Articles

Sequence-Selective Targeting of Long Stretches of the DNA Minor Groove by a Novel Dimeric Bis-benzimidazole†

Alexandra Joubert,‡ Xiao-Wen Sun,§ Eric Johansson,⊥ Christian Bailly,‡ John Mann,§ and Stephen Neidle*,⊥

INSERM U-524 et Laboratoire de Pharmacologie Antitumorale du Centre Oscar Lambret, IRCL, Place de Verdun, 59045 Lille, France, School of Chemistry, Queen's University Belfast, Belfast BT9 5AG, Northern Ireland, and Cancer Research UK Biomolecular Structure Group, The School of Pharmacy, University of London, 29−39 Brunswick Square, London WC1N 1AX, U.K.

Received September 30, 2002; Revised Manuscript Received March 21, 2003

ABSTRACT: A dimeric bis-benzimidazole molecule has been designed by computer modeling to bind to a DNA sequence via the DNA minor groove that covers a complete turn of B-DNA. A series of bis-benzimidazole dimers incorporating a $-O-(CH_2)_n-X-CH_2)_n-O-$ linker, with $n = 2$ or 3 and X = O or $N^+H(Me)$, were screened for their capacity to fit the DNA minor groove. The modeling studies enabled an optimal linker to be devised ($n = 3$, X = $N^+H(Me)$), and the synthesis of the predicted "best" molecule, *N*-methyl-*N*,*N*-bis-3,3-{4′-[5″-(2‴-*p*-methoxyphenyl)-5‴-1*H*-benzimidazolyl]-2″-1*H*-benzimidazolyl}-phenoxypropylamine (**5**), is reported. The optimized linker permits the two symmetric bis-benzimidazole motifs to maintain hydrogen-bonded contacts with the floor of the DNA minor groove. DNase I footprinting studies have shown that this ligand binds with high affinity to sequences representing approximately a complete turn of B-DNA, represented by the $[A\cdot T]_4$-[G·C]-$[A\cdot T]_4$ motif, and only poorly to sequences of half this site size, in accord with the computer modeling studies. Compound **5** does not show acute cellular cytotoxicity, in contrast with its monomeric bis-benzimidazole precursors, yet is rapidly taken up into cells.

There is currently much interest in the design of small molecules that bind to DNA with sequence selectivity (*1*−*6*). Almost all are targeted to the B-DNA minor groove. Such molecules would be expected to influence the binding of regulatory proteins and thus act as synthetic regulators of gene expression. This has been achieved in several instances, with down-regulation of transcription of a number of targeted genes being observed (*7*−*11*). Much of this work has involved the polyamide dimers pioneered by Dervan and colleagues, which contain both modified and unmodified pyrrole and imidazole subunits, and which bind in the minor groove as side-by-side dimers (*1*−*3*). Many of these compounds, as well as those based on other motifs, have only shown specificity for relatively short base pair sequences, in large part because of the difficulty in ensuring that longer ligands remain in register with all the base pairs in a target sequence of >10 base pairs. Some success has been achieved in recognizing longer sequences by the use of flexible amino acids to replace some ring units. High affinity for 10 base pair sites has been achieved, for example, with β-alanine-linked polyamides for 1:1 recognition (*12*) and also with head-to-head hairpin dimers (*13*).

Several groups have concentrated on a strategy that utilizes the pharmacophore-like benzimidazole motif derived from the well-established bis-benzimidazole drug Hoechst 33258 (**1**), which itself shows some antitumor activity and is an inhibitor of DNA topoisomerase I (*14*−*17*). Hoechst 33258 itself has been shown by footprinting and structural methods to recognize 3−4 consecutive A/T base pairs by hydrogen-bonding to base edges in 1:1 complexes (*18*−*24*). To recognize longer sequences, a number of benzimidazole trimers based on the Hoechst motif have been synthesized and evaluated for various biological activities (*25*−*30*). Several show experimental antitumor activity that is associated with inhibition of DNA topoisomerase enzymes. All are exclusively A/T selective, with the trimers (together with their substituents) recognizing up to seven base pairs, still as 1:1 complexes. Recognition of yet longer sequences has been achieved with the benzimidazole motif, by incorporating a tripyrrole peptide into the Hoechst 33258 conjugate (**2**), which resulted in high specificity and affinity over nine base pairs, again as a 1:1 complex (*31*, *32*).

Our own contribution to this area involved the design, synthesis, and evaluation of symmetric bis-benzimidazoles,

† This work was supported by Cancer Research UK (grant no. SP1384 to S.N.) and the Ligue Nationale Contre le Cancer (Equipe labellisée LA LIGUE) (to C.B.).

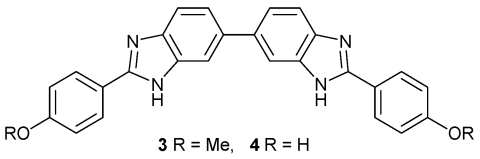
* To whom correspondence should be addressed. E-mail: stephen.neidle@ulsop.ac.uk. Tel.: 44 207 753 5969. Fax: 44 207 753 5970.

‡ Centre Oscar Lambret, IRCL.

§ Queen's University Belfast.

⊥ University of London.

10.1021/bi026926w CCC: $25.00
Published on Web 05/03/2003



3 R = Me, **4** R = H

FIGURE 1: (a) Schematic of the synthesis of compound **5**. (b) Structures of compounds **3** and **4**.

exemplified by molecules **3** and **4** (Figure 1), which not only exhibited high selectivity for the base sequence d(AATT) (a commonly studied motif of the minor groove of B-DNA) but also possessed marked selective cytotoxicity activity across a range of cancer cell lines (*33*, *34*), combined with favorable cellular uptake properties. We have now designed a linker that enables two such symmetric bis-benzimidazole molecules to be linked together such that a long stretch of DNA sequence can be recognized while maintaining all of the benzimidazole subunits in register with the DNA base edges. We describe in this paper the design, synthesis, and preliminary evaluation of the first of these dimeric bis-benzimidazole molecules (**5**), which has the ability to recognize a run of almost 10 consecutive base pairs.

