[15] Data for **2**: $[\alpha]_D^{26} = -61.4°$ ($c = 1.04$ in $CHCl_3$); ^1H NMR (400 MHz, $CDCl_3$): $\delta = 4.81$–4.59 (m, 6H), 4.10, (m, 2H), 3.47 (m, 2H), 3.39 (s, 3H), 3.38 (s, 3H), 3.36 (s, 3H), 3.33 (m, 1H), 3.21 (m, 1H), 3.10 (m, 1H), 2.47 (ddd, $J = 4.9, 2.9, 1.7$ Hz, 1H), 2.35 (ddd, $J = 7.3, 2.4, 1.7$ Hz, 1H), 2.24–1.22 (m, 12H), 1.25 (br s, 18H), 0.88 ppm (t, $J = 6.8$ Hz, 3H); ^{13}C NMR (75 MHz, $CDCl_3$): $\delta = 96.9, 96.8, 95.3, 81.6, 81.0, 80.9, 79.7, 79.5, 79.1, 77.1, 75.8, 69.6, 55.7, 55.5, 32.1, 31.9, 31.3, 30.1, 29.7, 29.6, 29.3, 28.2, 26.9, 26.6, 26.3, 25.5, 25.3, 22.7, 14.1$ ppm; IR (neat): $\tilde{\nu} =$ 3312, 3273, 2926, 2855, 2823, 1465, 1457, 1396, 1375, 1363, 1340, 1293, 1214, 1150, 1103, 1075, 1036, 918 cm^{-1}; HRMS calcd for $C_{32}H_{58}O_8Na$ $[M+Na]^+$: m/z 593.4029, found: m/z 593.4037.

[16] G. E. Keck, D. Krishnamurthy, M. C. Grier, *J. Org. Chem.* **1993**, *58*, 6543–6544.

[17] K. Takai, K. Nitta, K. Utimoto, *J. Am. Chem. Soc.* **1986**, *108*, 7408–7410.

[18] T. R. Hoye, Z. Ye, *J. Am. Chem. Soc.* **1996**, *118*, 1801–1802.

Structure of a Drug-Induced DNA T-Bulge: Implications for DNA Frameshift Mutations**

Michelle L. Colgrave, Huw E. L. Williams, and Mark S. Searle*

The biological function of nucleic acids is recognized to embrace a wide variety of structural topologies whose stability, conformation, and dynamics may be strongly influenced by the interaction of drug molecules. Single-base bulges, arising from slippage of DNA strands, particularly in homopolymeric tracts of DNA, can result in mutation "hotspots" due to DNA strand misalignment during replication.[1] Addition or deletion mutations at these sites can arise according to whether the extra base stacks within the DNA helix or is displaced into solution. Unpaired pyrimidines have been shown to adopt either an intrahelical or extrahelical conformation depending on flanking sequence, while purines preferentially stack within the helix.[2] The interaction of drug molecules at bulge sites has been of considerable interest since binding close to these sites has the potential to stabilize and increase the lifetime of the bulge state and its susceptibility to frameshift mutations. It has already been demonstrated that ethidium bromide and 9-aminoacridine have increased affinity for sites on duplex DNA carrying an extrahelical cytosine,[3] while neocarzinostatin induces highly efficient site-specific strand cleavage at bulge sites in folded single-stranded DNA.[4] In other cases, chemical modification (base alkylation) of DNA and RNA by drugs appears to occur primarily near bulge sites because of better binding site access,[5] while selective drug binding to RNA bulges has been shown to inhibit protein–RNA recognition.[6] In general, detailed structural information is lacking, although a number of low-resolution NMR studies of bulge recognition by simple DNA intercalators have been described.[3]

The anthracycline antibiotic nogalamycin (Figure 1) has a high sequence specificity for 5′-CG or 5′-TG sites with its bound orientation dictated by the requirement for a guanine (G) base on the 3′-side of the intercalation site.[7] Nogalamycin

