ChemComm

Symmetric bis-benzimidazoles: new sequence-selective DNA-binding molecules

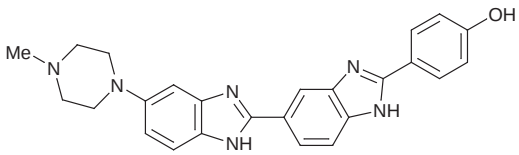
Stephen Neidle,*[a] John Mann,*[b] Emma L. Rayner,[a] Anne Baron,[b] Yaw Opoku-Boahen,[b] Ian J. Simpson,[a] Nicola J. Smith,[a] Keith R. Fox,[c] John A. Hartley[d] and Lloyd R. Kelland[e]

[a] *The CRC Biomolecular Structure Unit, The Institute of Cancer Research, Sutton, Surrey, UK SM2 5NG*
[b] *Department of Chemistry, University of Reading, Whiteknights Park, Reading UK RG6 6AD*
[c] *Division of Biochemistry and Molecular Biology, University of Southampton, Southampton, UK SO16 7PX*
[d] *CRC Drug-DNA Interactions Research Group, UCL Medical School, London, UK W1P 8BT*
[e] *CRC Centre for Cancer Therapeutics, The Institute of Cancer Research, Sutton, Surrey, UK SM2 5NG*

Received (in Liverpool, UK) 8th February 1999, Accepted 30th March 1999

A series of *bis*-benzimidazole compounds with a head-to-head orientation have been designed as sequence-specific DNA binders; crystallographic analysis of oligonucleotide complexes has been combined with DNase I footprinting to confirm that the predicted optimal site for the core bis-benzimidazole motif is the four-base-pair sequence 5′-AATT; this sequence specificity results in inhibition of transcription at A/T sites and may be responsible for the cytotoxic and antitumour effects shown by these head-to-head bis-benzimidazoles.

The minor groove of duplex DNA is the site of non-covalent interaction of a large number of anticancer drugs, antibiotics and antiviral agents,[1] which are believed to exert their action by competing with transcription factors[2] or architectural proteins,[3] such as E2F, TATA-box binding proteins or DNA topoisomerase I/II. The molecular basis of DNA recognition for a number of drugs in this super-family (notably distamycin, netropsin, berenil, pentamidine and Hoechst 33258) has been extensively studied, by crystallographic,[4] NMR[5] and footprinting methods.[6] These studies have shown that their AT preferences are a consequence of, in particular, (i) the sequence-dependent narrow width of the minor groove of B-form DNA, resulting in stabilisation by van der Waals interactions with the walls of the groove,[7] and (ii) their ability to form specific H-bonds with donor and acceptor atoms on the minor groove edge of A : T base pairs. These factors have been utilised in the design of molecules with altered and extended recognition properties, some of which are capable of sequence-specific gene regulation.[8]



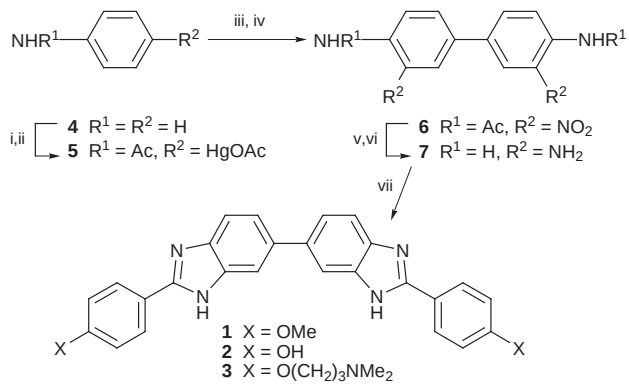
Hoechst 33258

We have previously reported crystallographic analyses of a number of oligonucleotide complexes with the head-to-tail bis-benzimidazole compound Hoechst 33258, and several of its derivatives,[9] including a tris-benzimidazole with an extended recognition site.[10] These have shown that each benzimidazole subunit interacts with two A : T base pairs by means of a pair of cross-strand H-bonds. The head-to-tail arrangement forces the site for each successive subunit to overlap the previous one by one base-pair (bp), so that each benzimidazole group in Hoechst 33258 and other head-to-tail analogues effectively recognises 1.5 A : T bps.

