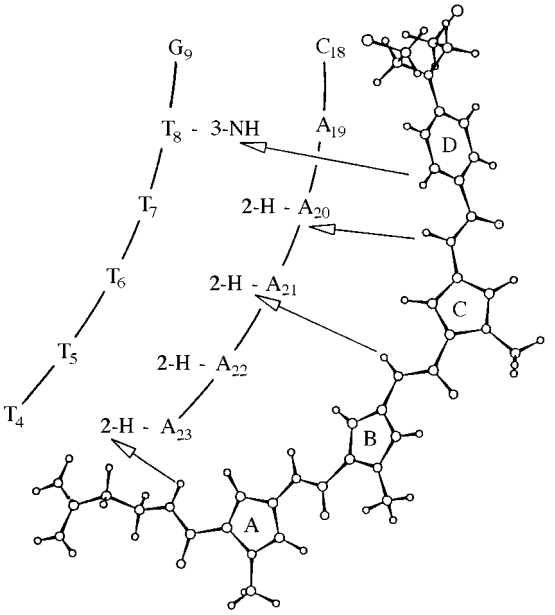


Fig. 4 Schematic representation of the complex between tallimustine and oligomer **3**. The arrows indicate the interactions derived from NOE and chemical shift analysis.

twist.[23] In the T_8–G_9–A_{10} tract the propeller twist was found to decrease to a value of 9° for A_{10} and T_{11} (Table 6).

^{1}H NMR and UV studies on the tallimustine binding

Under the conditions of the NMR experiments, tallimustine was found by HPLC to alkylate the oligomer very slowly (see following section). Thus, it was possible to study only the reversible interaction of tallimustine by NMR spectroscopy. The addition of increasing amounts of tallimustine to the oligonucleotide solution caused the splitting of many resonances of the oligomer. Fig. 3 shows the imino protons region of a 2D NOESY exchange experiment performed on a solution of the oligonucleotide in the presence of tallimustine at a drug : DNA ratio of 0.5. Under these conditions the imino proton resonances in the bound and free chemical environments were both present and displayed separate signals, as they are in slow exchange with respect to the NMR timescale. It was thus possible to assign the imino proton resonances also in the complex by means of the exchange cross-peaks (Table 3). The most perturbed signals reside in the 5′-TTTTT-3′ tract of the oligomer. The first three thymine imino protons experience a very low field shift ($\Delta\delta = +0.5$), characteristic of minor-groove binding. Note that the imino proton located at the T_8 : A_{19} base pair experiences an opposite shift at higher fields ($\Delta\delta = -0.5$). Presumably, tallimustine binds to the minor-groove and positions the positively charged amidinic moiety at the beginning of the 'AT'-rich region, close to C_3 residue, in a region known to possess a particularly negative electrostatic potential.[28] Consequently, the phenyl group resides close to the T_8 : A_{19} base-pair and can perturb the NH resonance by an up field ring-current shift.

As described for the oligomer in the free state, the 2-H aromatic protons of the adenines were assigned in the complex by means of their NOE interactions with the imino protons of the bound oligomer. Since the analysis of the NOESY spectrum was performed at a drug : DNA ratio equal to 0.5, many cross peaks were found to correlate the imino protons of the oligomer in the free species with 2-H of adenines in the bound species and *vice versa*. The origin of these peaks is due to a transferred NOE[29] caused by the presence of chemical exchange. The adenine 2-H protons reside in a key position within the minor groove, as many NOE interactions involving drug protons can be found, giving a detailed description of the drug orientation within the minor groove. The amide protons of tallimustine resonate between δ 8.8 and 10, in a characteristic region of the ^{1}H NMR spectrum free from oligonucleotide signals. Their sequential assignment is based upon the NOE interactions observed with the neighbouring 3-H pyrrole protons, resonating in a well characterized region of the ^{1}H NMR spectrum.[30–33] Observing a molecular model of tallimustine, it can be noted that each NH interacts either with the 3-H proton of the same pyrrole-carboxyamide moiety ($d = 2.7$ Å, expected weak NOE interaction), or with the 3-H pyrrole proton residing in the preceeding moiety ($d = 2.2$ Å, expected strong NOE interaction). The sequential assignment was thus performed on the basis of the observed patterns of the alternatively strong and weak interactions involving the tallimustine NH resonances. Having established the sequential assignments, it was easy to identify the intermolecular NOE interactions existing between the tallimustine amide and adenine 2-H protons (Fig. 4). We observed the following NOEs: NH(A) with 2-H of A_{23}, NH(C) with 2-H of A_{21} and NH(D) with 2-H A_{20}. These interactions completely define the position of tallimustine within the minor groove and confirm the previously suggested orientation with the phenyl group directed towards T_8, as suggested by chemical shift analysis.