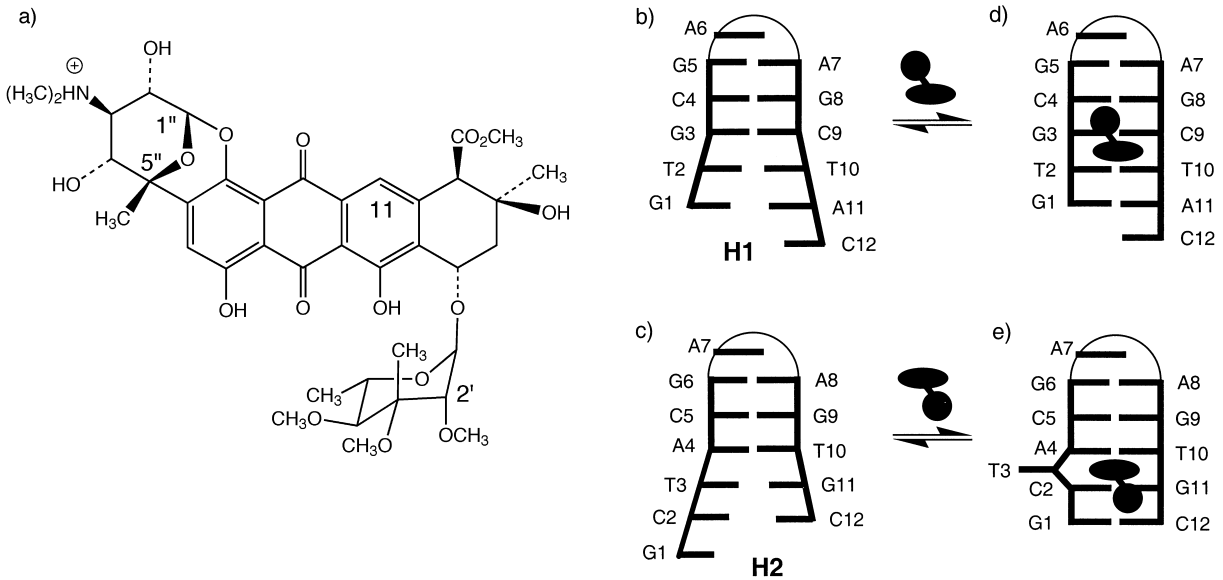

Figure 1. a) Structure of nogalamycin, and oligonucleotide sequences **H1** (b) and **H2** (c) in their proposed folded conformation with frayed ends. Conformations and base-pair alignment adopted in the bound stated with nogalamycin intercalated at the 5′-TpG step ((d) and (e)).

[*] Dr. M. S. Searle, Dr. M. L. Colgrave, Dr. H. E. L. Williams
School of Chemistry
University Park, Nottingham NG7 2RD (UK)
Fax: (+44) 115-951-3564
E-mail: mark.searle@nottingham.ac.uk.

[**] We thank the EPSRC of the UK, the Australian Research Council, and Boehringer Ingelheim Fonds for supporting this work.

threads its aglycon chromophore through the DNA helix and positions bulky sugar residues in both grooves simultaneously. Interactions in the major groove involving the two hydroxy groups on the bicycloaminoglucose sugar appear to account for the drugs orientational specificity through hydrogen bonding to guanine. Thus, binding to the 5′-TG site results in a unique binding orientation dictated by the G base. With

  0044-8249/02/4124-4754 $ 20.00+.50/0

this requirement in mind, we report NMR studies of two hairpin sequences each containing a single 5′-TG binding site incorporating an extra bulged thymine (T) base on the complementary strand (5′-C^{T}A; see Figure 1). We examine how G–C base pair specificity dictates the conformation of the extra T which could be accommodated through a drug-induced frameshift with base mispairing occurring downstream of the drug binding site, or through local stabilization of a T-bulge at the 5′-C^{T}A step preserving Watson–Crick base pairing downstream.

We have designed two novel DNA hairpins (**H1** and **H2**) containing a 5′-GAA loop sequence that has previously been reported to show remarkable thermal stability.[8] Both **H1** and **H2** contain the 5′-TG nogalamycin intercalation site but in different orientations with respect to the stabilizing loop sequence (Figure 1b and c). We have investigated the conformation of the two hairpins in the absence of drug using 1D and 2D NMR spectroscopy in H_2O and D_2O solutions.[9] The hairpins show characteristic NMR parameters that demonstrate that the 5′-GAA loop is formed in both cases.[8] Examination of the imino-proton region of the spectrum at 278 K reveals only two stabilized Watson–Crick pairs in each case. For **H1**, we see signals for G3 and G8 (Figure 2a); in spectra of **H2** signals are evident only for G9 and T10, with the latter showing exchange broadening (Figure 2b). Detailed NMR analysis shows that while the stem region close to the loop is well formed in both structures, the ends of the two hairpins are destabilized and poorly structured (Figure 1b and c).

