Site-specificity of bis-benzimidazole Hoechst 33258 in A-tract recognition of the DNA dodecamer duplex d(GCAAAATTTTGC)$_2$

Lee D. Higgins and Mark S. Searle*

Department of Chemistry, University of Nottingham, University Park, Nottingham, UK NG7 2RD. E-mail: mark.searle@nottingham.ac.uk

Received (in Liverpool, UK) 24th June 1999, Accepted 13th August 1999

Fluorescence titration measurements and NMR spectroscopy have identified a single molecule of Hoechst 33258 bound within the long A-tract of the dodecamer duplex d(GCAAAATTTTGC)$_2$: despite the possibility of multiple binding sites, and the apparent requirement from previous studies for the drug *N*-methylpiperazine ring to bind close to GC base pairs at the end of an A-tract (leading to an off-set asymmetric complex with this sequence), we identify a binding site that spans the central AAATTT sequence with the planar aromatic rings of the drug binding in the narrowest part of the minor groove.

The bis-benzimidazole Hoechst 33258 (H33258) binds with high affinity to the minor groove of AT rich DNA sequences.[1] Footprinting studies[2] and structural analysis of various H33258 complexes by NMR[3] and X-ray crystallography[4–6] have identified a number of key features of DNA recognition: (i) the drug binds to at least four AT base pairs, (ii) TpA steps are generally avoided, and (iii) a GC base pair is frequently accommodated at the 5′-end of the binding site. The conformation and orientation of the piperazine ring appears to account for preferential binding close to GC regions at the end of A-tracts of DNA where the wider groove has been proposed to more readily accommodate this bulky substituent.[4,5] In complexes with binding sites of only four AT base pairs (AATT)[4] the A-tract is not of sufficient length to examine this effect in detail, however, long A-tracts might enable significant site discrimination to be identified to examine the importance of the influence of the bulky *N*-methylpiperazine ring in site-specificity.

Here, we present an analysis of the interaction of H33258 with an eight AT base pair A-tract sequence within the dodecamer duplex d(GCAAAATTTTGC)$_2$ using a combination of fluorescence titration data and NMR spectroscopy to determine binding stoichiometry and site specificity within the A-tract. The sequence contains several possible high affinity sites within the A-tract (AATT, ATTT or AAAA) that could potentially complicate the analysis through multiple site occupancy with the ligand in exchange between them. Further, the A-tract is of sufficient length to accommodate two bound ligands each spanning four AT base pairs.

Fluorescence titration data were used to determine the binding stoichiometry of H33258 with the dodecamer duplex d(GCAAAATTTTGC)$_2$.[7,8] The method of continuous variation in ligand concentration (Job plot analysis) indicates a fluorescence maximum at a mole fraction of drug of 0.53, establishing a 1:1 (drug:duplex) binding stoichiometry (Fig. 1). Non-linear least-squares analysis of fractional binding saturation curves at 20 °C indicated tight binding ($\sim 10^8$ M^{-1}) in agreement with previous estimates using AATT[9] and AAATTT[7,8] containing duplexes. Despite the possibility of accommodating two drug molecules in adjacent sites, the fluorescence data indicate a single bound ligand per duplex. The fluorescence data are unlikely, however, to be able to distinguish between a number of different 1:1 binding modes in equilibrium if the ligand is bound in each case with a similar affinity.

NMR titration analysis was used to examine the possibility of multiple binding modes. In Fig. 2, the thymine methyl region of the spectrum of d(GCAAAATTTTGC)$_2$ is illustrated (1.0–2.0 ppm) for the free duplex, 0.5:1 ratio of drug:duplex and for the 1:1 complex. Addition of drug results in the four signals from the free duplex (T7, T8, T9 and T10) being replaced by a set of eight new resonances from an asymmetric 1:1 complex in which the two strands of the duplex are no longer equivalent and are in slow exchange through ligand dissociation and reassociation. We see no evidence for multiple sets of resonances that

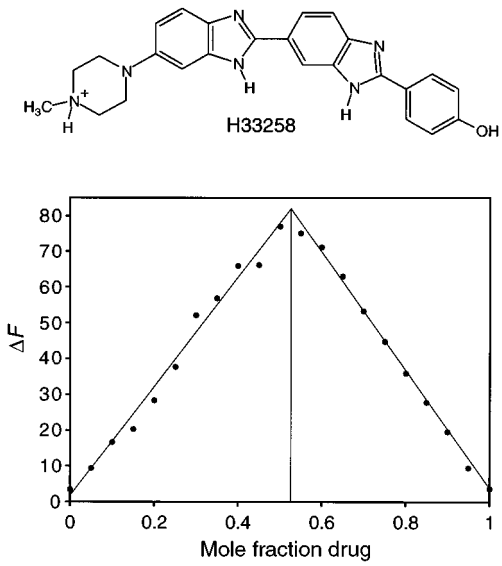

Fig. 1 Structure of H33258. Continuous variation analysis (Job plot) for binding of H33258 to d(GCAAAATTTTGC)$_2$ at pH 7.2, 100 mM NaCl and 10 mM phosphate buffer using a constant value for [ligand] + [DNA] = 0.1 μM; change in fluorescence intensity (ΔF) is plotted against mole fraction of drug. Spectra were recorded on a Perkin-Elmer Luminescence spectrometer LS 50 B at a temperature of 20 °C in 1 cm path-length polymethacrylate cuvettes using fluorescence excitation wavelengths between 330 to 360 nm and emission wavelengths from 390 to 420 nm.