METHODS

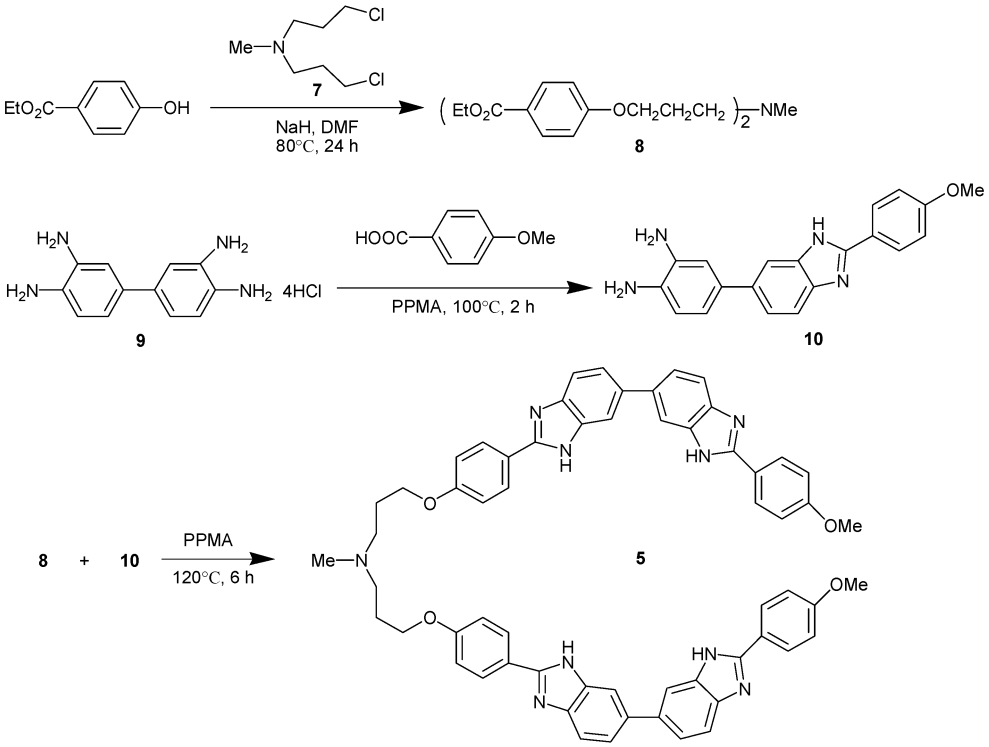
Molecular Modeling. The goal of the modeling was to ascertain whether symmetric bis-benzimidazole units could be linked together in a simple manner to actively recognize the bases in a complete turn of B-DNA. The strategy adopted was to link two of the head-to-head benzimidazole dimers that we have previously designed, and which we have established are effective binders at A/T sequences. Several different linkers were systematically examined, on the basis of the need to be able to adapt optimally to the curvature of the minor groove, and to maintain hydrogen-bonding as far as possible. The general composition of the linker region was finally chosen to be $-O(CH)_nX(CH)_nO-$, with the value of *n* and the nature of X being the variables. The symmetrical DNA sequence d(CGAATTCGAATTCG) was used as the template for ligand design, since it contains two of the d(AATT) motifs that we have previously shown to be hydrogen-bonded to the benzimidazole subunits. A canonical B-DNA duplex structure was assumed as a starting point for the sequence, which was constructed using the BIOPOLYMER module in the InsightII package (*35*). The duplex was energy-minimized using the AMBER99 (*36*) force field in the HYPERCHEM (version 7) program (*37*) until convergence was achieved, with a gradient of <0.1 kcal^{-1} mol^{-1} Å^{-1}. A distance-dependent dielectric constant was used, $\epsilon = 4.0r_{ij}$. The various ligands were constructed using the HYPERCHEM package, and partial charges were calculated semiempirically using the MOPAC method (*38*). The energy-minimized ligands were manually docked into the DNA and positioned such that the benzimidazole subunits were available to interact in the two AATT A-tracts in the same manner as observed from our previous crystallographic studies. Care was taken to ensure that the starting positions for each ligand were as similar as possible and that each subunit was maximally in register with DNA base edges. Each DNA complex structure was then energy-minimized until convergence was judged to have occurred, with a gradient of <0.1 kcal^{-1} mol^{-1} Å^{-1}. Equivalent calculations were also performed on structures for eight other sequences, including several that were suggested by the subsequent footprinting studies (see below). Canonical B-DNA duplex structures were also used as starting points for these calculations. In each case, the initial ligand starting positions were identical.

Synthetic Chemistry. Compound **5** was assembled as shown in Scheme 1. Thus, *N*-methyl-*N*,*N*-bis(3-hydroxypropyl)amine (**6**) was reacted with thionyl chloride to produce the corresponding dichloride **7** (60% yield), and this was reacted with 2 equiv of the sodium salt of ethyl 4-hydroxybenzoate to produce the amine **8** (95%). The hydrochloride salt of 3,4,3′,4′-tetraamino-*p*-biphenyl (**9**) was condensed with 4-methoxybenzoic acid using phosphorus pentoxide dissolved in methanesulfonic acid (PPMA) as both solvent and reagent to provide the benzimidazole **10** (32%). This condensation between diamine and ester using PPMA is far superior to our previous method (*33*), where condensation occurred between the diamine and the corresponding aryl aldehyde when nitrobenzene was used as solvent and oxidant, and gram quantities of bis-benzimidazoles can now be prepared. Finally, benzimidazole **10** (2 equiv) was reacted with amine **8**, again in solution in PPMA. This produced the desired dimeric bis-benzimidazole **5** (31%) as a yellow solid after crystallization from DMF−MeOH−water. Further details and analytical data are given elsewhere (*39*).

Purification of DNA Restriction Fragments and Radiolabeling. The different plasmids were isolated from *Escherichia coli* by a standard sodium dodecyl sulfate−sodium hydroxide lysis procedure and purified by banding in cesium chloride−ethidium bromide gradients. The 117-bp DNA fragment was prepared by 3′-[^{32}P]-end labeling of the *Eco*RI-*Pvu*II double digest of the pBS plasmid (Stratagene) using α-[^{32}P]-dATP (Amersham, 3000 Ci/mmol) and AMV reverse transcriptase (Roche). Similarly, the 160-bp *tyr*T fragment was prepared by 3′-end-labeling of the *Eco*RI−*Ava*I digest of plasmid pKMΔ98 (*40*). The two 198-bp fragments were obtained from plasmids pMS1 and pMS2 (kindly provided by Professor K. R. Fox, University of Southampton) after digestion with the restriction enzymes *Hin*dIII and *Xba*I (*41*). In each case, the labeled digestion products were separated on a 6% polyacrylamide gel under nondenaturing conditions in TBE buffer (89 mM Tris−borate, pH 8.3, 1 mM EDTA). After autoradiography, the requisite band of DNA was excised, crushed, and soaked in water overnight at 37 °C. This suspension was filtered through a Millipore 0.22-μm filter, and the DNA was precipitated with ethanol. Following washing with 70% ethanol and vacuum-drying of the precipitate, the labeled DNA was resuspended in 10 mM Tris, adjusted to pH 7.0, containing 10 mM NaCl.

DNase I Footprinting. Bovine pancreatic deoxyribonuclease I (DNase I, Sigma Chemical Co.) was stored as a 7200 units/mL solution in 20 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$, pH 8.0. The stock solutions of DNase I was kept at −20 °C and freshly diluted to the desired concentration immediately prior to use. Footprinting experiments were performed essentially as previously described (*42*). Briefly, reactions were conducted in a total volume of 10 μL. Samples (3 μL) of the labeled DNA fragments were incubated with 5 μL of the buffered solution containing the ligand at the appropriate concentration. After 30 min of incubation at 37 °C to ensure equilibration of the binding reaction, the

Scheme 1: Synthesis of the Dimeric Bis-benzimidazole **5**



digestion was initiated by the addition of 2 μL of a DNase I solution, the concentration of which was adjusted to yield a final enzyme concentration of about 0.01 unit/mL in the reaction mixture. The reaction was stopped by freeze-drying after 3 min. Samples were lyophilized and resuspended in 5 μL of an 80% formamide solution containing tracking dyes. The DNA samples were then heated at 90 °C for 4 min and chilled in ice for 4 min prior to electrophoresis.