At low drug : DNA ratios the number of observed NH resonances was consistent with the formation of only one type of complex. However, up to a drug : DNA ratio equal to one, the number of detected amide resonances increased with increasing amounts of tallimustine, precluding any further analysis because of extensive broadening of all the NMR resonances (see Fig. 5). At drug : DNA ratios greater than one, the excess of tallimustine precipitated because of its low solubility in water, showing that the stoichiometry of the complex is one drug for each double helix. The high number of the NH amide resonances proved that the binding to the minor groove is not characterized by a single mode. A multiple mode of binding of tallimustine was also detected by us for the interaction with the shorter alternating sequence $d(CGTATACG)_2$,[30] while distamycin-A itself was shown to slide within the 'AT'-rich region of DNA.[33]

Tallimustine[30] and distamycin-A[34] were found to move 'in and out' from the minor groove, with the so-called 'flip-flop' mechanism, which corresponds to a head-to-tail rearrangement of the drug within the helix. This type of movement was also detected by the symmetric pattern of cleavage on both strands caused by Fe(EDTA)–distamycin[35,36] and has also been found

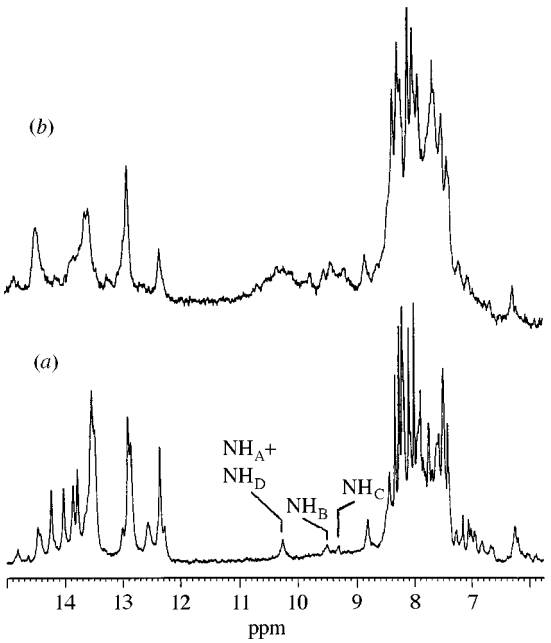


Fig. 5 Low field region of 1D ^{1}H NMR spectra in H_2O buffer solution (pH 6.7), 0.1 mol dm^{-3} NaCl, 1 mmol dm^{-3} NaN_3 at 15 °C of oligomer **3** in the presence of tallimustine at a drug : DNA ratio *R* (*a*) 0.5 and (*b*) 1.0

in some X-ray structures of oligonucleotide complexes with netropsin.[37,38] In our case, one orientation is more favourable than the other, *i.e.* the one which puts the charged amidinic group in contact with the most negative electrostatic potential at the level of the two consecutive CG base pairs, as previously described. However, at high drug : DNA ratios other non-equivalent orientations may also be populated enough to be detected by NMR spectroscopy and to produce the observed complexity of the amidic protons region and the overall exchange line-broadening.