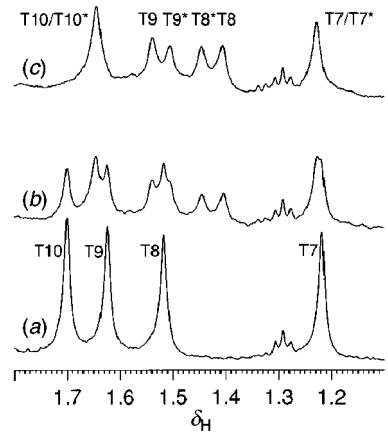

Fig. 2 ^{1}H NMR spectra (500 MHz) of the thymine methyl region (1.0–2.0 ppm) of d(GCAAAATTTTGC)$_2$: (*a*) ligand-free DNA, (*b*) drug:duplex ratio 0.5:1, and (*c*) 1:1 complex at 298 K; the two strands are distinguished by the use of asterisks, assignments are shown.

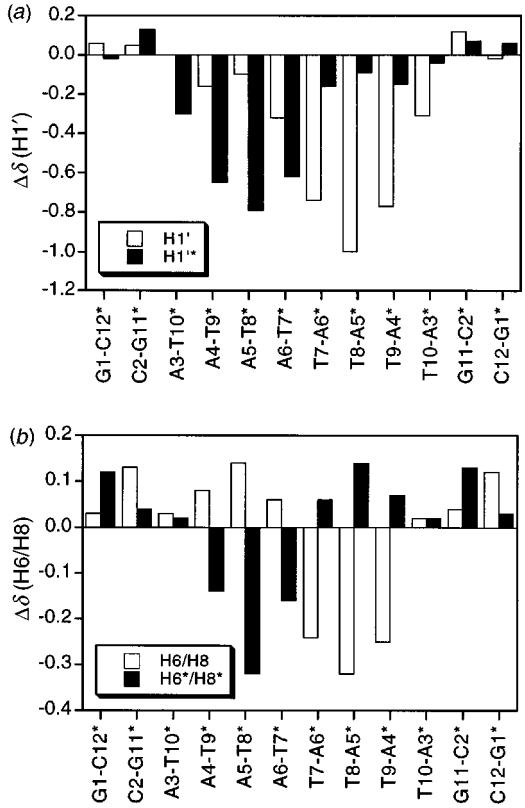

Fig. 3 Plots of drug-induced changes in ^{1}H NMR chemical shifts ($\Delta\delta = \delta_{complex} - \delta_{free}$) for (*a*) deoxyribose H1′ and (*b*) base H6/H8 for the H33258 complex with d(GCAAAATTTTGC)$_2$. The two strands of DNA read in antiparallel directions, one of which is distinguished by the use of asterisks and filled columns.

would indicate the presence of a number of slowly exchanging conformations, leading us to conclude that the drug either binds at a single high affinity site, or that the bound drug is in fast exchange between a number of overlapping sites.

To examine more closely the position of the drug within the groove we have used ^{1}H chemical shift data (assigned through 2D TOCSY and NOESY data sets) to provide an NMR 'footprint' from ligand-induced perturbations to deoxyribose H1′ and base H6/H8 resonances ($\Delta\delta = \delta_{complex} - \delta_{free}$). $\Delta\delta$ values are plotted against sequence position in Fig. 3. Many deoxyribose H1′ protons come into direct contact with the face of one or other aromatic ring of the bound ligand and experience large upfield ring current perturbations to their chemical shifts of up to 1.0 ppm. The $\Delta\delta$ values of these signals are particularly sensitive to the position of the ligand in the minor groove, with the thymine sugars showing the largest perturbations. In contrast, the effects on base H6/H8 resonances, which are located in the major groove, are smaller ($\leqslant$0.3 ppm), and probably arise from changes in base stacking interactions, but reflect the same general pattern of perturbations. The striking observation is that both sets of chemical shift perturbations suggest a highly centrosymmetric binding of the drug across the dyad axis of the duplex spanning primarily the central AAATTT sequence. The AT base pairs A3*-T10 and A3-T10* at the ends of the A-tract are relatively unperturbated by the binding interaction.