Molecular modelling, based on these structural studies, has suggested to us that the novel symmetric head-to-head benzimidazole arrangement could also effectively bind in the minor groove. This arrangement would extend the size of the *bis*-benzimidazole recognition site from three (in Hoechst 33258 and analogues) to four consecutive A:T bps, with distinctive cross-strand H-bonding involving each base pair. This arrangement thus extends the effective recognition of each benzimidazole sub-unit to two A:T bps. The modelling used the structures of the self-complementary duplex sequences d(CGCGAATTCGCG) and d(CGCAAATTTGCG), as found in several relevant drug complexes.[9,10] It suggested that the central 5'-AATT sequence would be an optimal site for the head-to-head motif and for maintenance of helical register with all four consecutive bps without necessitating significant DNA conformational change. An initial range of head-to-head bis-benzimidazoles have now been synthesised, and is reported here.

The highly convergent synthesis of these bis-benzimidazoles (Scheme 1) involves the condensation between the requisite 4-substitituted aryl aldehyde and 3,3′,4,4′-tetraaminobiphenyl **7**. The latter is commercially available, but is of variable quality, and we preferred to prepare it by the route shown. Thus, **1** and **2** were easily obtained from 4-methoxybenzaldehyde and 4-hydroxybenzaldehyde, respectively, while **3** was obtained from 4-(3-dimethylaminopropoxy)benzaldehyde prepared from 4-hydroxybenzaldehyde and 3-dimethylaminopropanol using a Mitsunobu reaction. In each instance the aryl aldehyde and 3,3′,4,4′-tetraaminobiphenyl were heated in nitrobenzene at 150 °C for 8–10 h to effect the condensation, and the yields of bis-benzimidazoles were typically 25–40% on the 10 mmol scale.

Crystallographic analyses have been undertaken on (i) the bis(hydroxyphenyl) compound **2** co-crystallised with the oligonucleotide sequence d(CGCGAATTCGCG), and (ii) its Me_2N derivative **3** with the sequence d(CGCAAATTTGCG). Both crystal structures[11] show a B-form DNA double helix with ligand bound in the A/T region of the minor groove [Fig. 1(*a*)]. They show H-bonding between all four central A : T bps and the benzimidazole subunits [Fig. 1(*b*)], in full accord with the modelling predictions. Each inner-facing ring nitrogen atom of



Scheme 1 *Reagents and conditions*: i, Ac_2O, AcOH; ii, $Hg(OAc)_2$, AcOH, $HClO_4$, *ca.* 20% (2 steps); iii, Cu, $PdCl_2$, Py; iv, HNO_3, AcOH, *ca.* 50% (2 steps); v, H_2SO_4; vi, Raney Ni, H_2, *ca.* 70% for 2 steps; vii, 4-XC_6H_4CHO [X+MeO (**1**), OH (**2**) or $Me_2N(CH_2)_3O$ (**3**)], $PhNO_2$, 150 °C.

