Altered sequence specificity identified from a library of DNA-binding small molecules

Vladimir M Guelev, Matthew T Harting, R Scott Lokey* and Brent L Iverson

Background: The ability to target specific DNA sequences using small molecules has major implications for basic research and medicine. Previous studies revealed that a *bis*-intercalating molecule containing two 1,4,5,8-napthalenetetracarboxylic diimides separated by a lysine-*tris*-glycine linker binds to DNA cooperatively, in pairs, with a preference for G + C-rich sequences. Here we investigate the binding properties of a library of *bis*-intercalating molecules that have partially randomized peptide linkers.

Results: A library of *bis*-intercalating derivatives with varied peptide linkers was screened for sequence specificity using DNase I footprinting on a 231 base pair (bp) restriction fragment. The library mixtures produced footprints that were generally similar to the parent *bis*-intercalator, which bound within a 15 bp G + C-rich repeat above 125 nM. Nevertheless, subtle differences in cleavage enhancement bands followed by library deconvolution revealed a derivative with novel specificity. A lysine-*tris*-β-alanine derivative was found to bind preferentially within a 19 bp palindrome, without substantial loss of affinity.

Conclusions: Synthetically simple changes in the *bis*-intercalating compounds can produce derivatives with novel sequence specificity. The large size and symmetrical nature of the preferred binding sites suggest that cooperativity may be retained despite modified sequence specificity. Such findings, combined with structural data, could be used to develop versatile DNA ligands of modest molecular weight that target relatively long DNA sequences in a selective manner.

Address: Department of Chemistry and Biochemistry, University of Texas at Austin, Austin, TX 78712, USA.

*Present address: Institute of Chemistry and Cell Biology, Harvard Medical School, Boston, MA 02115, USA.

Correspondence: Brent L Iverson
E-mail: biverson@utxvms.cc.utexas.edu

Key words: combinatorial chemistry, cooperative effects, DNA recognition, screening methods, sequence specificity

Received: **19 July 1999**
Revisions requested: **29 July 1999**
Revisions received: **6 October 1999**
Accepted: **12 October 1999**

Published: **8 December 1999**

Chemistry & Biology 2000, **7**:1–8

1074-5521/00/$ – see front matter

Introduction

The ability to target specific genes *in vivo* and regulate their expression using small molecules has major implications for medicine, as well as for basic research [1–4]. In the past few years, impressive advances in the design of sequence-selective DNA-binding agents have been achieved [1,4–12]. Polyamides have been developed that exhibit a one-to-one correspondence code for the recognition of DNA base pairs via specific hydrogen-bonding patterns in the minor groove [3,5–7]. Other successful approaches include using triple-helix-forming oligonucleotides [8], modified zinc-finger proteins [9], peptide nucleic acids (PNAs) [10], and small molecules such as other minor-groove binders and intercalators [1,11,12].

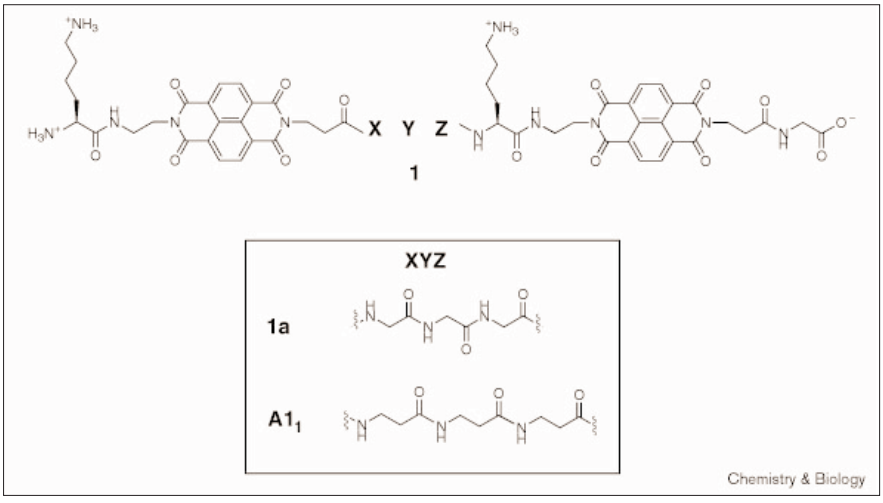
Still, much remains to be learned about the details of DNA local structure and dynamics that affect a so-called 'indirect DNA readout' [1,2,4,13–15]. In this respect, combinatorial synthesis and screening methods [16–22] can reveal unanticipated binding motifs, as well as lead to improvement and fine-tuning of the properties of existing ligands.

Derivatives of 1,4,5,8-naphthalenetetracarboxylic diimide and related intercalators have been studied extensively [23–28] because of their potential as chemotherapeutic agents [23,24], and because these molecules have been implicated in unusual DNA-binding modes such as threading [25,26,29] and major-groove binding [27]. We recently described the synthesis and DNA-binding properties of a series of polyintercalators [30], in which the intercalating diimide units are linked head-to-tail by amino acid linkers [26,29] (Figure 1).

These polyintercalators were synthesized on solid phase using N-9-fluorenylmethoxycarbonyl (FMOC)-based peptide chemistry [31] and were intended as a modular scaffold for the generation of combinatorial libraries of DNA ligands. Although most known simple intercalators have a rudimentary specificity for certain base pair steps (e.g. 5′-Py–Pu-3′) [2,32], this specificity can be overridden by noncovalent contacts with the DNA, usually via complex, rigid sidechains [2], or, we reasoned, by amino acid linkers held in place by two intercalating moieties. Analogous to zinc-finger proteins [9,20,21], a modular structure could, in principle, allow coverage of extended DNA sequences by combining the appropriate amino acid linkers, and intercalating units.