Electrophoresis and Quantitation by Storage Phosphor Imaging. DNA cleavage products were resolved by polyacrylamide gel electrophoresis under denaturating conditions (0.3 mm thick, 8% acrylamide containing 8 M urea). After electrophoresis (about 2.5 h at 60 W, 1600 V in Tris−borate−EDTA buffered solution, BRL sequencer model S2), gels were soaked in 10% acetic acid for 10 min, transferred to Whatman 3MM paper, and dried under vacuum at 80 °C. A Molecular Dynamics 425E PhosphorImager was used to collect data from the storage screens exposed to dried gels overnight at room temperature. Baseline-corrected scans were analyzed by integrating all the densities between two selected boundaries using ImageQuant (version 3.3) software. Each resolved band was assigned to a particular bond within the DNA fragments by comparison of its position relative to sequencing standards generated by treatment of the DNA with dimethylsulfate followed by piperidine-induced cleavage at the modified guanine bases in DNA (G-track).

RESULTS AND DISCUSSION

Molecular Modeling. The modeling has shown that a simple linker group can be devised such that both bis-benzimidazole dimer groups can simultaneously bind into the minor groove of the target DNA (Figure 2), with all four benzimidazole groups interacting with A·T base pairs by hydrogen-bonding. The four complexes examined have been ranked on the basis of two criteria; first, their calculated relative binding energies with the target DNA sequence used in the design phase of the study (Table 1), and second, their ability to establish hydrogen bonds with the base-edge acceptor atoms in the minor groove. This latter was assessed

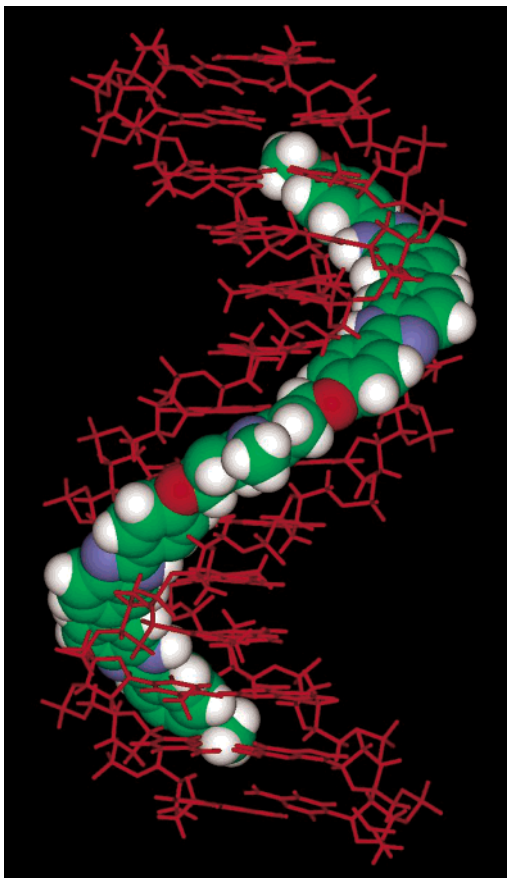


FIGURE 2: Energy-minimized model for the interaction of compound **5** with DNA.

Table 1: Calculated Relative Interaction Energies (kcal mol^{-1}) for Bis-benzimidazole Dimers with Various Linkers, for Minimized Structures

linker	interaction energy
$-(CH_2)_2-N^+(HMe)-(CH_2)_2-$	−181.5
$-(CH_2)_3-N^+(HMe)-(CH_2)_3-$	−191.2
$-(CH_2)_2-O-(CH_2)_2-$	−159.2
$-(CH_2)_3-O-(CH_2)_3-$	−141.7

[a] Compound **5** incorporates the linker $-(CH_2)_3-N^+(HMe)-(CH_2)_3-$.

Table 2: Calculated Relative Interaction Energies (kcal mol^{-1}) for the Bis-benzimidazole Dimer Compound **5** with Various 14-mer DNA Sequences

sequence	interaction energy
CGAATTCGAATTCG	−184.6
CGATATCAAATTCG	−187.7
CGAATTGTAATACG	−194.3
CGTTATGTAAAACG	−192.8
CGAAAATGAACTCG	−186.2
CGAAAAATTTTTCG	−188.9
CGATATATATATCG	−194.5
CGCGCGCGCGCGCG	−162.3
CGATTAGAATTACG	−188.4

by the pattern of distances between the corresponding hydrogen-bond donors and acceptors. In view of the lack of explicit solvent in the calculations, we have used the modeling only as an indicative guide for the design and subsequent solution studies. However since (i) the experimental binding sites are generally in accord with the predictions (see below) and (ii) the experimental site size found, of ∼9−10 base pairs, is in accord with the model, we suggest that the modeled structures are realistic approximations to the complexes in solution.

The two ligands having the central positively charged nitrogen atom in the linker were predicted to be best able to bind in the minor groove, since their calculated relative interaction energies (Table 2) are significantly greater than those of the two ligands with the −O− linker groups. This difference is primarily a consequence of the smaller intermolecular electrostatic energy contributions of the latter two ligands to the total binding energy compared to the energy contributed by those containing the charged linker nitrogen atom. All four of the linkers examined have at least some degree of effective van der Waals and hydrogen-bond contacts with the DNA minor groove floor, for all four benzimidazole subunits in all four compounds. We find that compound **5**, which has a more favorable interaction energy than its analogue with $-(CH_2)_2-$ linker groups, also has hydrogen-bonding geometries overall superior to those of the bases in the target sequence, and we therefore considered it to be the optimal compound of the four. Visual inspection shows that the full length of this optimized dimer ligand covers a total of 11−12 base pairs (Figure 2), although the total direct base−ligand hydrogen-bonding site size is nine base pairs (Figure 3). The four benzimidazole subunits are required to be highly twisted in relation to each other, to achieve hydrogen-bonding with A/T base edges (Figures 2 and 3; Table 3). The crystal structures of the head-to-head bis-benzimidazole ligands (*33*, *34*) have each dimer hydrogen-bonding to four consecutive A/T bases. The model presented here has two benzimidazole units hydrogen-bonding in this way (Figure 3), with the other two hydrogen-bonding to three

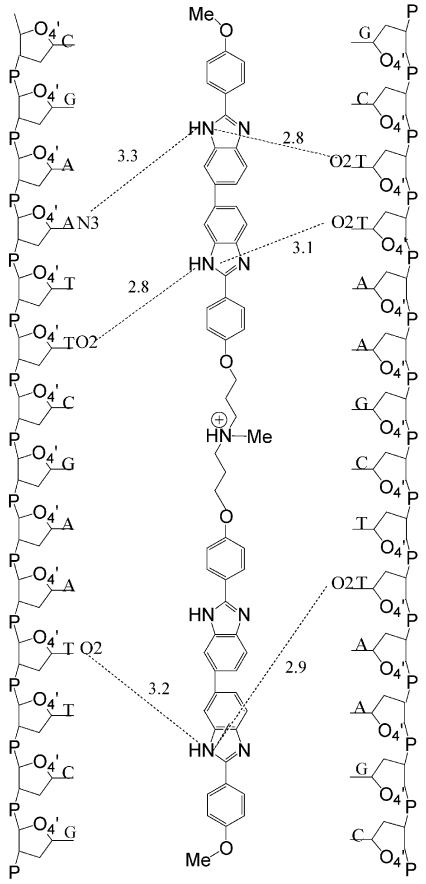
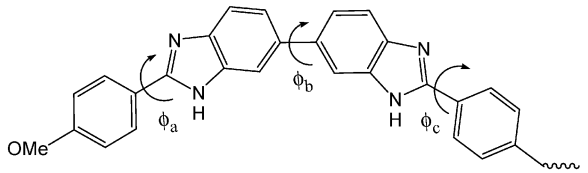


FIGURE 3: Schematic representation showing the possible hydrogen-bond contacts of **5** with minor groove base atoms in the DNA sequence d(CGAATTCGAATTCG).