The quantitation of the exchange cross-peaks between imino protons in the bound and free species, observed in the NOESY experiment performed at R = 0.5 (Fig. 3), lead to the measurement of the exchange rate. The ratio of NOESY cross- and diagonal-peak volumes A_{cross}/A_{diag} depends on the exchange rate constant (k_{ex}) given in eqn. (1) where t_{mix} is the mixing time

$$\frac{A_{cross}}{A_{diag}} = \frac{1 - \exp(-2k_{ex}t_{mix})}{1 + \exp(-2k_{ex}t_{mix})} \quad (1)$$

of the NOESY experiment. k_{ex} was found to be equal to 1.5 ± 0.5 s^{-1} corresponding to a lifetime of the complex equal to 0.7 ± 0.1 s. This value is relatively high but is consistent with the measurements performed on the complex of d(CGCAAATTTGCG)$_2$ with distamycin-A and analogues.[34,39]

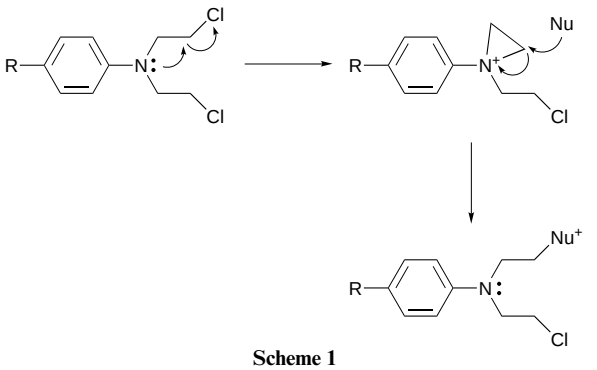
The melting temperatures of the oligomer **3**, measured in the absence and in the presence of the drug, are 41.5 and 52.5 °C, respectively. From these data, following the equations reported in ref. 40, a Van't Hoff enthalpy variation was found to be equal to 394.5 kJ mol^{-1}, in good agreement with the value (400.4 kJ mol^{-1}) predicted on the basis of the base sequence.[41] Consequently, the association constant, K_a, of 2×10^7 dm^3 mol^{-1} was obtained from the same calculations. A similar high value was measured for the interaction between distamycin-A and netropsin with d(GCGAATTCGC)$_2$, polyd(AT)–polyd(AT) and polyd(A)–polyd(T).[42] On the contrary, lower values (K_a in the order of 10^5 dm^3 mol^{-1}) were reported for tallimustine bound to the 5′-TATA-3′ site within sequence d(CGTATACG)$_2$[30] and for distamycin bound to the closely related oligomer d(GGTATACC)$_2$[43] Breslauer and co-workers found that the binding of netropsin and distamycin to the homopolymer poly(dA)–poly(dT) is dominated by the entropy term.[42] Actually, the homopolymer is more heavily hydrated, and the disruption of a water spine in the minor groove accounts for the entropic binding. Moreover, the affinity of tallimustine for sequence T_5GA does not uniquely reflect the disruption of the hydration of the minor groove but is also due to the larger width of the minor groove, which allows the bulky phenyl moiety to come into close contact with the minor groove floor.

HPLC studies on the adduct formation

A solution of the oligomer containing tallimustine at a drug : DNA ratio equal to one was allowed to incubate at 37 °C at the same concentration and ionic strength used for the NMR experiments and aliquots of the solution were analysed by reversed phase HPLC at different times. HPLC peaks eluted between 6.7 and 8 min were all identified as alkylation products by means of their characteristic UV spectra, containing both the absorption maxima of DNA and tallimustine chromophores at 260 and 314 nm, respectively. Yields in alkylation were very low, reaching 13.5% of the total peak area measured at 314 nm after nine days incubation. From these data the value of the pseudo-first-order rate constant for alkylation was estimated to be $k_{alk} = \ln(A_0/A_1)/(t_1 - t_0) = 0.63 \pm 0.15 \times 10^{-6}$ s^{-1}, A_0 and A_1 being the concentrations of the oligomer measured at t_0 and t_1, respectively. Low yields in alkylation were caused also by the competing reaction of hydrolysis, which converted tallimustine into the mono-substituted and then into the unreactive dihydroxy derivative, which might exert a competing action against alkylation, by binding itself to the minor groove. Actually, the quantitation of the hydrolysis peaks gave an apparent hydrolysis rate constant $k_{hyd} = (0.3 \pm 0.1) \times 10^{-6}$ s^{-1}, in the presence of the oligomer, whereas in its absence the measured value was 0.6×10^{-6} s^{-1}.