The reported polyintercalators, in which the diimide units were connected via lysyl-*tris*-glycyl linkers, were shown to

Figure 1



General structure of the *bis*-intercalators. Compound **A1**$_1$ was identified from library mixture **A1**. X, Y and Z are amino acids in the linker (see Figure 2). **1a** is a *tris*-glycine control.

bind DNA with full intercalation of all diimide units, while showing a preference for G + C-rich regions of DNA, not unlike the naphthalene diimide monomers themselves [25,27,28,30]. Importantly, like other *bis*-intercalating molecules such as echinomycin [33], our *bis*-diimide derivative bound cooperatively in pairs to the preferred G + C sequences. Relatively small molecules (1200 Da) were therefore found to recognize and bind relatively long DNA sequences (8 bp).

Here we investigate the effect of altered amino acid linkers on DNA binding in these compounds, with the initial construction and screening of a library of *bis*-intercalators.

Results and discussion

Synthesis

A 360-member (with some degeneracy) library was synthesized as 30 mixtures (Figure 2) of derivatives, in which structural parameters such as linker length, rigidity, orientation of the amide bonds, spacing of the amide bonds, and the presence of certain 'functional' sidechains were varied. All library members contained two 1,4,5,8-naphthalenetetracarboxylic diimide units and two positively charged lysine residues to normalize intercalative and electrostatic binding interactions, respectively. Amino acids **X** and **Z** in the linker were randomized with the structures shown (Figure 2), including the option to not add a residue in order to alter total linker length. Position **Y** contained one type of amino acid for every mixture in an attempt to detect any general trends of binding imposed by a specific sidechain. Note that the compounds described here, including the *tris*-glycine control (**1a**), differ from the original *bis*-intercalator by an inverted amino acid sequence [30]. The reasons for this change were twofold. We found that coupling of a lysine–diimide dipeptide was less problematic than that of the glycine–diimide adduct used in our previous study. In addition, using a lysine–diimide building block allowed us to incorporate the library invariant residues in one step.

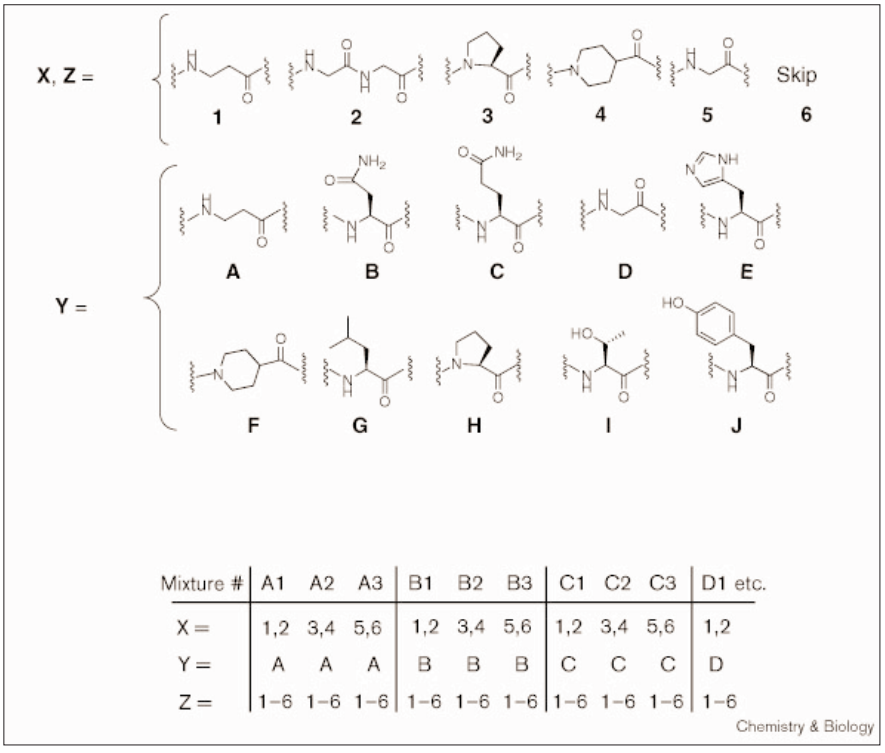
Several mixtures were analyzed using reverse-phase high-performance liquid chromatography (RP–HPLC) and electrospray ionization (ESI) mass spectrometry (Tables 1–3). The majority of the expected library members were identified in those mixtures (27 out of the expected 31 *m/z* ratios were matched, as both the mono- and di-protonated species). Three of the six expected compounds containing the 'skip codon' (mixture **D3**, Table 1) could not be identified. This was probably a result of an overestimate of the amount of glycine added to mixtures **3** and to complete couplings. In mixture **F2**, containing 'difficult' sequences of two or three secondary amino acids (proline or isonipecotic acid [Inp]; Table 2), proline couplings appeared problematic, whereas isonipecotic acid couplings did not pose a problem. All members of mixture **G1** were matched to the corresponding peaks (Table 3). In all three mixtures most of the remaining unidentified peaks were matched to monomeric diimide products, resulting from incomplete second coupling of the diimide moiety, suggesting that the coupling conditions for this amino acid need further optimization. In view of the relatively low binding affinity of the parent mono-intercalator (~10^{-4} M as compared with ~10^{-7} M for the *bis*-intercalator) [30] and the small number of library members in a mixture, we proceeded under the assumption that monomeric impurities would not interfere with the library screening.

Screening

Each mixture was assayed in the nanomolar to micromolar concentration range by DNase I footprinting [34] on a 231 bp *Eco*RI–*Nhe*I restriction fragment of plasmid pBR322. The overall appearance of the mixture footprints

Figure 2

Structure of the 360-member library (refer to structure **1** in Figure 1).



Mixture #	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1 etc.
X =	1,2	3,4	5,6	1,2	3,4	5,6	1,2	3,4	5,6	1,2
Y =	A	A	A	B	B	B	C	C	C	D
Z =	1–6	1–6	1–6	1–6	1–6	1–6	1–6	1–6	1–6	1–6

Chemistry & Biology

was similar to that of the *tris*-glycine control **1a** (Figure 3), indicating that, in most instances, changes in the linker had little, or even a deleterious effect, on DNA binding. We did not observe any general trends in specificity dictated by functional groups present in position **Y** in the linker.