Table 3: Benzimidazole Torsion Angles in the Optimal Conformation of Compound **5**[a]



	ϕ_{a1}	ϕ_{b1}	ϕ_{c1}	ϕ_{a2}	ϕ_{b2}	ϕ_{c2}
torsion angle (deg)	13	16	11	9	14	12

[a] Each angle is defined by the four atoms facing into the minor groove, as shown for one bis-benzimidazole moiety. The two bis-benzimidazoles are designated 1 and 2, respectively, in the table.

consecutive A/T bases and not making the maximum number of possible hydrogen bonds. These differences are due to the slight winding of DNA at this point in the structure, where A-like features are apparent.

The modeling data indicate that the dimers containing the $-N^+H(Me)-$ linker would have a binding site that is fully in register with the DNA B helix. The overall occluded site size, of effectively 11.5 base pairs, is a marked increase in length compared to the conventional head-to-tail bis-benzimidazole dimers, as in Hoechst 33258, with at most a 4.5-base-pair site, or a tris-benzimidazole, with a 7.5-base-pair site size. The predicted ability of both benzimidazole ends of ligand **5** to effectively and specifically hydrogen-bond to the target DNA sequence throughout the binding site shows

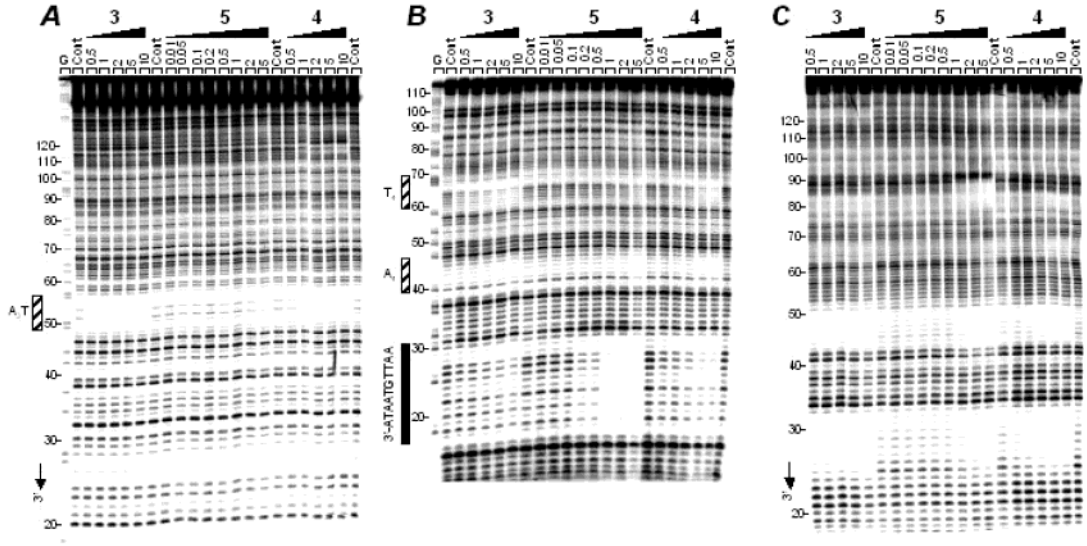


FIGURE 4: Gels showing DNase I footprinting with (A) 198-mer, (B) 117-mer, and (C) the 160-mer restriction fragments cut from plasmids pMS2, pBS, and pMΔ98, respectively. In each case, the DNA was 3′-end-labeled with [α-^{32}P]-dATP in the presence of AMV reverse transcriptase. The products of nuclease digestion were resolved on an 8% polyacrylamide gel containing 7 M urea. Control tracks (Cont) contained no drug. The concentration (μM) of the test drug is shown at the top of the appropriate gel lanes. Tracks labeled "G" represent dimethylsulfate−piperidine markers specific for guanines. Numbers on the left side of the gels refer to the standard numbering scheme for the nucleotide sequence of the DNA fragment, as indicated in Figure 5. The black and hatched boxes on gel B indicate the positions of the two binding sites, with the corresponding DNA sequence, used for the quantitative analysis presented in Figure 6.

that the problem of helical phasing being a limitation to the recognition of long sequences has been overcome in this instance. Accordingly, the synthesis of the dimeric bis-benzimidazole **5**, incorporating a propylaminopropyl-type linker, was undertaken on the basis of these results from the modeling study. For synthetic reasons, we chose early in the project to prepare compound **5** with a central *N*-methyl group, as indicated in Scheme 1 (and as modeled).

The interactions of compound **5** with a range of sequences having variations in the central 10 base pairs were modeled after the footprinting studies, using the same protocol as employed with the four different ligands. Comparison of the interaction energies (Table 2) shows first that the site with all C·G base pairs is energetically highly disfavored. It is striking that sites containing one or two central C·G base pairs are not disfavored compared to all-A·T ones. There appears to be a slight preference for sites with a central sequence 5′-TGT, although this conclusion must be tentative in view of the fact that only a small sampling of the possible sequences has been performed. The modeling also shows that compound **5** has high affinity for a purely alternating A/T sequence. However, this is probably more a reflection of the sequence in the vicinity of the linker and is consistent with the slight preference for an alternating pyrimidine−purine−pyrimidine sequence in the center of the binding site.

DNA Footprinting. DNase I footprinting methodology has been used to investigate the sequence selectivity of the dimeric bis-benzimidazole compound **5** and its monomeric bis-benzimidazole precursors. Four DNA restriction fragments were prepared by 3′-end-radiolabeling with ^{32}P. On one hand, we used fragments with random DNA sequences such as the 117-mer from plasmid pBS and the 160-mer *tyr*T fragment containing the tyrosine tRNA promoter. These two DNA pieces have been used previously to study sequence recognition by a variety of minor groove binders, including bis-benzimidazoles (*43*, *44*). On the other hand, we employed two designed fragments of 198 bp, referred to as "universal footprinting substrates" (*41*), containing all 136 distinguishable tetranucleotide sequences [$(4^4)/2 + (4^{4/2})/2 = 136$]. These two fragments, MS1 and MS2, contain the same sequence of 136 bp but cloned in the opposite orientations, 5′ → 3′ and 3′ → 5′. With each fragment, the products of digestion by DNase I in the absence and in the presence of the test drugs were resolved by polyacrylamide gel electrophoresis. Typical DNase I digestion patterns observed in the absence and in the presence of the benzimidazole compounds are illustrated in Figure 4. Significant differences between the dimeric compound and the monomeric analogues can easily be seen. For example, the simple bis-benzimidazole compounds **3** and **4** both produce a marked footprint at site 5′-AAAT located near position 50 within the MS2 fragment, whereas the dimer **5** fails to bind to this site (Figure 4A). On the other hand, a clear footprint can be detected with **5**, but not with the monomers, around position 85 on the 160-mer at a site encompassing the sequence 5′-ATATCAAAT (Figure 4C). Another marked difference can be seen with the 117-mer fragment, where compound **5** strongly protects the sequence 5′-AATTGTAATA from cleavage by DNase I, whereas the protection is considerably weaker with **3** or **4** (Figure 4B). A densitometric analysis of the different patterns was performed to estimate the location and relative strength of binding at particular DNA sites. Figure 5 shows the differential cleavage plots determined with the four DNA fragments in the presence of the bis-benzimidazole compounds. The dips in these plots (negative values) indicate sites of protection from DNase I cleavage, whereas peaks (positive values) indicate regions of drug-induced enhancement of nuclease cleavage. A total of eight binding sites were identified for compounds **3** and **4**. These two bis-benzimidazoles produce identical cleavage patterns, but in some cases the footprints are slightly more pronounced with the OH compound **4** compared to its methoxy derivative **3**. With the 117-mer and 160-mer DNA fragments, the binding sites observed with these two compounds are identical to those