In preceeding studies of DNA-footprinting the kinetic reactivity of tallimustine was measured by means of incubation experiments performed with ^{14}C-labelled tallimustine and salmon sperm DNA.[44] The number of covalent adducts formed in one hour was in the order of 10^{-4} adducts per base pairs per hour. Taking into account that oligomer **3** contains 13 base pairs, the value of k_{alk} can be transformed into a value of 1.7×10^{-4} adducts per base pairs per hour, in agreement with the DNA-footprinting results.

The mechanism of alkylation of nitrogen mustards has been studied in detail[45] and it has been proved that the rate determining step is the formation of the aziridinium ion (Scheme 1),



Scheme 1

which is therefore dependent on the delocalization of the nitrogen lone-pair. In principle, the formation of the aziridinium ion can be proved by the dependence of the alkylating rate constant on the ionic strength.[46] A value of $(0.8 \pm 0.3) \times 10^{-6}$ s^{-1} was measured for k_{alk} in the absence of NaCl, but unfortunately its uncertainty did not allow us to sufficiently prove this hypothesis. However, it is well established that, within the context of the aziridinium-based mechanism, the chemical reactivity of aromatic nitrogen mustards is modulated by substituents in the *para* position. Thus the carboxamide group, being an electron withdrawing substituent, is expected to exert a deactivating

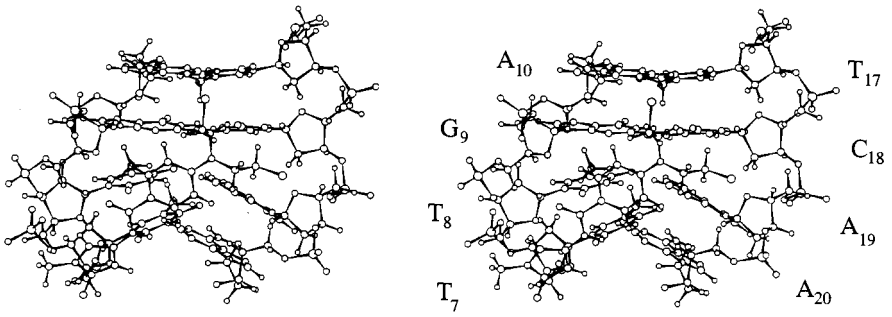


Fig. 6 Stereo drawing of an energy minimized molecular model of the complex between tallimustine and oligomer **3**

effect. Denny and co-workers[47] evaluated the chemical reactivity and the cytotoxicity of a series of DNA-targeted aniline mustards, where the alkylating groups were anchored to the DNA-intercalating 9-aminoacridine chromophore. They devised a correlation between the hydrolysis and alkylation rates (determined at 66 °C in 50% aqueous acetone) of the mustards towards model nucleophiles and the Hammett constant values of the *para* substituent. Under Denny's conditions the hydrolysis rate constants of tallimustine and phenylalanine mustard (melphalan) were measured respectively as 3.8×10^{-5} and 4.7×10^{-4} s^{-1}. Considering that the carboxamide group has a Hammett constant σ equal to 0.36, the theoretical hydrolysis rate constant values calculated on the basis of Denny's equations are between 1.2×10^{-5} s^{-1} and 32×10^{-5} s^{-1}, in line with our experimental results. Thus, the low reactivity towards alkylation found for tallimustine could be explained by the electron-withdrawing effect of the carboxyamidic group in the *para* position with respect to the nitrogen mustard.