As evident in Figure 3, footprinting of mixtures suffers inherently low signal-to-noise ratios. The absence of nonspecific competitor DNA (we are screening multiple sequences at once) and the nature of a footprinting signal in the assay (i.e. an actual lack of cleavage bands) make it difficult to detect subtle differences in protected regions between the mixtures. Fortunately, cleavage enhancement bands are often observed adjacent to a binding site in footprinting experiments [35]. Indeed, careful comparison revealed that differences between the mixtures

Table 1

Mass-spectral data for library mixture D3.

Mixture **D3** linker (X Y Z)	Calculated m/z $(M+H)^+$	Found	% Relative abundance	Calculated m/z $(M+2H)^{++}$	Found	% Relative abundance	Corresponding monomer*	Found	% Relative abundance
Gly–Gly–β-Ala	1243.5	1243.7	29	622.2	623.0	78	752.3	752.8	10
Gly–β-Ala	**1186.5**	**–**	**–**	**593.7**	**–**	**–**	695.3	695.6	15
Gly–Gly–Gly–Gly	1286.5	1286.7	42	643.7	644.5	100	795.3	794.7	35
Gly–Gly–Gly	1229.5	1229.7	27	615.2	615.9	54	738.3	738†	<5
Gly–Gly	1172.4	1172.7	27	586.7	587.4	41	681.2	681.7	17
Gly–Gly–Inp	1283.5	1283.7	29	642.3	642.9	69	792.3	792.7	10
Gly–Inp	**1226.5**	**–**	**–**	**613.7**	**–**	**–**	735.3	735.7	19
Gly–Gly–Pro	1269.5	1269.7	30	635.2	635.9	51	778.3	–	–
Gly–Pro	1212.5	1213.4	9	606.7	606.9	10	721.3	–	–
Gly	**1115.4**	**1115†**	**<5**	**558.2**	**–**	**–**	624.2	624.8	60

Library members that could not be identified in the mixture are bold. *Resulting from incomplete coupling of a diimide–lysine adduct and calculated from the corresponding library member. †Peak at background level.

Table 2

Mass-spectral data for library mixture F2.

Mixture **F2** linker (X Y Z)	Calculated m/z $(M + H)^+$	Found	% Relative abundance	Calculated m/z $(M + 2H)^{++}$	Found	% Relative abundance	Corresponding monomer*	Found	% Relative abundance
Inp–Inp–β-Ala	1351.6	1351.7	73	676.8	676.5	15	860.4	860†	< 10
Pro–Inp–β-Ala	1337.6	1337.7	90	669.8	668.7	20	846.4	847.1	26
Inp–Inp–Gly–Gly	1394.6	1394.6	76	698.3	698.2	50	903.4	902.8	55
Pro–Inp–Gly–Gly	1380.6	1380.0	59	691.3	690†	< 10	889.4	889.1	57
Inp–Inp–Gly	(1337.6)			(669.8)			(846.4)		
Pro–Inp–Gly	1323.5	1323.7	24	662.8	661.1	15	832.3	832.0	69
Inp–Inp–Inp	1391.6	1391.7	95	696.8	697.2	46	900.4	900.8	20
Pro–Inp–Inp	1377.6	1376.6	22	689.8	689.9	39	886.4	886.2	27
Inp–Inp–Pro	(1377.6)			(689.8)			(886.4)		
Pro–Inp–Pro	**1363.6**	**–**	**–**	**682.8**	**–**	**–**	**872.4**	873†	< 10
Inp–Inp	1280.5	1280.7	100	641.3	641.4	12	789.3	789.7	40
Pro–Inp	1266.5	1266.7	17	634.3	633.5	10	775.3	775†	< 10

Parentheses indicate isobaric compounds. Library members that could not be identified in the mixture are listed in bold. *Resulting from incomplete coupling of a diimide–lysine adduct and calculated from the corresponding library member. †Peak at background level.

exist (Figure 3). The most apparent deviation from the general appearance of the footprints was a unique enhancement at base position 37 (Figure 3, arrow), observed for mixture **A1**. Consequently, the members of library **A1**, containing β-alanine in the **Y** position, were synthesized and assayed individually.

Footprinting of the individual members of mixture **A1**, designated $\mathbf{A1}_1$–$\mathbf{A1}_{12}$ (Figure 4), showed that a single compound ($\mathbf{A1}_1$) containing three contiguous β-alanine residues, produced the enhancement at position 37 and protected the DNA from cleavage at a novel site, immediately adjacent to the enhancement.

Comparison of compounds 1a and $\mathbf{A1}_1$

Figure 5 shows a more detailed footprinting comparison of compounds **1a** and $\mathbf{A1}_1$ at 125–1000 nM. Compound **1a** recognizes several G + C-rich DNA regions as expected, but shows highest affinity for an 11 bp site that is part of the repeat 5′-*TGCCGG*TAC*TGCCGG*-3′ (better observed with the opposite, '–' strand labeled). In contrast, the new derivative $\mathbf{A1}_1$ has less affinity for the **1a** binding sites,