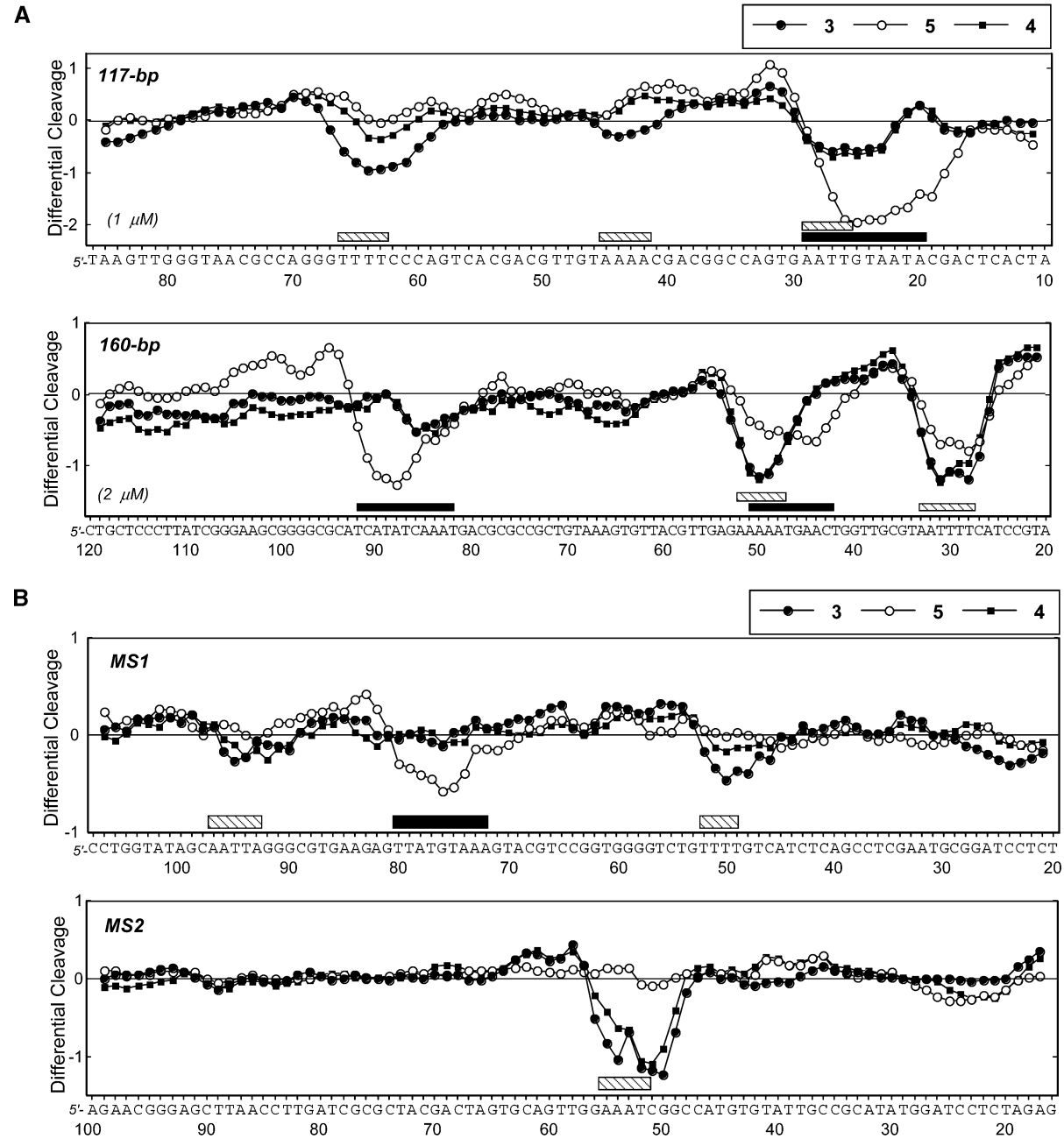


FIGURE 5: Differential cleavage plots comparing the susceptibility of the different restriction fragments to DNase I cutting in the presence of the bis-benzimidazole compounds [(A) 117-bp at 1 μM and 160-bp at 2 μM; (B) MS1 and MS2 each at 2 μM]. Negative values correspond to a ligand-protected site, and positive values represent enhanced cleavage. Vertical scales are in units of $\ln(f_a) - \ln(f_c)$, where f_a is the fractional cleavage at any bond in the presence of the drug and f_c is the fractional cleavage of the same bond in the control, given closely similar extents of overall digestion. Each line drawn represents a three-bond running average of individual data points, calculated by averaging the value of $\ln(f_a) - \ln(f_c)$ at any bond with those of its two nearest neighbors. Only the region of the restriction fragment analyzed by densitometry is shown. Black boxes indicate the positions of inhibition of DNase I cutting in the presence of the dimeric compound **5**, and hatched boxes refer to the positions of binding sites for the monomeric bis-benzimidazole compounds **3** and **4**.

recently reported with the same pieces of DNA with Hoechst 33258, Hoechst 33342, and related head-to-tail bis-benzimidazoles (*43*). The eight binding sites for **3** and **4** contain at least four contiguous A·T base pairs, such as 5′-TTTT, 5′-AAAT, and 5′-AATT, for example. The footprints are strictly restricted to AT-rich sequences and in particular those that lack a TpA step. For example, compounds **3** and **4** both bind strongly to the sequence 5′-AAAT near position 52 but not to the sequence 5′-TTAA around position 87 within the MS2 fragment (Figure 5). Similarly, sequences such as 5′-ATAT and 5′-TTAT are poorly recognized. As recently discussed (*45*), a TA step at one end of a 4-bp AT sequence is quite detrimental to minor groove binding, presumably because it has the effect of widening the minor groove and disrupting the spine of hydration motif (*46*). $d(TTAA)_2$ tracts display a wider minor groove than $d(AATT)_2$ segments and a disordered hydration pattern (*47*).