In contrast, in one X-ray structure of H33258 bound to d(CGCAAATTTGCG)$_2$,[5] containing a shorter A-tract, the drug appears to be displaced off-centre to bind across the ATTTG sequence. This binding locus is reported to accommodate the piperazine ring close to the GC base pairs at the end of the A-tract.[5] However, in a second X-ray structure with a similar AAATTT binding site the displacement is less pronounced.[6] Our data place the *N*-methylpiperazine ring within the A-tract with the bis-benzimidazole portion of the ligand binding across the AT dyad axis. Although our results seem at variance with some structural data, they do concur with some footprinting studies which also consider the interaction with longer AT tracts and which show that AATT is an unusually good binding site for H33258.[10] On a purely statistical basis, for a ligand binding site of four base pairs where there is no discrimination between sites, the central AT base pairs are expected to show the highest time-averaged site occupancy. Is a model involving fast exchange between overlapping sites consistent with the NMR data? Our preliminary analysis of intermolecular NOE data, and structural modelling, suggests that the bound ligand is restricted to a relatively small binding locus rather than shuffling between multiple (symmetrical and asymmetrical) overlapping sites. Recent studies on the binding of ammonium cations to A-tract structures,[11] together with earlier results,[12] suggest that the minor groove is at its narrowest at the centre of A_nT_n sequences. In the present context, maximising van der Waals interactions between the bound ligand and the walls of the narrowest part of the groove may be the primary determinant of A-tract site-specificity. A more detailed analysis of the structure and dynamics of this and other A-tract complexes in solution is currently in progress.

The EPSRC, the Department of Chemistry at the University of Nottingham, and the Royal Society are gratefully acknowledged for financial support. We thank John Keyte for oligonucleotide synthesis.

Notes and references

1 J. Portugal and M. J. Waring, *Biochem. Biophys. Acta*, 1988, **949**, 158.
2 K. D. Harshman and P. B. Dervan, *Nucleic Acids Res.*, 1985, **13**, 4825; A. Abu-Daya, P. M. Brown and K. R. Fox, *Nucleic Acids Res.*, 1995, **23**, 3385.
3 M. S. Searle and K. J. Embrey, *Nucleic Acids Res.*, 1990, **18**, 3753; J. A. Parkinson, J. Barber, K. T. Douglas, J. Rosamond and D. Sharples, *Biochemistry*, 1990, **29**, 10 181; A. Fede, A. Labhardt, W. Bannwart and W. Leupin, *Biochemistry*, 1991, **30**, 11 377; K. J. Embrey, M. S. Searle and D. J. Craik, *Eur. J. Biochem.*, 1993, **211**, 437; A. Fede, M. Billeter, W. Leupin and K. Wuthrich, *Structure*, 1993, **1**, 177; C. E. Bostock-Smith, C. A Laughton and M. S. Searle, *Nucleic Acids Res.*, 1998, **26**, 1660.
4 P. E. Pjura, K. Grzeskowiak and R. E. Dickerson, *J. Mol. Biol.*, 1987, **197**, 257; M.-K. Teng, N. Usman, C. A. Frederick and A. H. J. Wang, *Nucleic Acids Res.*, 1988, **16**, 2671; M. A. A. F. deC. T. Carrondo, M. Coll, J. Aymami, A. H. J. Wang, G. A. van der Marel, J. H. van Boom and A. Rich, *Biochemistry*, 1989, **28**, 7849; J. R. Quintana, A. A. Lipanov and R. E. Dickerson, *Biochemistry*, 1991, **30**, 10 294; G. R. Clark, C. J. Squire, E. J. Gray, W. Leupin and S. Neidle, *Nucleic Acids Res.*, 1996, **24**, 4882; G. R. Clark, D. W. Boykin, A. Czarny and S. Neidle, *Nucleic Acids Res.*, 1997, **25**, 1510; G. R. Clark, E. J. Gray, S. Neidle, Y.-H. Li and W. Leupin, *Biochemistry*, 1996, **35**, 13 745.
5 N. Spink, D. G. Brown, J. V. Skelly and S. Neidle, *Nucleic Acids Res.*, 1994, **22**, 1607.
6 M. C. Vega, I. Garcia Saez, J. Aymami, R. Eritja, G. A. van der Marel, J. H. van Boom, A. Rich and M. Coll, *Eur. J. Biochem.*, 1994, **222**, 721.
7 C. E. Bostock-Smith and M. S. Searle, *Nucleic Acids Res.*, 1999, **27**, 1619.
8 I. Haq, J. E. Ladbury, B. Z. Chowdhry, T. C. Jenkins and J. B. Chaires, *J. Mol. Biol.*, 1997, **271**, 244.
9 F. G. Loontiens, L. W. McLaughlin, S. Diekmann and R. M. Clegg, *Biochemistry*, 1991, **30**, 182.
10 A. Abu-Daya and K. R. Fox, *Nucleic Acids Res.*, 1997, **25**, 4962.
11 N. V. Hud, V. Sklenar and J. Feigon, *J. Mol. Biol.*, 1999, **286**, 651.
12 A. M. Burkhoff and T. D. Tullius, *Cell*, 1987, **48**, 935.

Communication 9/05232A