Molecular modelling of the complex between tallimustine and oligomer 3

The oligomer conformation, determined by means of the NMR experiments and molecular dynamics calculations, was used as a starting structure for the analysis of the intermolecular interactions between tallimustine and the double helix. By means of computer graphics the drug was positioned in the 5′-TTTTTGA-3′ site, with the phenyl ring and the charged amidinic moiety positioned respectively at the level of T_8 and T_4 residues. After a 50 ps MD run at 500 K and energy minimization, keeping fixed the conformation of the double helix, tallimustine was found to interact preferentially with the 'A-rich' strand, with each 3-H pyrrole proton in close van der Waals contact with the 2-H of successive adenines (d <2.5 Å). Amide protons were found at longer distances from adenine 2-H and N-3 (d <3.5 Å). It is generally assumed that the main stabilizing forces of minor groove binders are van der Waals interactions between pyrrole aromatic protons and adenine 2-H, as well as hydrogen bonds between amides and thymine O-2 and/or adenine N-3. In our model, tallimustine shows bifurcated hydrogen bonds between each amide group, NH(A), NH(B) and NH(C), and the base edge nitrogens and oxygens of the DNA. This pattern is reminiscent of the hydrogen bonding pattern of the spine of hydration, bridging the same N-3 atoms of adenine and O-2 atoms of thymine. Hydrogen bond distances were found to be longer (2.8–3.2 Å) than usually expected for hydrogen bonded atoms, but were consistent with the X-ray structure of distamycin and netropsin complexes with other oligonucleotides.[26] Tallimustine, however, does not reside in the middle of the minor groove, as NH(A) was found to interact mainly with O-2 of T_5 (d = 2.6A) while NH(D) was found close to N-3 of A_{20}. In this orientation the phenyl protons of tallimustine were found in close van der Waals contact with the 2-H protons of A_{19} and A_{20} (Fig. 6). The minor groove at this level is sufficiently wide to allow the entrance of the bulky phenyl ring. The reduced propeller twist found for these two residues might help the phenyl ring to build up additional attractive van der Waals forces instead of repulsive ones. Thus the local conformation of the double helix in the $T_8G_9A_{10}$ tract is an important factor for the formation of the reversible complex. The sequence specificity of tallimustine must therefore originate from a fine balance of repulsive and attractive forces exerted by T_8:A_{19} and G_9:C_{18} base pairs towards the aniline mustard.

Conclusions

The tridecamer contains the recognition sequence of tallimustine, characterized by a long homopyrimidine tract followed by a purine tract, 'T_5GA'. In this oligomer the minor groove width progressively expands going from 5′ to the 3′ end of the 'TTTTT' tract. The local geometry of the double helix changes at the level of G_9 because of purine–pyrimidine alternation. The steric clash originated by residues located in opposite strands causes a progressive widening of the minor groove and the concomitant unwinding of the double helix, as monitored by a high value of the water exchange rate of the A_{10}:T_{17} imino proton.

The change in propeller twist and the consequent widening of the minor groove at the level of G_9:C_{18} and A_{10}:T_{17} seems to be an important factor for the stabilization of the complex with tallimustine. It is probable that the binding of tallimustine is entropy-driven: the molecule competes with the water molecules for the formation of bifurcated hydrogen bonds. The drug interacts mainly by positioning the distamycin moiety exactly in the 'TTTTT' tract, with the phenyl ring below the 2-NH_2 group of G_9 and the charged amidinic group at the beginning at the homopyrimidinic tract. The phenyl group perturbs the binding of the tallimustine from a steric point of view, but in addition it stabilizes the binding by means of new attractive van der Waals interactions. In the TGA tract the groove width becomes larger, thus allowing it to accommodate the bulky aromatic ring of tallimustine in the groove floor. In this orientation tallimustine seems to be quite far from the supposed alkylation site at the level of A_{10}. However, there is evidence from our NMR studies and from the HPLC analyses of the alkylation products that the drug interacts with the double helix with a multiple mode of binding. Thus, it is possible that tallimustine transiently 'slides' in a secondary binding site which has a more favourable geometry for alkylation. Work is in progress for the HPLC isolation and structural characterization of the alkylation products.