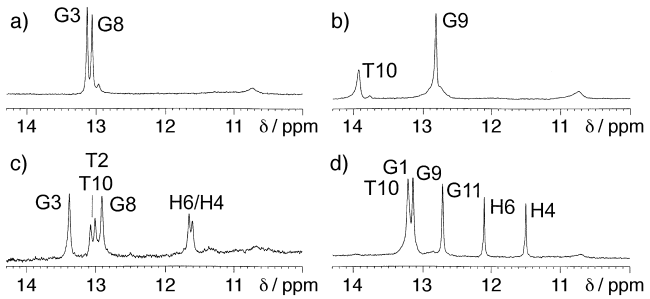

Figure 2. Imino-proton region of the 500 MHz 1D NMR spectra (278 K) of a) free **H1** and b) free **H2**, and in the presence of bound drug **Nog-H1** (c) and **Nog-H2** (d); labeling scheme in accord with Figure 1.

Complex formation with nogalamycin (**Nog**) was monitored by 1D ^{1}H NMR spectroscopy and showed that drug binds cleanly to each sequence in a 1:1 stoichiometry with the free and bound species in slow exchange on the chemical shift time scale. The imino-proton region of the **Nog-H1** complex reveals two sharp signals corresponding to G3 and G8, as evident for uncomplexed **H1,** plus two additional weaker signals of similar intensity at $\delta \approx 13.0$ ppm (Figure 2c), which simultaneously broaden and disappear above 298 K. The site of intercalation and the bound orientation of the drug is established from drug–DNA NOEs; for example, Nog-H11 shows clear interactions with the deoxyribose H1′ of T2 and G3 indicating that the drug is intercalating at the 5′-TG step. Numerous NOEs from methyl and methoxy groups on the nogalose sugar identify hydrophobic interactions within the minor groove consistent with previous studies.[7] NOEs from Nog-H1″ and Nog-5″-CH_3 on the bicycloamino glucose sugar to T2 CH_3 confirm its orientation in the major groove with the drug hydroxy groups positioned to interact with the G3–C9 base pair. The signals for T10 show evidence for exchange broadening, however, weak NOEs from Nog-2′-OCH_3 to T10 H2″/H4′, together with the observation of the two broadened imino-proton signals in Figure 2c, suggest that the drug stacks with a weakly stabilized T2–T10 mismatched pair with the 3′- and 5′-termini poorly structured (Figure 1d).

In contrast, drug binding to hairpin **H2** produces a highly stabilized 1:1 complex. Four imino-proton signals are identified (G1 and T10 overlapped in Figure 2d) indicating that G1, G9, T10, and G11 are involved in stable Watson–Crick base pairs with only the terminal G1 base showing evidence for exchange broadening at 308 K. The chemical shift and NOE data readily confirmed that drug binding stabilizes a bulged conformation in which the T3 base is flipped-out of the helical stack allowing complementary pairing downstream of the intercalation site (Figure 1e). We see that both the base and sugar proton signals of T3 are significantly downfield of those of the uncomplexed DNA ($\Delta\delta$(H1′) = 0.75 ppm; $\Delta\delta(CH_3)$ = 0.48 ppm) suggesting an absence of ring current effects from neighboring stacked bases. In addition, we see only a very weak sequential NOE (base H6/H8 to sugar H2′/2″) from T3 to C2, and none from A4 to T3 (Figure 3) providing strong evidence that T3 is not accommodated within the helical stack. Weak NOEs from T3 CH_3 to C2 H6 and C2 H3′, but very weak or an absence of NOEs to C2 H1′ and C2 H2′/2″, indicate that the extrahelical T base is located in the major groove. A detailed analysis of the NOE data substantiates this conclusion as well as identifying many drug–DNA interactions in both grooves that are analogous to those described above; Nog-H11 unambiguously pins down the bound ori-

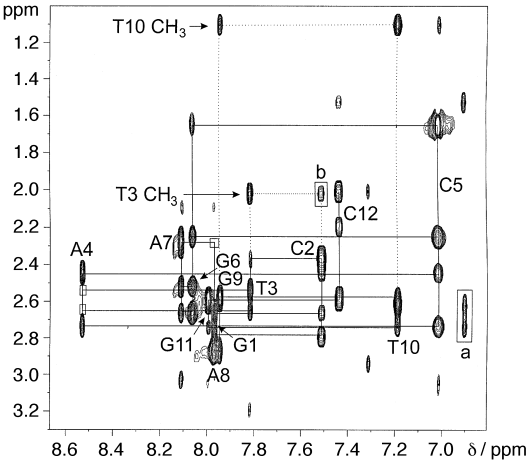

Figure 3. Portion of the 600 MHz NOESY spectrum (300 ms mixing time) of the **Nog-H2** complex showing NOEs between base H6/H8 and sugar H2′/2″. Vertical lines identify the positions of the base H6/H8 signals, while horizontal lines show sequential connectivities. Intranucleotide H6/H8–H2′/H2″ are labeled G1 through to C12. Boxes mark the positions of sequential NOEs that are too weak to observe (A4 H8–T3 H2′/H2″ and A8 H8–A7 H2′/H2″, the latter in the GAA loop). Dotted lines show NOEs from T10 CH_3 and T3 CH_3. Additional cross-peaks are labeled a) Nog-H11–T10 H2′/H2″ and b) T3 CH_3–C2 H6.