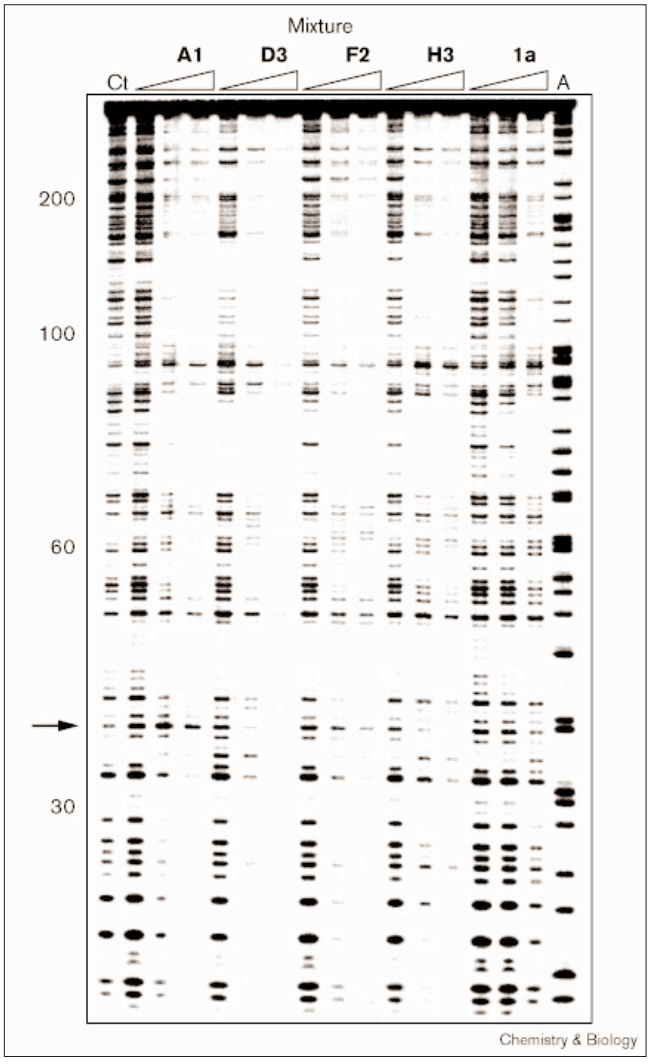
Table 3

Mass-spectral data for library mixture G1.

Mixture **G1** linker (X Y Z)	Calculated m/z $(M + H)^+$	Found	% Relative abundance	Calculated m/z $(M + 2H)^{++}$	Found	% Relative abundance	Corresponding monomer*	Found	% Relative abundance
β-Ala–Leu–β-Ala	1313.6	1313.7	29	657.3	657.1	12	822.4	–	–
Gly–Gly–Leu–β-Ala	1356.6	1356.6	100	678.8	679.2	53	865.4	–	–
β-Ala–Leu–Gly–Gly	(1356.6)			(678.8)			(865.4)	–	–
Gly–Gly–Leu–Gly–Gly	1399.6	1399.6	69	700.3	701.9	36	908.5	907.7	100
β-Ala–Leu–Gly	1299.6	1299.7	38	650.3	650†	< 10	808.4	809.0	30
Gly-Gly-Leu–Gly	1342.6	1342.7	68	671.8	671.2	20	851.4	851.8	48
β-Ala–Leu–Inp	1353.7	1353.6	33	677.3	676.7	52	862.5	859.7	28
Gly–Gly–Leu–Inp	1396.7	1396.6	67	698.8	701.0	35	905.5	905.8	34
β-Ala–Leu–Pro	1339.6	1339.7	35	670.3	671†	< 10	848.5	848.8	45
Gly–Gly–Leu–Pro	1382.7	1382.6	53	691.8	692.7	35	891.5	890.8	56
β-Ala–Leu	1242.6	1242.8	47	621.8	621.5	19	751.4	751.9	31
Gly–Gly–Leu	1285.6	1285.7	95	643.3	643†	< 10	794.4	794.6	41

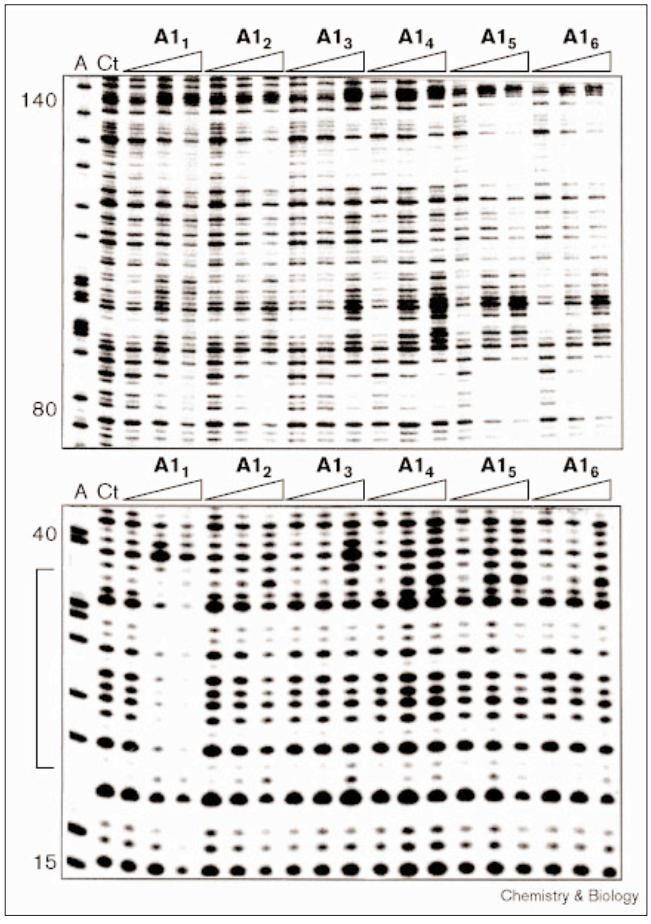
Parentheses indicate isobaric compounds. *Resulting from incomplete coupling of a diimide-lysine adduct and calculated from the corresponding library member. †Peak at background level.

Figure 3



DNase I footprinting analysis of selected mixtures on a 231 bp pBR322 restriction fragment (see the Supplementary material for all other library mixtures). Approximate compound concentrations (per chromophore, see text) are 0.8 μM, 3 μM and 10 μM. Compound **1a** concentrations are 0.5 μM, 1 μM and 2.5 μM. Lane Ct contains DNase I but no compound. Lane A contains A-specific chemical sequencing reaction [42]. Arrow indicates enhancement at bp 37, observed for mixture **A1**.