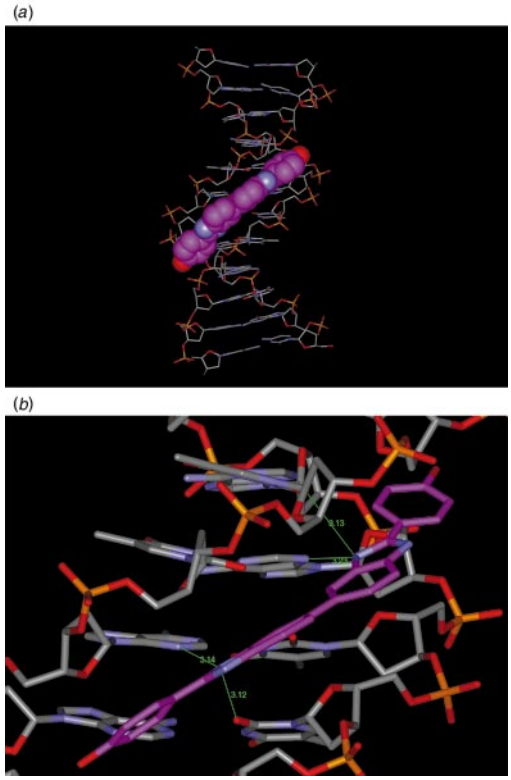


Fig. 1 Computer plots of the crystal structure of the complex of **2** with d(CGCGAATTCGCG): (*a*) shows the ligand, in space-filling mode, centrally bound in the minor groove of the 12-mer double helix, while (*b*) shows a detailed view of **2** and the H-bond contacts to A : T bp edges.

Table 1 Cytotoxicity of *bis*-benzimidazole compounds, for 96 h exposure in a panel of four ovarian cell lines (as $IC_{50}/\mu M$)

Compound	A2780	A2780-Pt^R	CH1	SKOV-3
1	13.5	5.1	3.8	16.5
2	4.3	2.7	1.05	16
3	0.235	0.115	0.24	0.375
Hoechst 33258	12.0	9.5	26.5	> 100

a benzimidazole ring H-bonds to a thymine O2 atom of one bp and an adenine N3 atom on the opposite strand of an adjacent one. It is notable that the absence of formal cationic charge in **2**, does not affect its sequence selectivity, suggesting that the role of charge is to modulate affinity rather than to determine sequence selectivity.

These preferences have also been observed in DNase I footprinting studies with a natural DNA fragment, which have shown that **2** and **3** both interact preferentially with some A/T sites. Compound **3** binds to AT sites between 7–10-fold better than **2**, and has an especially high affinity for the 5′-AATT site (30 nM). Quantitative footprinting data indicated that the order of preference is 5′-AATT > 5′-ATAT > 5′-TAAT > 5′-TATA ~ 5′-TTAA, similar to that previous determined for Hoechst 33258. The increased affinity of **3** relative to **2** does not affect its sequence selectivity. This enhanced binding strength may be a consequence of its cationic charge or its increased DNA site size, since the crystal structure of the complex with $d(CGCAAATTTGCG)_2$ shows it to cover 6 bps. The ability of the head-to-head bis-benzimidazoles to discriminate between the 136 possible tetranucleotide sequences may be further enhanced by flanking sequence preferences. These remain to be systematically explored. Whilst this paper was in preparation, a report has appeared which describes compound **2** and its interactions with tRNA.[12]

These new bis-benzimidazoles have been examined for potential cytotoxic effects in a group of ovarian carcinoma cell lines. Table 1 shows that all compounds are cytotoxic at the μM level, with activity significantly greater than that shown by Hoechst 33258. The Me_2N derivative **3** is over 10-fold more active than the others in three out of the four lines. **3** also shows significant activity in several ovarian lines in the NCI 60-cell line panel (to be published). The compound with least *in vitro* activity, the hydrophobic dimethoxy derivative **1**, has been examined in an *in vivo* tumour model with the ADJ/PC6 plasmacytoma, where it showed 44% tumour inhibition with ip administration. Compound **3** has been examined in an *in vitro* transcription assay. It effectively inhibits transcription at a number of A/T sites, consistent with the above data. In addition, compound **3** inhibited transcription at least 10-fold more efficiently than compound **2**, consistent with its increased cytoxicity. We speculate that the activity of **3** *in vitro* and *in vivo* may be related to its potent ability to specifically inhibit transcription at a small sub-set of A/T sites. The relatively low affinity of compound **3** for 5′-TATA sites suggests that it is less likely to block TATA-box binding proteins at the start of transcription.

This work has been supported by grants from the Cancer Research Campaign, the Royal Society of Chemistry (Adrien Albert Memorial Fund) and the Ghanaian Government.