  0044-8249/02/4124-4755 $ 20.00+.50/0

entation giving strong NOEs to T10 H1′ and G11 H1′. 340 NOE distance restraints (including some 39 drug–DNA interactions) have enabled us to calculate a family of structures of the **Nog-H2** complex using a combination of distance geometry simulated annealing and restrained molecular dynamics to model this unusual structure (Figure 4).[10]

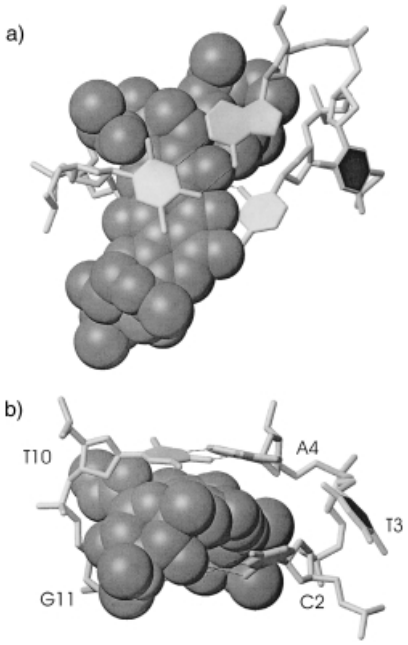

Figure 4. NOE-refined structure of the **Nog-H2** complex showing base-pair alignment and orientation of the extrahelical T-bulge at the drug intercalation site: a) view of the complex down the helix axis illustrating the orientation of the extrahelical T base located in the major groove and drug stacking with A4–T10 (top) and C2–G11 (below), b) view into the major groove showing close van der Waals contacts with the drug at the A4–C2 step with the T3 base flipped out of the helical stack.

Earlier structural studies have rationalized nogalamycin binding specificity for the 5′-TG site in the context of specific interactions in the major groove between the polar bicycloaminoglucose sugar and an intact G–C base pair.[7] These interactions appear to be primary determinants in stablizing an extrahelical base in the **Nog-H2** complex where forcing the bulged-T out of the stack facilitates formation of the required binding site G–C base pair. In contrast, for **H1** the required G–C pairing is already present in the free hairpin such that the nogalamycin–DNA recognition criteria are already partially satisfied. However, the drug appears to be noninstructional downstream of the binding site in the sense that a weakly stabilized T2–T10 mispaired conformation prevails mimicking a "frameshift" (Figure 1 d). Earlier studies of nogalamycin binding to 5′-CG terminal bulge sites in two DNA heptamers, C^{T}GTACG and CGTAC^{T}G,[11] have reported a strand misalignment with T–G and C–T mispairs formed at the terminal sites, leaving 5′-C and 3′-G dangling nucleotides, respectively. However, we demonstrate that the structural distortion observed is context-dependent as a consequence of base-specific recognition, leading to either a highly localized T-bulge, or a frameshift that can potentially propagate along the DNA sequence.

DNA intercalators have already been shown to have a significant impact in inducing mutations during replication by perturbing the recognition of DNA defects by proofreading and repair enzymes.[12] The ability of some intercalators to act as mutagens has already been reported.[3b,13] Here we describe one of the first detailed NMR structural analyses of a drug-stabilized DNA bulge.

Received: August 14, 2002 [Z19970]

[1] a) G. Streisinger, J. Owen, *Genetics* **1985**, 633–659; b) S. A. Woodson, D. M. Crothers, *Biochemistry* **1987**, *26*, 904–912; c) S. A. Woodson, D. M. Crothers, *Biochemistry* **1988**, *27*, 436–445.