Figure 4

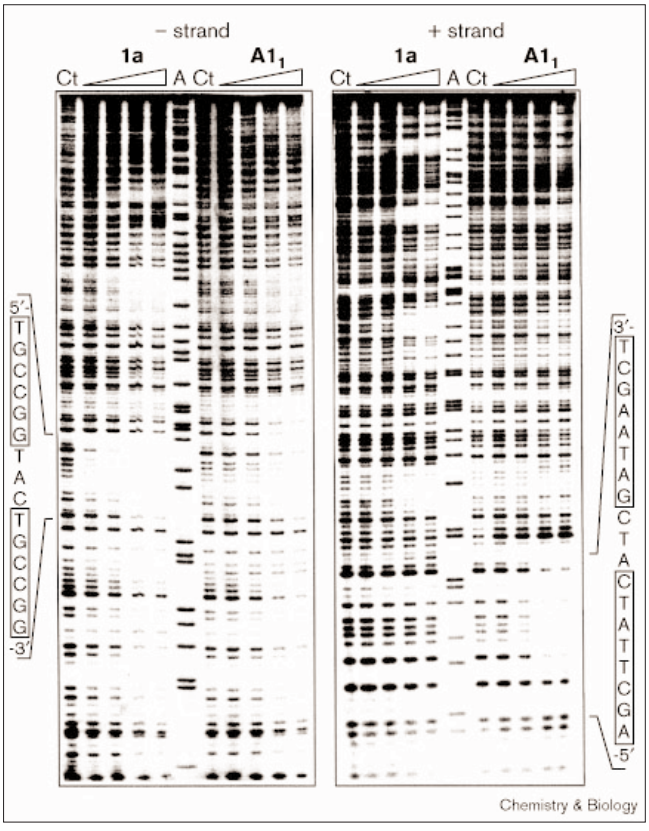


Footprinting analysis of the first six members of mixture **A1** on bp 15–42 and 75–145. Approximate concentrations (see text) are as follows: 0.25 μM, 1 μM and 2.5 μM. Lane Ct contains DNase I but no compound. Lane A contains A-specific chemical sequencing reaction [42]. The region protected from DNase I cleavage by $\mathbf{A1}_1$ is indicated by the bracket. Linker compositions are listed in the Materials and methods section; see the Supplementary material for complete footprinting data.

but binds most tightly to a 12 bp region, for which the *tris*-glycine control (**1a**) displays no apparent affinity. Interestingly, the preferred $\mathbf{A1}_1$ binding site is also symmetric, overlapping the palindrome (inverted repeat) 5′-*AGCTTATC*ATC*GATAAGCT*-3′. The data clearly indicate that compound $\mathbf{A1}_1$ shows novel specificity for a non-G + C-rich sequence, and a weakened affinity towards G + C regions, compared with compound **1a**. At concentrations above 1 μM the DNase I protection pattern of both compounds extends over the entire DNA fragment, as a result of nonspecific binding (data not shown). The footprinting data suggest sub-micromolar binding affinities for the preferred binding sites ($K_d \approx 1–3 \times 10^{-7}$ M) and a specificity range of ≤ tenfold (corresponding to a free energy difference ≤ 1.3 kcal/mol for specific versus nonspecific sites on the DNA).

Several observations suggest that the sequence-specificity difference between **1a** and $\mathbf{A1}_1$ may originate in the orientation and spacing of the amide bonds in the linkers, namely, α- versus β-peptide, and/or differences in conformational flexibility. Compound $\mathbf{A1}_1$ differs from all other library members by having a β-peptide linker and, as such, is probably the most conformationally flexible derivative. Limitations of the screening method notwithstanding, major specificity trends corresponding to the presence of 'functional' amino acid sidechains were not observed, possibly

Figure 5



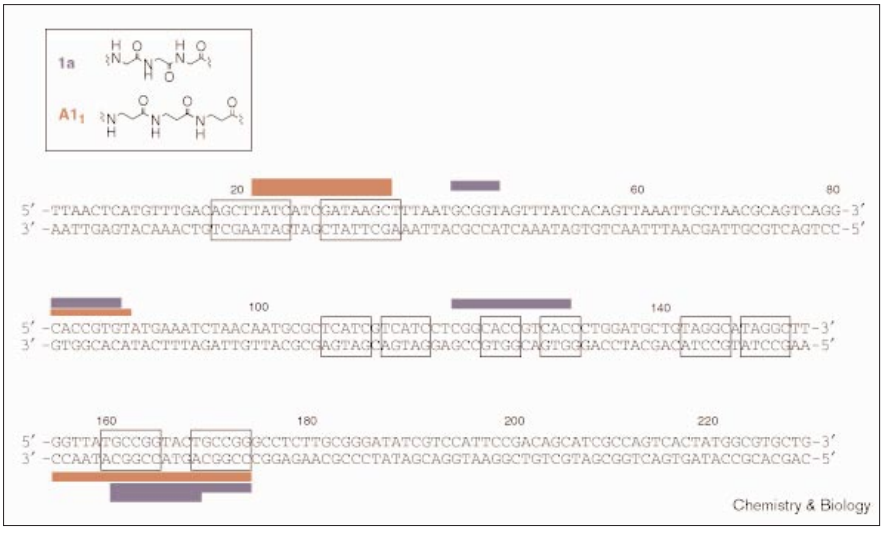
Footprinting of compounds **1a** and $\mathbf{A1_1}$ on 231 bp restriction fragment, 5′-end labeled at *Eco*RI ('+' strand) and *Nhe*I ('–' strand) overhangs. Compound concentrations are as follows: 0.13 μM, 0.25 μM, 0.5 μM and 1 μM. The symmetrical regions of interest are listed as read from the '+' strand, and the boxes indicate the repeated bases in those sites.

indicating that sidechains may not be as important for recognition as the amide backbone itself. Recently obtained two-dimensional nuclear magnetic resonance (NMR) data for compound **1a** (to be published elsewhere) suggest that various hydrogen-bonding contacts could be made between the DNA bases and the amino acid linker. In fact, the presence or absence of a single hydrogen bond would suffice to explain the observed extent of sequence specificity. It is therefore likely that a flexible *tris*-β-alanine linker with a different amide bond orientation allows new contacts with the DNA, directing a novel binding specificity, while retaining the original, albeit weakened, specificity of the *tris*-glycyl compound. Nevertheless, the possibility that the increased linker length of $\mathbf{A1_1}$ contributes to the new specificity cannot be ruled out. Another compound with the same linker length (in terms of number of atoms), containing four glycine residues, is present in mixture **D3**, however. This mixture did not produce the enhancement at base pair 37, observed for mixture **A1** and compound $\mathbf{A1_1}$ (Figure 3).