Acknowledgements

We kindly acknowledge Professor R. Mondelli for helpful discussions. We are also indebted to MURST and to CNR (Comitato per le ricerche tecnologiche e l'innovazione) for financial support.

References

1 F. C. Giuliani, B. Barbieri, G. Pezzoni, E. Lazzari, F. M. Arcamone and N. Mongelli, *Proc. Am. Ass. Cancer Res.*, 1988, **29**, 329.

2 G. Pezzoni, M. Grandi, G. Biasoli, L. Capolongo, D. Ballinari, F. C. Giuliani, B. Barbieri, A. Pastori, E. Pesenti, N. Mongelli and F. Spreafico, *Br. J. Cancer*, 1991, **64**, 1047.
3 R. D'Alessio, C. Geroni, G. Biasoli, E. Pesenti, M. Grandi and N. Mongelli, *Bioorg. Med. Chem. Lett.*, 1994, **4**, 1467.
4 D. Abigerges, G. P. Armand, L. Da Costa, E. Fadel, D. Mignard, C. Lhomme, M. G. Zurlo and D. Gandia, *Proc. Am. Ass. Cancer Res.*, 1993, **34**, 1589.
5 F. M. Arcamone, F. Animati, B. Barbieri, E. Configliacchi, R. D'Alessio, C. Geroni, F. C. Giuliani, E. Lazzari, M. Menozzi, N. Mongelli, N. S. Penco and M. A. Verini, *J. Med. Chem.*, 1989, **32**, 774.
6 C. L. Propst and T. J. Perun, *Nucleic Acid Targeted Drug Design*, Marcel Dekker, New York, 1992, p. 303.
7 M. D. Wyatt, M. Lee, B. J. Gabiras, R. L. Souhami and J. A. Hartley, *Biochemistry*, 1995, **34**, 13 034.
8 M. Broggini, H. M. Coley, N. Mongelli, E. Pesenti, M. D. Wyatt, J. A. Hartley and M. D'Incalci, *Nucleic Acids Res.*, 1995, **23**, 81.
9 C. Szezylik, T. Skorski, N. C. Nicoladeis, L. Manzella, L. Malaguarnera, D. Venturelli, A. M. Gewirtz and B. Calabretta, *Science*, 1991, **253**, 562.
10 F. X. Mahon, F. Belloc and J. Reiffers, *Lancet*, 1993, **341**, 566.
11 J. Jeener, B. M. Meier, P. Bachmann and R. R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546.
12 S. Macura and R. R. Ernst, *Mol. Phys.*, 1980, **41**, 95.
13 L. Braunschweiler and R. R. Ernst, *J. Magn. Res.*, 1983, **53**, 521.
14 G. Otting, R. Grutter, W. Leupin, C. Minganti, K. N. Ganesh, B. S. Sproat, M. J. Gait and K. Wütrich, *Eur. J. Biochem.*, 1987, **166**, 215.
15 A. C. Wang, S. G. Kim, P. F. Flynn, E. Sletten and B. R. Reid, *J. Magn. Reson.*, 1992, **100**, 358.
16 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
17 P. Rajagopal, D. E. Gilbert, G. A. van der Marel, J. H. van boom and J. Feigon, *J. Magn. Reson.*, 1988, **78**, 526.
18 R. Boelens, T. M. G. Koning, G. A. van der Marel, J. H. van Boom and R. Kaptein, *J. Magn. Reson.*, 1989, **82**, 290.
19 A. Saran, D. Perahia and B. Pullman, *Theor. Chim. Acta* (*Berlin*), 1973, **30**, 31.
20 R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 1981, **149**, 761.
21 D. G. Gorenstein, *Chem. Rev.*, 1994, **94**, 1315.
22 R. E. Dickerson, *J. Mol. Biol.*, 1983, **166**, 419.
23 C. R. Calladine, *J. Mol. Biol.*, 1982, **161**, 343.
24 M. Katahira, H. Sugeta, Y. Kyogoku, S. Fujisawa and R. Tomita, *Nucleic Acids Res.*, 1988, **16**, 8619.