The situation is radically different with compound **5**. On one hand, this dimeric bis-benzimidazole compound does not bind to short $[A{\cdot}T]_4$ tracts. This is particularly obvious with the MS2 fragment, which does not provide any good binding site for it. On the other hand, this compound produces marked footprints at several well-defined A/T-rich sites where the monomeric compounds **3** and **4** bind only

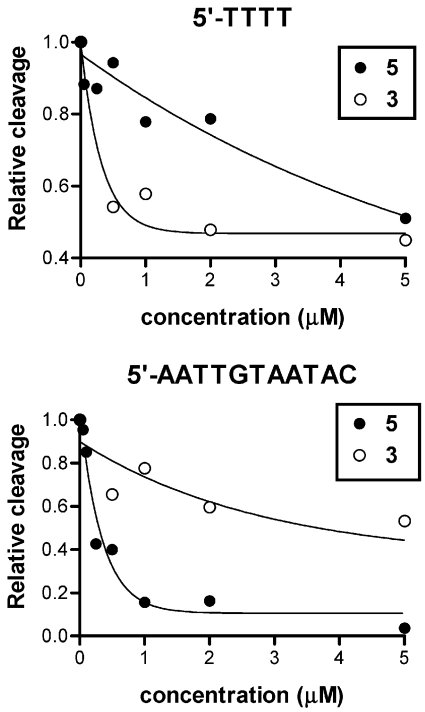


FIGURE 6: Footprinting plots for the binding of the monomeric (**3**) and dimeric (**5**) compounds to the sites 5′-TTTT and 5′-AATTGTAATAC. The relative cleavage intensity corresponds to the ratio I_c/I_o, where I_c is the intensity of the given bands at the ligand concentration c and I_o is the intensity of the same band in the absence of the test drug.

weakly. This is particularly clear for the sequences 5′-TTATGTAAA (positions 72−80 in MS1), 5′-ATATCAAAT (positions 82−90 on the 160-mer), and 5′-AATTGTAATA (positions 20−29 on the 117-mer). These three sequences present a common motif [A·T]$_4$-[G·C]-[A·T]$_4$ which is well-suited to accommodate the extended minor groove binder, in accord with the expected recognition site for **5**. These have the two bis-benzimidazole moieties each recognizing an [A·T]$_4$ tract and the linker bridging the central G·C base pair. The [A·T]$_4$-[G·C]-[A·T]$_4$ motif clearly provides a high-affinity binding site for **5** but not for the monomeric compounds **3** and **4**, which at best recognize only one-half of the motif. It should be noted that weak footprints were also detected with **5** at two sequences that do not match with the optimal motif: 5′-AAAATGAACT (positions 42−51) and 5′-AATTTTC (positions 27−33) within the *tyr*T fragment. In the first case, the large size of the footprint suggests that the two halves of the dimer are both effectively bound to DNA, whereas in the second case, it is possible that only one-half of the dimer is inserted in the minor groove because the footprint at this site is comparable with those produced by **3** and **4**. Nevertheless, footprints at these two sites are much weaker compared to those at the sites corresponding to the preferred binding site sequence motif [A·T]$_4$-[G·C]-[A·T]$_4$. It is also apparent that A/T sequences of 7−8 consecutive A·T base pairs, such as 5′-AAATTAA or 5′-AAAAATAT, are not protected from DNase I cleavage by **5** (data not shown). This is unsurprising since these are not quite long enough to accommodate the predicted binding site size of compound **5**. Thus, at present, we cannot exclude the possibility that longer A/T tracts, with nine or more A·T base pairs, would provide suitable binding sites for **5**. A site-selection method is well-suited to define more precisely the sequences of all potential binding sites, and we propose to examine this question in future studies.

A concentration-dependent analysis was performed to better compare the magnitude of binding of the two bis-benzimidazole monomeric compounds **3** and **4** to two distinct binding sites within the 117-mer fragment. The footprint plots presented in Figure 6 show most clearly that **5** binds poorly to the site containing 5′-TTTT but shows a high affinity for the sequence 5′-AATTGTAATAC. On the other hand, compound **3** binds much more weakly to this later site and, in fact, recognizes only the 5′-half (i.e., AATT) of this sequence (Figure 5). For **5**, the drug concentration required for half-maximal footprinting (the so-called C_{50} value, which approximates the dissociation constant for binding to an individual site) is about 0.15 μM for the 5′-AATTGTAATAC site, whereas it is >5 μM for the 5′-TTTT site. We can

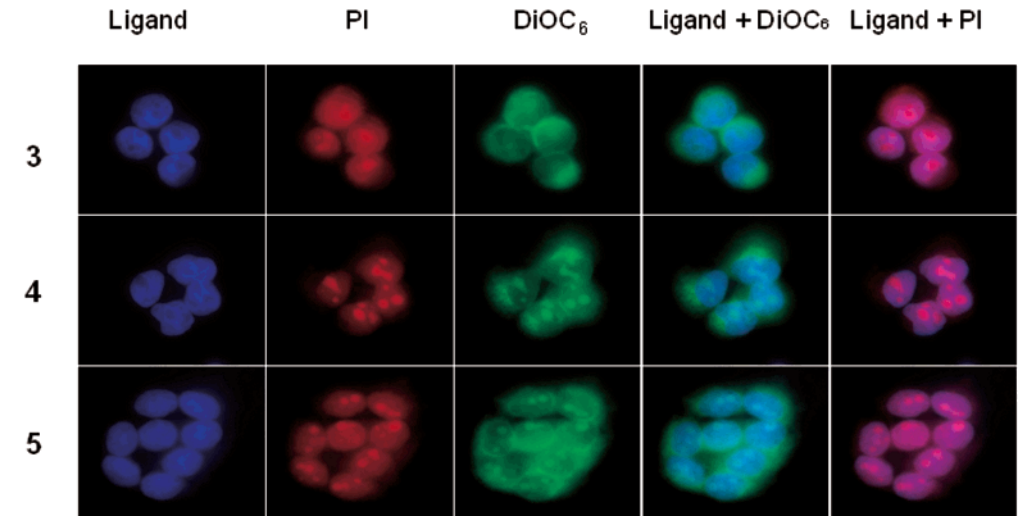


FIGURE 7: Fluorescence micrographs of HT-29 colon carcinoma cells stained with the indicated compound (5 μM each in blue), with propidium iodide (PI in red), or with 3,3-dihexyloxacarbocyanine iodide (DiOC$_6$ in green). Images on the right side of the figure show the overlay of the minor groove binder with PI (blue + red) or with DiOC$_6$ (blue + green). The cells were incubated with the test drug for 24 h, washed, fixed with 2% paraformaldehyde, and then labeled with PI (0.2 μg/mL PI) or DiOC$_6$ (20 nM) prior to the microscopy observation (×63). Results are representative of three independent stainings. The experimental procedure used for these experiments has been recently described in detail (*48*).

therefore conclude that the $[A\cdot T]_4$-[G·C]-$[A\cdot T]_4$ motif represents a high-affinity and selective binding sequence for compound **5**. Molecular modeling studies, as described above, were subsequently undertaken on compound **5** bound to a number of duplex sequences. While these studies do not fully explain the preference for only a single G·C base pair in the $[A\cdot T]_4$-[G·C]-[A·T] motif in energetic terms, they do suggest that the exocyclic N2 amino group of the G does not hinder optimal nonbonded interactions. This finding is at variance with the preferences shown by small, A/T-selective ligands such as the bis-benzimidazoles **3** and **4** (*33*) and Hoechst 33258 itself (*49*). We suspect that this may be a consequence of the lack of structural perturbations induced in DNA by the linker in compound **5**. We are currently studying these issues in more detail.