Interestingly, computer analysis revealed that the above-mentioned binding sites are the most significant direct or inverted repeats in the entire 231 bp fragment (Figure 6). Given that the original *bis*-intercalator bound cooperatively to DNA [30], the large size and symmetrical nature of the preferred binding sites suggest a binding model for $\mathbf{A1_1}$, in which two ligand molecules cooperatively recognize pairs of their preferred sequences.

Although the remaining mixtures and single **A1** members produced generally similar footprinting patterns, other subtle enhancement bands exist. For example, mixture **H3** exhibits a reproducible enhancement around base pair position 90 (Figure 3). The remaining single members of

Figure 6



Sequence of the 231 bp *Eco*RI–*Nhe*I restriction fragment. The bars show the approximate regions protected from DNase I digestion by compounds **1a** (blue) and $\mathbf{A1_1}$ (red). These were estimated from footprinting experiments with labeled '–' (bottom) and '+' (top) strand [34] (Figure 5). The boxed regions indicate all direct or inverted repeats of four or more bases, separated by no more than five bases. Note that the two largest such sites coincide with the two most preferred binding sites for compounds **1a** and $\mathbf{A1_1}$ (thick bars).

A1 also differ somewhat in their DNase I protection patterns (Figure 4).

The lack of more striking deviations in the apparent sequence specificity for other compounds and mixtures does not necessarily rule out the presence of new sequence specificities. Aside from limitations in the screening method, it is likely that the 231 bp fragment used for our studies does not contain other optimal binding sites, such as the repeats and palindromes that appear important for binding. We are currently re-screening the libraries against DNA sequences designed to contain varied repeats and palindromes in order to probe the scope and limitations of the findings reported here. Efforts are also underway to determine the structural basis of the different sequence specificities observed with **1a** and $\mathbf{A1_1}$.

Significance

The results reported here show that our modular poly-intercalator system can produce new derivatives with unique DNA-binding properties, by simply varying the amino acid linker connecting the intercalating moieties.

The binding affinities and specificity range of our *bis*-intercalators are rather modest, when compared with the polyamides described by Dervan and coworkers [5,6] that bind DNA with affinities comparable to DNA-binding proteins ($K_d \approx 10^{-11}$ M) and in which a single base-pair–ligand 'mismatch' results in a greater than tenfold decrease in the binding constant. They are also modest when compared with peptide nucleic acids (PNAs), in which base-pair recognition is more energetically favored than DNA–DNA base pairing [10]. The binding properties of the *bis*-diimides are comparable, however, to other *bis*-intercalators [1,2,4], such as echinomycin ($K_d \approx 10^{-5}$ M and ~30-fold specificity over other sites, as judged by footprinting experiments) [33], with the notable exception of a recently developed *bis*-daunomycin derivative. The latter binds DNA with $K_d \approx 10^{-11}$ M [12] and reveals the extent to which *bis*-intercalator binding can be fine tuned, using structural data.

Likewise, further combinatorial synthesis and screening of the *bis*-diimides, based on our results, can be used in conjunction with structural information from specific ligands bound to their preferred DNA for the design of derivatives with improved affinity and specificity, potentially expandable over a broad range of DNA sequences.

The diimide *bis*-intercalators may have a general propensity to bind cooperatively in pairs to repeats of their preferred binding sites. On the basis of this property, a cooperativity-based strategy for recognizing larger sequences [36], analogous to DNA-binding proteins [37], may be possible for a variety of small molecules.

Materials and methods

Library synthesis

Synthesis was performed manually [31], using a parallel synthesizer assembled in our laboratory, on a Tentagel™ resin, functionalized with FMOC-glycine. A naphthalene diimide-lysine adduct, prepared in solution [30], was used to incorporate the invariant residues in one step. The couplings of this amino acid were somewhat problematic, and best results were obtained with double couplings of the free acid in the presence of PyBOP/N-methylmorpholine [38]. The variable amino acids (**X**, **Y** and **Z**) were incorporated using the FMOC-protected pentafluorophenyl esters (purchased or prepared following published methods [39,40]) in the presence of HOBT. The first variable residue (**X**) was added in separate vessels, the resin was then combined and re-divided and all subsequent couplings were performed in parallel. The third variable position (**Z**) was added as mixtures of two amino acids, combining β-alanine and diglycine (mixtures 1), and proline and isonipecotic acid (mixtures 2), to minimize differences in reactivity. Mixtures 3 were coupled with less than one equivalent of glycine to allow for shorter linkages ('skip'). The mixtures were cleaved with TFA:phenol (95:5), extracted between water and diethyl ether and the aqueous layer was neutralized with triethylamine. Following lyophilization, the product was dissolved in water, loaded on a Waters Sep-Pak cartridge, eluted with a 70:30 mixture of water:acetonitrile (0.1% TFA each) and again lyophilized. The mixtures were characterized by RP–HPLC and ESI mass spectrometry (see the Supplementary material).

DNA preparation

Plasmid pBR322 was digested with *Eco*RI, dephosphorylated with CIAP, 5′-^{32}P-end labeled with [γ-^{32}P]-ATP and T4 kinase, digested with NheI and purified by native polyacrylamide gel electrophoresis (PAGE), following standard protocols [41].

Screening

Library mixture stock solutions were prepared to uniform UV absorbance ($\lambda = 386$ nm) and the final concentrations were calculated per total chromophore, assuming an average extinction coefficient $\varepsilon_{386} = 2.6 \times 10^4$ (as determined for compound **1a**). Based on the mass spectral data for mixture F2 (Table 2), assuming similar basicity among all compounds present in a mixture, we estimated that the compound present at lowest concentration in the mixture will be screened between 15 nM (at 0.8 μM total chromophore) and 200 nM (at 10 μM total chromophore). The DNA was dissolved in 20 mM sodium phosphate (pH 7.5) containing 2 mM $MgCl_2$ to a final estimated concentration of 0.5–2 nM and incubated with the library mixtures for at least 30 min prior to digestion with DNase I (0.5 μg/ml, 3 min). The reactions were stopped by freeze-drying and separated on a 8% denaturing polyacrylamide gel. The gels were exposed on phosphor screen and analyzed on ImageQuaNT 4.1 software from Molecular Dynamics.

