Extending the recognition site of designed minor groove binding molecules

A linked, imazole/pyrrole minor-groove ligand has been shown to bind sequence specificially to a 13 base-pair target sequence in a mixed 1:1/2:1 mode.

Sir—Small molecules that specifically bind to any unique DNA sequence would be useful tools in molecular biology and potentially in medical therapy. The natural product distamycin has provided a good starting point for design of

such molecules. Distamycin preferentially binds in the DNA minor groove at stretches of A,T base pairs[1] with the 1:1 ligand:DNA complexes extending over four base pairs (Fig. 1a)[2–4]. Until recently attempts to create a general design for small molecules that bind designated DNA sequences has had limited success[5]. Rules for controlling the minor groove binding specificity of polyamides, based on pairing of 2-methylimidazole (Im) and N-methylpyrrole (Py) amides, have now been developed[6–10]. Within this framework, an imidazole on one ligand complemented by a pyrrole on the second ligand recognizes a G·C base pair, while a pyrrole/imidazole combination targets a C·G base pair[6,7,11]. A pyrrole/pyrrole pair is degenerate for A·T and T·A base

pairs. Footprinting, affinity cleavage, NMR and X-ray studies have established the high structural similarity in such antiparallel, side-by-side 2:1 complexes at several different sites containing five and six base pairs (Fig. 1b)[6–18]. Here we show that 1:1 and 2:1 polyamide·DNA motifs can be combined to bind extended sequences of DNA containing both A,T and G,C base pairs.

It had previously been shown that polyamides with four to six pyrrole rings, $(Py)_{4-6}$, bound sequences containing six to eight contiguous A,T pairs[19,20]. A decrease in affinity occurs as the polyamide size increases beyond five rings, indicating that the shape of the N-methylpyrroleamide subunit does not match perfectly the B-DNA helix[20]. By incorporating simple hydrocarbon spacers[21,22], a 14 residue polyamide $(Py)_4$-(β-alanine)-$(Py)_4$-(β-alanine)-$(Py)_4$ was designed, and was shown to bind 16 contiguous A,T pairs[22]. These extensions of binding site size utilizing the Py subunits are restricted