[2] a) D. J. Patel, S. A. Kozlowski, L. A. Marky, J. A. Rice, C. Broka, K. Itakura, K. Breslauer, *Biochemistry* **1982**, *21*, 445–451; b) K. M. Morden, Y. G. Chu, F. H. Martin, I. Tinoco, *Biochemistry* **1983**, *22*, 5557–5563; c) S. A. Woodson, D. M. Crothers, *Biochemistry* **1988**, *27*, 3130–3141; d) M. W. Kalnik, D. G. Norman, M. G. Zagorski, P. F. Swann, D. J. Patel, *Biochemistry* **1989**, *28*, 294–303; e) J. Zhu, R. M. Wartell, *Biochemistry* **1999**, *38*, 15986–15993.

[3] a) J. W. Nelson, I. Tinoco, *Biochemistry* **1985**, *24*, 6416–6421; b) S. A. Woodson, D. M. Crothers, *Biochemistry* **1988**, *27*, 8904–8914; c) K. Nakatani, S. Sando, I. Saito, *J. Am. Chem. Soc.* **2000**, *122*, 2172–2177.

[4] a) L. S. Kappen, I. H. Goldberg, *Biochemistry* **1997**, *36*, 14861–14867; b) Z. Xi, Q. K. Mao, I. H. Goldberg, *Biochemistry* **1999**, *38*, 4342–4354.

[5] a) L. D. Williams, I. H. Goldberg, *Biochemistry* **1988**, *27*, 3004–3011; b) A. Stassinopoulous, J. Ji, X. Gao, I. H. Goldberg, *Science* **1996**, *272*, 1943–1946; c) K. Nakatani, A. Okamoto, I. Saito, *Angew. Chem.* **1999**, *111*, 3581–3584; *Angew. Chem. Int. Ed.* **1999**, *38*, 3378–3381.

[6] F. Hamy, E. R. Felder, G. Heizmann, J. Lazdins, F. Aboul-ela, G. Varani, J. Karn, T. Klimkait, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3548–3553.

[7] a) X. L. Zhang, D. J. Patel, *Biochemistry* **1990**, *29*, 9451–9466; b) C. K. Smith, J. A. Brannigan, M. H. Moore, *J. Mol. Biol.* **1996**, *263*, 237–258; c) H. E. Williams, M. S. Searle, *J. Mol. Biol.* **1999**, *290*, 699–716.

[8] a) I. Hirao, G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. A. Watanabe, K.-I. Miura, *Nucleic Acids Res.* **1994**, *22*, 576–582; b) S. Yoshizawa, G. Kawai, K. A. Watanabe, K.-I. Miura, I. Hirao, *Biochemistry* **1997**, *36*, 4761–4767.

[9] Oligonucleotides were synthesized by using standard solid-phase methods and purified by reverse-phase HPLC "trityl-on". NMR data were collected at oligonucleotide sample concentrations of 3 mM in 100 mM NaCl and 10 mM Na_2HPO_4 pH 7.0 at the temperatures indicated by using Bruker DRX500 and Avance600 spectrometers and standard phase-sensitive pulse sequences.

[10] Distance restraints (232 DNA–DNA, 39 drug–DNA and 69 drug–drug NOEs) were derived from NOESY data collected at mixing times of 75, 100, and 125 ms by extrapolating to zero mixing time (J. Baleja, J. Moult, B. D. Sykes, *J. Magn. Reson.* **1990**, *87*, 375–384). Candidate structures were obtained from a set of 100 randomly generated starting structures using NMR-derived distance restraints and a simulated annealing/torsion angle dynamics protocol using DYANA (P. Guntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* **1997**, *273*, 283–292). Subsequently, NOE-restrained MD simulations were carried out for a total of 700 ps, using an explicit solvent model, with the structures minimized by cooling from 300 K to 0 K over the final 100 ps resulting in well-defined structures with a low NOE restraint violation energy; previously reported procedures were used throughout (E. Gavathiotis, G. J. Sharman, M. S. Searle, *Nucleic Acids Res.* **2000**, *28*, 728–735). Structures were displayed and analyzed using MOLMOL (R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51–55).

[11] J. Caceres-Cortes, A. H.-J. Wang, *Biochemistry* **1996**, *35*, 616–625.

[12] a) W.-C. Lam, E. J. C. Van der Schans, L. C. Sowers, D. P. Millar, *Biochemistry* **1999**, *38*, 2661–2668; b) V. A. Malkov, I. Biswas, R. D. Camerini-Otero, P. Hsieh, *J. Biol. Chem.* **1997**, *272*, 23811–23817.

[13] B. Ames, F. D. Lee, W. E. Durston, *Proc. Natl. Acad. Sci. USA* **1973**, *70*, 782–786.

  0044-8249/02/4124-4756 $ 20.00+.50/0