25 M. Katahira, H. Sugeta and Y. Kyogoku, *Nucleic Acids Res.*, 1990, **18**, 613.
26 M. Coll, C. A. Frederick, A. H. J. Wang and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 8385.
27 W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984. p. 266 and references therein.
28 (*a*) K. Zakrzewska and B. Pullman, *Nucleic Acids Res.*, 1983, **11**, 8841; (*b*) R. Lavery, A. Pullman and B. Pullman, *Theor. Chim. Acta*, 1982, **62**, 93; (*c*) K. W. Kohn, J. A. Hartley and W. B. Mattes, *Nucleic Acids Res.*, 1987, **15**, 10 531.
29 D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis,* VCH, New York, 1986, p. 175.
30 S. Mazzini, G. Musco and E. Ragg, *Magn. Reson. Chem.*, 1994, **32**, 139.
31 J. G. Pelton and D. Wemmer, *Biochemistry*, 1988, **27**, 8088.
32 T. J. Dwyer, B. H. Geierstanger, Y. Bathini, J. W. Lown and D. E. Wemmer, *J. Am. Chem. Soc.*, 1992, **114**, 5911.
33 J. G. Pelton and D. E. Wemmer, *J. Am. Chem. Soc.*, 1990, **112**, 1393.
34 R. E. Klevit, D. E. Wemmer and B. Reid, *Biochemistry*, 1986, **25**, 3296.
35 P. G. Schutz and P. B. Dervan, *J. Am. Chem. Soc.*, 1983, **105**, 7748.
36 R. P. Hertzberg and P. B. Dervan, *J. Am. Chem. Soc.*, 1982, **104**, 313.
37 (*a*) M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *J. Mol. Biol.*, 1985, **183**, 553; (*b*) M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *Proc. Natl. Acad. Sci. USA*, 1985, **82**, 1376.
38 M. Coll, J. Aymami, G. A. van der Marel, J. H. van Boom, A. Rich and A. H. J. Wang, *Biochemistry*, 1989, **28**, 310.
39 A. Blaskò, K. A. Browne and T. C. Bruice, *J. Am. Chem. Soc.*, 1994, **116**, 3726.
40 (*a*) L. A. Marky and K. J. Breslauer, *Biopolymers*, 1987, **26**, 1601; (*b*) D. M. Crothers, *Biopolymers*, 1971, **10**, 2147.
41 K. J. Breslauer, R. Frank, H. Blocker and L. A. Marky, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 3746.
42 (*a*) L. A. Marky and K. J. Breslauer, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4359; (*b*) K. J. Breslauer, D. P. Remeta, W. Y. Chou, R. Ferrante, J. Curry, D. Zaunczkowski, J. G. Snyder and L. A. Marky, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 8922.
43 E. L. Fish, M. L. Lane and J. N. Vournakis, *Biochemistry*, 1988, **27**, 6026.
44 M. Fontana, M. Lestingi, C. Mondello, A. Braghetti, A. Montecucco and G. Ciarrocchi, *Anti-Cancer Drug Des.*, 1992, **7**, 131.
45 C. J. O'Connor, W. A. Denny, J. Y. Fan, G. Lance Gravatt, B. A. Grigor and D. J. McLennan, *J. Chem. Soc.*, *Perkin Trans. 2*, 1991, 1933.
46 J. A. Hartley, S. M. Forrow and R. L. Souhami, *Biochemistry*, 1990, **29**, 2985.
47 T. A. Gourdie, K. K. Valu, G. Lance Gravatt, T. J. Boritzki, B. C. Baguley, L. P. G. Wakelin, W. R. Wilson, P. D. Woodgate and W. A. Denny, *J. Med. Chem.*, 1990, **33**, 1177.

Paper 7/02943H
Received 29*th April* 1997
Accepted 29*th September* 1997