Single compound synthesis and screening

Compounds $\mathbf{A1_1}$–$\mathbf{A1_{12}}$ were synthesized in parallel, analyzed by HPLC and high resolution FAB mass spectrometry (see the Supplementary material), and screened without further purification. Stock concentrations were calculated per total chromophore, assuming an average extinction coefficient $\varepsilon_{386} = 2.6 \times 10^4$ (as determined for compound **1a**) and adjusted for the presence of monomeric impurities, as determined by HPLC. Linker compositions for the compounds shown in Figure 4 as follows: $\mathbf{A1_1}$(β-Ala–β-Ala–β-Ala); $\mathbf{A1_2}$(Gly–Gly–β-Ala–β-Ala); $\mathbf{A1_3}$(β-Ala–β-Ala); $\mathbf{A1_4}$(Gly–Gly–β-Ala); $\mathbf{A1_5}$(β-Ala–β-Ala–Gly); $\mathbf{A1_6}$(Gly–Gly–β-Ala–Gly).

Compounds **1a** and $\mathbf{A1_1}$ were purified by FPLC on a reversed phase C_{18} column (PepRPC 15 m, Pharmacia) with a gradient of 10 mM TFA/$H_2$0 → 0.1% TFA/CH_3CN over 120 min with a flow rate of 0.75 ml/min. Compound **1a.** ^{1}H NMR (500 MHz, D_2O) δ 8.36–8.19 (m, 8 H), 4.37–4.28 (m, 4 H), 4.11–4.02 (m, 6 H), 3.95–3.85 (m, 4 H), 3.87 (s, 2 H), 3.82–3.74 (m, 2 H), 3.34–3.60 (m, 4 H), 2.88 (m, 4 H), 2.81 (t, 2 H, J = 7.6 Hz), 2.68 (t, 2 H, J = 7.1 Hz), 1.76–1.2 (m, 12 H); HRMS (FAB) 1229.4632 (1229.4652 calc'd for C_{58} $H_{65}N_{14}O_{17}$). Compound $\mathbf{A1_1}$. ^{1}H NMR (500 MHz, D_2O) δ 8.33–8.25 (m, 8 H), 4.28

(br t, 2 H), 4.21 (m, 4 H), 4.15 (br q, 2 H), 4.12–4.06 (m, 2 H), 3.81 (s, 2 H), 3.66–3.25 (m, 10 H), 2.90 (t, 2 H, J = 7.8 Hz), 2.83 (t, 2 H, J = 7.8 Hz), 2.67 (t, 2 H, J = 7.2 Hz), 2.57 (t, 2 H, J = 7.1 Hz), 2.42 (dt, 2 H, J_1 = 2.6, J_2 = 6.8 Hz), 2.36 (t, 2 H, J = 7.0 Hz), 2.33 (dt, 2 H, J_1 = 2.6, J_2 = 6.7 Hz)1.78–1.15 (m, 12 H); HRMS (FAB) 1271.5112 (1271.5122 calc'd. for $C_{61}H_{71}N_{14}O_{17}$).

Differential cleavage plots

Volume integration of the bands in the footprint and control lanes was performed in ImageQuant 4.1, and the plots were generated as $Ln(f_a/f_c)$, where f_a is the fractional cleavage of the DNA at the specified position in the presence of drug, and f_c– of the free DNA.

DNA sequence analysis

The 231 bp restriction fragment sequence was analyzed by a program written in Mathematica 3.0 (Wolfram Research, Champaign, Ill) to establish the frequency and size of repeat and palindromic (inverted repeat) sites.

Supplementary material

Supplementary material including mass spectral and HPLC data, full footprinting data, and differential cleavage plots is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

This work was supported by the Welch Foundation (F-1188), the American Cancer Society (RPG-97-085-01-CDD) and the National Institutes of Health (RO1 GM55646). We thank Dr Mehdi Moini for mass spectral analysis, Mark Cubberley, Chandra Miller and Dr Gang Chen for technical assistance, and Andy Zych for many lively discussions.