CONCLUSION AND BIOLOGICAL IMPLICATIONS

We have demonstrated that rational structure-based design principles can be used to design a molecule based on a dimer of the bis-benzimidazole motif that will bind to and selectively recognize a complete turn of double-helical DNA, thus validating the overall design strategy. Footprinting studies have shown not only that this compound binds to the predicted size of site, but also that there are unexpected requirements for particular types of sequences within the site. The ligand has been found experimentally to bind with high affinity to the motif $[A\cdot T]_4$-[G·C]-$[A\cdot T]_4$, whereas the modeling studies suggested that up to two C·G base pair could be tolerated in the central linker region. This is in contrast with the behavior of typical minor groove ligands such as Hoechst 33258 or distamycin, which strongly disfavor C·G base pairs.

Previous compounds in this series studied by us, exemplified by the monomeric bis-benzimidazole compounds **3** and **4**, are biologically active and show significant cytotoxicity in a range of tumor cell lines with IC_{50} values $\ll 1\ \mu$M. By contrast, compound **5** shows poor levels of acute cytotoxicity, with typical values in the $>10\ \mu$M range. Preliminary studies on cellular uptake (Figure 7) show that it is rapidly transported into tumor cells, despite the large size of this molecule. We suggest that the poor cytotoxicity may reflect the reduced number of possible genomic binding sites available to **5**, by contrast to the much larger number for **3** or **4**, as well as the fact that a purely A/T site is no longer necessarily preferred, excluding the ligand from binding at the TATA box and thus inhibiting the initiation of transcription.

These results give confidence that the same approach can be used to design a further set of compounds that would extend the recognition site to 16−18 base pairs. As yet, specificity to unique sequences is not readily achievable with the bis-benzimidazole motif. However, if combined with other methodologies, such as those using polyamides, that are more selective for particular sequences, it should be possible to target a small number of unique genomic sites. The fact that compound **5** is so readily transported into cells suggests that it and its descendents may also be of therapeutic value in the future.

ACKNOWLEDGMENT

We are grateful to The Institute of Cancer Research (where the majority of the molecular modeling studies were undertaken, when E.J. and S.N. were based there) for a Research Studentship to Eric Johansson, to The McClay Trust for a postdoctoral fellowship in support of Xiao-Wen Sun, and to the Service Commun d'Imagerie Cellulaire de l'IMPRT (IFR114, Lille) for access to the fluorescence microscope.

REFERENCES

1. Dervan, P. B. (2001) Molecular recognition of DNA by small molecules. *Bioorg Med. Chem. 9*, 2215−2235.
2. Dervan, P. B., and Buerli, R. W. (1999) Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol. 3*, 688−693.
3. Wemmer, D. E., and Dervan, P. B. (1997) Targeting the minor groove of DNA. *Curr. Opin. Struct. Biol. 7*, 355−361.
4. Reddy, B. S. P., Sondhi, S. M., and Lown, J. W. (1999) Synthetic DNA minor-groove binding drugs. *Pharmacol. Ther. 84*, 1−111.
5. Neidle, S. (2001) DNA minor-groove recognition by small molecules. *Natural Prod. Rep. 18*, 291−309.
6. Neidle, S. (1997) Crystallographic insights into DNA minor groove recognition by drugs. *Biopolymers 44*, 105−121.
7. Chiang, S.-Y., Welch, J., Rauscher, F. J., III, and Beerman, T. A. (1994) Effects of minor groove binding drugs on the interaction of TATA box binding protein and TFIIA with DNA. *Biochemistry 33*, 7033−7040.
8. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Acad. Sci. U.S.A. 95*, 12739−12741.
9. Kwok, Y., Zhang, W., Schroth, G. P., Liang, C. H., Alexi, N., and Bruice, T. W. (2001) Allosteric interaction of minor groove binding ligands with UL9-DNA complexes. *Biochemistry 40*, 12628−12638.
10. White, C. M., Satz, A. L., Bruice, T. C., and Beerman, T. A. (2001) Inhibition of transcription factor−DNA complexes and gene expression by a microgonotropen. *Proc. Natl. Acad. Sci. U.S.A. 98*, 10590−10595.
11. Ehley, J. A., Melander, C., Herman, D., Baird, E. E., Ferguson, H. A., Goodrich, J. A., Dervan, P. B., and Gottesfield, J. M. (2002) Promoter scanning for transcription inhibition with DNA-binding polyamides. *Mol. Cell Biol. 22*, 1723−1733.
12. Urbach, A. R., Love, J. J., Ross, S. A., and Dervan, P. B. (2002) Structure of a β-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1:1 motif. *J Mol. Biol. 320*, 55−71.
13. Weyermann, P., and Dervan, P. B. (2002) Recognition of ten base pairs by head-to-head hairpin dimers. *J. Am. Chem. Soc. 124*, 6872−6878.
14. Ferguson, L. R., and Denny, W. A. (1995) Microbial mutagenic effects of the DNA minor groove binder Pibenzimol (Hoechst 33258) and a series of mustard analogues. *Mutat. Res. 329*, 19−27.
15. Kraut, E., Fleming, T., Segal, M., Neidhart, J., Behrens, B. C., and MacDonald, J. (1999) Phase II Study of Pibenzimol in pancreatic cancer−a Southwest Oncology Group study. *Invest. New Drugs 9*, 95−96.
16. Tolner, B., Hartley, J. A., and Hochhauser, D. (2001) Transcriptional regulation of topoisomerase II alpha at confluence and pharmacological modulation of expression by bis-benzimidazole drugs. *Mol. Pharmacol. 59*, 699−706.
17. Soderlind, K. J., Gorodetsky, B., Singh, A. K., Bachur, N. R., Miller, G. G., and Lown, J. W. (1999) Bis-benzimidazole anticancer agents: targeting human tumour helicases. *Anticancer Drug Des. 14*, 19−36.
18. Harshman, K. D., and Dervan, P. B. (1985) Molecular recognition of B-DNA by Hoechst 33258. *Nucleic Acids Res. 13*, 4825−4835.
19. Spink, N., Brown, D. G., Skelly, J. V., and Neidle, S. (1994) Sequence dependant effects in drug−DNA interaction: the crystal structure of Hoechst 33258 bound to the $d(CGCAAATTTGCG)_2$ duplex. *Nucleic Acids Res. 22*, 1607−1612.
20. Vega, M. C., Coll, M., and Aleman, C. (1996) Intrinsic conformational preferences of the Hoechst dye family and their influence on DNA binding. *Eur. J. Biochem. 239*, 376−383.