Notes and references

1 C. Zimmer and U. Wähnert, *Prog. Biophys. Mol. Biol*., 1986, **47**, 31; P. B. Dervan, *Science*, 1986, **232**, 464.
2 See for example: J. J. Welch, F. J. Rauscher III and T. A. Beerman, *J. Biol. Chem*., 1994, **269**, 31 051; S.-Y. Chiang, J. C. Azizkhan and T. A. Beerman, *Biochemistry*, 1998, **37**, 3109.
3 For example: Z. Xu, T.-K. Li, J. S. Kim, E. J. LaVoie, K. J. Breslauer, L. F. Liu and D. S. Pilch, *Biochemistry*, 1998, **37**, 3558; J. S. Kim, Q. Sun, C. Yu, A. Liu, L. F. Liu and E. J. LaVoie, *Bioorg. Med. Chem*., 1998, **6**, 163.
4 S. Neidle, *Biopolymers*, 1997, **44**, 105.
5 A. Fede, A. Labhardt, W. Bannwarth and W. Leupin, *Biochemistry*, 1991, **30**, 11 377; J. A. Parkinson, S. E. Ebrahimi, J. H. McKie and K. T. Douglas, *Biochemistry*, 1994, **33**, 8442; C. E. Bostock-Smith, C. A. Laughton and M. S. Searle, *Nucleic Acids Res*., 1998, **26**, 1660.
6 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.
7 S. Neidle, *FEBS Lett.*, 1992, **298**, 97; A. Czarny, D. W. Boykin, A. A. Wood, C. M. Nunn, S. Neidle, M. Zhao and W. D. Wilson, *J. Am. Chem. Soc*., 1995, **117**, 4716.
8 J. M. Gottesfield, L. Neely, J. W. Trauger, E. E. Baird and P. B. Dervan, *Nature*, 1997, **387**, 205; S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird and P. B. Dervan, *Nature*, 1998, **391**, 468; L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. M. Mosier, J. M. Gottesfield and P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 12 890.
9 N. Spink, D. G. Brown, J. V. Skelly and S. Neidle, *Nucleic Acids Res*., 1994, **22**, 1607; G. R. Clark, D. W. Boykin, A. Czarny and S. Neidle, *Nucleic Acids Res*., 1997, **25**, 1510.
10 G. R. Clark, E. J. Gray, S. Neidle, Y.-H. Li and W. Leupin, *Biochemistry*, 1996, **35**, 13 745.
11 Compound **2** was co-crystallised with d(CGCGAATTCGCG) (from Oswell DNA Service, Southampton) using the hanging-drop method at 286 K. Elongated prismatic crystals were obtained from a 1.5 mM solution of 1 in 50% w/w MPD, in a droplet containing 10 mM $MgCl_2$, 60 mM KCl, 0.75 mM DNA and 6 mM spermine, equilibrated against 30% w/w MPD. The crystals are orthorhombic, space group $P2_12_12_1$, with cell dimensions a = 25.46(2), b = 39.96(4) and c = 65.66(7) Å. Intensity data were collected from a flash-frozen crystal maintained at 100 K with a RAXIS-4 image plate detector, using Cu-Kα radiation from a rotating-anode generator and mirror-focussing optics. A total of 6380 unique reflections were obtained after merging Friedel equivalents (R_{merge} = 0.044) to a resolution of 1.8 Å. The structure was solved using molecular replacement, and refined with the SHELX-97 package. The final model included 174 water molecules, and had an R of 0.229 for 6157 reflections with $F > 4\sigma(F)$, and 23.2% for all data to 1.8 Å. R_{free} was 25.1%. Coordinates and reflection data have been deposited at the Nucleic Acid Database as entry no. dd0013. See http://www.rsc.org/suppdata/cc/1999/929/ for a 3D image in .pdb format.
12 E. V. Bichenkova, S. E. Sadat-Ebrahimi, A. N. Wilton, N. O'Toole, D. S. Marks and K. T. Douglas, *Nucleosides Nucleotides*, 1998, **17**, 1651.

Communication 9/01074B