References

1. Chaires, J.B. (1998). Drug–DNA interactions. *Curr. Opin. Struct. Biol.* **8**, 314-320.
2. Hurley, L.H. & Chaires, J.B (1996). *Advances in DNA-Sequence Specific Agents*. Vol. 2, JAI Press, Greenwich, CT.
3. Dervan, P.B. (1986). Design of sequence-specific DNA-binding molecules. *Science* **232**, 464-471.
4. Nielsen, P.E. (1997). Design of sequence-specific DNA-binding ligands. *Chem. Eur. J.* **3**, 505-508.
5. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E., & Dervan, P.B. (1998). Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* **391**, 468-471.
6. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1998). Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *J. Am. Chem. Soc.* **120**, 3534-3535.
7. Dickinson, L.A., *et al.*, & Dervan, P.B. (1998). Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl Acad. Sci. USA* **95**, 12890-12895.
8. Gowers, D.M. & Fox, K.R. (1999). Towards mixed sequence recognition by triple helix formation. *Nucleic Acids Res.* **27**, 1569-1577.
9. Desjarlais, J.R. & Berg, J.M. (1992). Toward rules relating zinc-finger protein sequences and DNA binding site preferences. *Proc. Natl Acad. Sci. USA* **89**, 7345-7349.
10. Egholm, M., *et al.*, & Nielsen, P.E. (1993). PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566-568.
11. Bailly, C. & Chaires, J.B. (1998). Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconj. Chem.* **9**, 513-538.
12. Leng, F., Priebe, W. & Chaires, J.B. (1998). Ultratight DNA binding of a new *bis*-intercalating anthracycline antibiotic. *Biochemistry* **37**, 1743-1753.
13. Dickerson, R.E. (1992). DNA structure from A to Z. *Methods Enzymol.* **211**, 67-109.
14. Patel, D.J. (1997). Structural analysis of nucleic acid aptamers.*Curr. Opin. Chem. Biol.* **1**, 32-46.
15. Neidle, S. & Nunn, C.M. (1998). Crystal structures of nucleic acids and their drug complexes. *Nat. Prod. Reports* **15**,1-15.
16. Shipps, G.W. Jr., Pryor, K.E., Xian, J., Skyler, D.A., Davidson, E.H. & Rebek J., Jr. (1997). Synthesis and screening of small molecule libraries active in binding to DNA. *Proc. Natl Acad. Sci. USA* **94**, 11833-11838.
17. Klekota, B., Hammond, M.H. & Miller, B.L. (1997). Generation of novel DNA-binding compounds by selection and amplification from self-assembled combinatorial libraries. *Tetrahedron Lett.* **38**, 8639-8642.
18. Sandman, K.E., Fuhrmann, P. & Lippard, S.J. (1998). A mechanism-based, solution phase method for screening combinatorial mixtures of potential platinum anticancer drugs. (1998). *J. Biol. Inorg. Chem.* **3**, 74-80.
19. Lescrinier, T., *et al.*, & Herdewijn, P. (1998). DNA-binding ligands from peptide libraries containing unnatural amino acids. *Chem. Eur. J.* **4**, 425-433.
20. Choo, Y. & Klug, A. (1994). Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl Acad. Sci. USA* **91**, 11163-11167.
21. Desjarlais, J.R. & Berg, J.M. (1994). Length-encoded multiplex binding site determination: application to zinc finger proteins. *Proc. Natl Acad. Sci. USA* **91**, 11099-11103.
22. Tong, G. & Nielsen, J. (1996). A convergent solid-phase synthesis of actinomycin analogues-towards implementation of double-combinatorial chemistry. *Bioorg. Med. Chem.* **4**, 693-698.
23. Stevenson, K.A., Yen, S., Yang N., Boykin, D.W. & Wilson, W.D. (1984). A substituent constant analysis of the interaction of substituted naphthalene monoimides with DNA. *J. Med. Chem.* **27**, 1677-1682.
24. Braña, M.F., *et al.*, & Keilhauer, G. (1993). *Bis*-naphthalimides: a new class of antitumor agents. *Anti-Cancer Drug Des.* **8**, 257-268.
25. Tanious, F.A., Yen, S. & Wilson, W.D. (1991). Kinetic and equilibrium analysis of a threading intercalation mode: DNA sequence and ion effects. *Biochemistry* **30**, 1813-1819.
26. Lamberson, C. (1991). Topologically constrained DNA intercalators: synthesis and DNA binding studies. PhD thesis. pp 23-26. University of Illinois, USA.
27. Bailly, C., Braña, M. & Waring, M.J. (1996). Sequence-selective intercalation of antitumour *bis*-naphthalimides into DNA. Evidence for an approach via the major groove. *Eur. J. Biochem.* **240**, 195-208.
28. Liu, Z., Hecker, K.H., & Rill, R.L. (1996). Selective DNA binding of (N-alkylamine)-substituted naphthalene imides and diimides to G + C-rich DNA. *J. Biomol. Struct. Dyn.* **14**, 331-339.
29. Takenaka, S. Nishira, S., Tahara, K., Kondo, H. & Takagi, M. (1993). Synthesis and characterization of novel *tris*-intercalators having potentially two different DNA binding modes. *Supramol. Chem.* **2**, 41-46.
30. Lokey, R.S., Kwok, Y., Guelev, V., Pursell, C.J., Hurley, L.H. & Iverson, B.L. (1997). A new class of polyintercalating molecules. *J. Am. Chem. Soc.* **119**, 7202-7210.
31. Atherton, E. & Sheppard, R.C. (1989). *Solid Phase Peptide Synthesis – A Practical Approach*. IRL Press, Oxford.
32. Gago, F. (1998). Stacking interactions and intercalative DNA binding. *Methods* **14**, 277-292.
33. Bailly, C., Hamy, F., & Waring, M.J. (1996). Cooperativity in the binding of echinomycin to DNA fragments containing closely spaced CpG sites. *Biochemistry* **35**, 1150-1161.
34. Fox, K.R. (1997). DNase I footprinting. In *Drug-DNA Interaction protocols*. (Fox, K.R., ed.), pp. 1-22, Humana Press, Totowa, NJ.
35. Goodisman, J. & Dabrowiak, J.C. (1992). Structural changes and enhancements in DNase I footprinting experiments. *Biochemistry* **31**, 1058-1064.
36. Trauger, J.W., Baird, E.E. & Dervan, P.B. (1998). Cooperative hairpin dimers for recognition of DNA by pyrrole-imidazole polyamides. *Angew. Chem. Int. Ed. Engl.* **37**, 1421-1423.
37. Travers, A.A. (1993). *DNA-Protein Interactions*. Chapman & Hill, London.
38. Coste, J., Le-Nguyen, D. & Castro, B. (1990). PyBOP: a new peptide coupling reagent devoid of toxic by-product. *Tetrahedron Lett.* **31**, 205-208.
39. Lapatsanis, L., Milias, G., Froussios, K. & Kolovos, M. (1983). Synthesis of N-2,2-(trichlorethoxycarbonyl)-L-amino acids and N-(9-fluorenylmethoxycarbonyl)-L-amino acids involving succinimidoxy anion as a leaving group in amino acid protection. *Synthesis* 671-673.
40. Green, M. & Berman, J. (1990). Preparation of pentafluorophenyl esters of FMOC-protected amino acids with pentafluorophenyl trifluoroacetate. *Tetrahedron Lett.* **31**, 5851-5852.
41. Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989). *Molecular Cloning. A Laboratory Manual* (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
42. Iverson, B.L. & Dervan, P.B. (1993). Adenine-specific chemical sequencing reaction. *Methods Enzymol.* **218**, 222-227.