21. Parkinson, J. A., Barber, J., Douglas, K. T., Rosamond, J., and Sharples, D. (1990) Minor-groove recognition of the self-complementary duplex d(CGCGAATTCGCG)$_2$ by Hoechst 33258: a high-field NMR study. *Biochemistry 29*, 10181−10190.
22. Embrey, K. J., Searle, M. S., and Craik, D. J. (1993) Interaction of Hoechst 33258 with the minor groove of the A + T-rich DNA duplex d(GGTAATTACC)$_2$ studied in solution by NMR spectroscopy. *Eur. J. Biochem. 211*, 437−447.
23. Czarny, A., Boykin, D. W., Wood, A. A., Nunn, C. M., Neidle, S., Zhao, N., and Wilson, W. D. (1995) Analysis of van der Waals and electrostatic contributions in the interactions of minor-groove binding benzimidazoles with DNA. *J. Am. Chem. Soc. 117*, 4716−4717.
24. Wood, A. A., Nunn, C. M., Czarny, A., Boykin, D. W., and Neidle, S. (1995) Variability in DNA minor groove width recognised by ligand binding. The crystal structure of a bis-benzimidazole compound bound to the DNA Duplex d(CGCGAATTCGCG)$_2$. *Nucleic Acids Res. 23*, 3678−3684.
25. Pilch, D. S., Xu, Z., Sun, Q., LaVoie, E. J., Liu, L. F., and Breslauer, K. J. (1997) A terbenzimidazole that preferentially binds and conformationally alters structurally distinct DNA duplex domains: a potential mechanism for topoisomerase I poisoning. *Proc. Natl. Acad. Sci. U.S.A 94*, 13565−13570.
26. Kim, J. S., Sun, Q., Chaing, Y., Liu, A., Liu, L. F., and LaVoie, E. J. (1998) Quantitative structure−activity relationships on 5-substituted ter-benzimidazoles as topoisomerase in poisons and antitumor agents. *Bioorg. Med. Chem. 6*, 163−172.
27. Xu, Z., Tsai-Kun, L., Kim, J. S., LaVoie, E. J., Breslauer, K. L., Liu, L. F., and Pilch, D. S. (1998) DNA minor groove binding-directed poisoning of human DNA topoisomerase I by ter-benzimidazoles. *Biochemistry 37*, 3558−3566.
28. Ji, Y. H., Bur, D., Hasler, W., Runtz, V., Dorn, A., Bailly, C., Waring, M. J., Hochstrasser, R., and Leupin, W. (2001) Tris-benzimidazole derivatives: design, synthesis and DNA sequence recognition. *Bioorg. Med. Chem. 9*, 2905−2919.
29. Clark, G. R., Gray, E. J., Li, Y.-H., Leupin, W., and Neidle, S. (1996). Isohelicity and phasing in drug−DNA sequence recognition: crystal structure of a tris (benzimidazole)−oligonucleotide complex. *Biochemistry 35*, 13745−13752.
30. Aymami, J., Nunn, C. M., and Neidle, S. (1999) DNA minor groove recognition of a non-self-complementary AT-rich sequence by a tris-benzimidazole ligand. *Nucleic Acids Res. 27*, 2691−2698.
31. Satz, A. L., and Bruice, T. C. (2001) Recognition of nine base pairs in the minor groove of DNA by a tripyrrole peptide−Hoechst conjugate. *J. Am. Chem. Soc. 123*, 2469−2477.
32. Satz, A. L., and Bruice, T. C. (2002) Recognition of nine base pair sequences in the minor groove of DNA at subpicomolar concentrations by a novel microgonotropen. *Bioorg. Med. Chem. 10*, 241−252.
33. Neidle, S., Mann, J., Rayner, E. L., Baron, A., Opoku-Boachen, Y., Simpson, I. J., Smith, N. J., Fox, K. R., Hartley, J. A., and Kelland, L. R. (1999) Symmetric bis-benzimidazoles, a new class of sequence-selective DNA-binding molecules. *J. Chem. Soc., Chem. Commun.* 929−930.
34. Mann, J., Baron, A., Opoku-Boahen, Y., Johansson, E., Parkinson, G., Kelland, L. R., and Neidle, S. (2001) A new class of symmetric bis-benzimidazole-based DNA minor groove-binding agents showing antitumor activity. *J. Med. Chem. 44*, 138−144.
35. INSIGHT II Modelling Environment, Molecular Simulations Inc., San Diego, 1999.
36. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Cardwell, J. W., and Kollman, P. A. (1995) A second-generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc. 117*, 5179−5197.
37. HYPERCHEM, v. 7.03, Hypercube Inc., Gainesville, FL, 2002.
38. Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., and Stewart. J. J. P. (1985) AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc. 107*, 3902−3909.
39. Sun, X.-W., Mann, J., and Neidle, S. (2002) Synthesis of a novel dimeric bis-benzimidazole with site-selective DNA-binding properties. *Tetrahedron Lett. 43*, 7239−7241.
40. Drew, H. R., and Travers, A. A. (1984) DNA structural variations in the *E. coli tyr* T promoter. *Cell 37*, 491−502.
41. Lavesa, M., and Fox, K. R. (2001) Preferred binding sites for [N-MeCYs(3), N-MeCys(7)] TANDEM determined using a universal footprinting substrate. *Anal. Biochem. 293*, 246−250.
42. Bailly, C., and Waring, M. J. (1995) Comparison of different footprinting methodologies for detecting binding sites for a small ligand on DNA. *J. Biomol. Struct. Dyn. 12*, 869−898.
43. Ji, Y. H., Bur, D., Hasler, W., Bailly, C., Waring, M. J., Hochstrasser, R., and Leupin, W. (2001) Tris-benzimidazole derivatives: Design, synthesis and DNA sequence recognition. *Bioorg. Med. Chem. 9*, 2905−2919.
44. Portugal, J., and Waring, M. J. (1987) Assignment of DNA binding sites for 4′,6-diamidine-2-phenylindole and bisbenzimide (Hoechst 33258). A comparative footprinting study. *Biochim. Biophys. Acta 949*, 158−168.
45. Nguyen, B., Tardy, C., Bailly, C., Colson, P., Houssier, C., Kumar, A., Boykin, D. W., and Wilson, W. D. (2002) Influence of compound structure on affinity, sequence selectivity and mode of binding to DNA for unfused aromatic dications related to furamidine. *Biopolymers 63*, 281−297.
46. Quintana, J. R., Grzeskowiak, K., Yanagi, K., and Dickerson, R. E. (1992) Structure of a B-DNA decamer with a central T−A step: C-G-A-T-T-A-A-T-C-G. *J. Mol. Biol. 225*, 379−385.
47. Liepinsh, E., Leupin, W., and Otting, G. (1994) Hydration of DNA in aqueous solution: NMR evidence for a kinetic destabilization of the minor grove hydration of d-(TTAA)$_2$ versus d-(AATT)$_2$ segments. *Nucleic Acids Res. 22*, 2249−2254.
48. Lansiaux, A., Dassonneville, L., Facompré, M., Kumar, A., Stephens, C. E., Bajic, M., Tanious, F., Wilson, W. D., Boykin, D. W., and Bailly, C. (2002) Distribution of furamidine analogues in tumor cells: Influence of the number of positive charges. *J. Med. Chem. 45*, 1994−2002.
49. Abu-Daya, A., Brown, P. M., and Fox, K. (1995) DNA sequence preferences of several AT-selective minor groove binding ligands. *Nucleic Acids Res. 11*, 3385−3392.

BI026926W