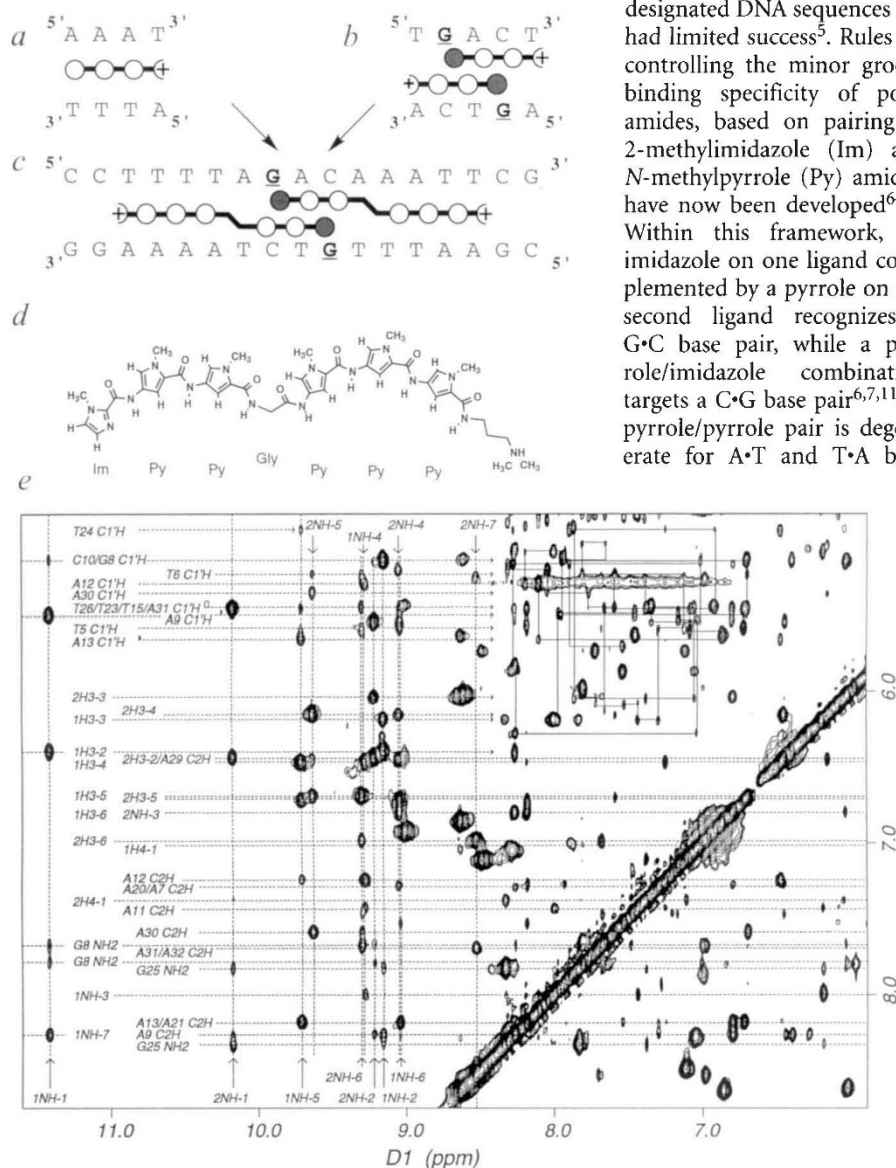**Fig. 1** Schematic representation of 1:1 and 2:1 polyamide–DNA binding motifs. a, 1:1 PyPyPy complex. b, Cooperative 2:1 ImPyPy complex. c, Cooperative 2:1 ImPyPy–Gly–PyPyPy complex on the designated 13 base-pair binding site. d, Structure of ImPyPy–Gly–PyPyPy. Open circles represent N-methylpyrrole amino acid (Py) subunits, while shaded-circles indicate N-methylimidazole amino acid (Im) subunits. e, Expanded region from a NOESY spectrum of the 2:1 complex (90% H_2O/10% D_2O, 25°C, τ_{mix}=100 ms). Sequential assignments are drawn in the DNA aromatic-H1' region with a solid line, dashed lines indicate intermolecular connectivities and are labelled with the protons involved.

correspondence

to pure A,T sequences. Targeting longer sequences containing both A,T and G,C base pairs requires use of the 2:1 motif. To explore binding at larger sites containing both A,T and G,C base pairs, we consider here combining 2:1 and 1:1 complexing motifs. We report that the polyamide ImPyPy–Gly–PyPyPy forms a stable, cooperative 2:1 complex with a 13 base pair site of sequence composition 5'-(A,T)$_5$G(A,T)C(A,T)$_5$-3' (Fig. 1 c,d).

Quantitative DNase I footprint titrations show that the polyamide ImPyPy-Gly-PyPyPy binds specifically at the sequences 5'-AAAAAGA-

CAAAAA-3' and 5'-ATATAGA-CATATA-3' with high affinities ($K \approx 10^8$ M^{-1} and 6×10^6 M^{-1} respectively) (J. Trauger & P.B.D., unpublished results). The region protected against DNase cleavage covers the entire 13 base-pair target sequence, consistent with the proposed binding motif.

Direct characterization of a complex was carried out by NMR. A duplex: 5'-CCTTTTAGA-CAAATTCG- 3'•5'-GCAATTTGTC-TAAAAGG-3' was synthesized to provide a central five base pair recognition sequence for a side-by-side 2:1 arrangement of the

ImPyPy, flanked by high affinity 1:1 A,T binding sites for the PyPyPy moieties. As ImPyPy–Gly–PyPyPy is added to a solution of this duplex a single complex, with two ligands per DNA, forms with high cooperativity. Using NOESY data it was possible to identify many intermolecular contacts, all of which occur in the minor groove. A region of such a NOESY spectrum is shown in Fig. 1e. The two ImPyPy ring systems bind the central 5'-AGACA-3'•5'-TGTCT-3' sequence in an antiparallel side-by-side arrangement identical to that of the unlinked 2:1 (ImPyPy)$_2$•DNA complex with the
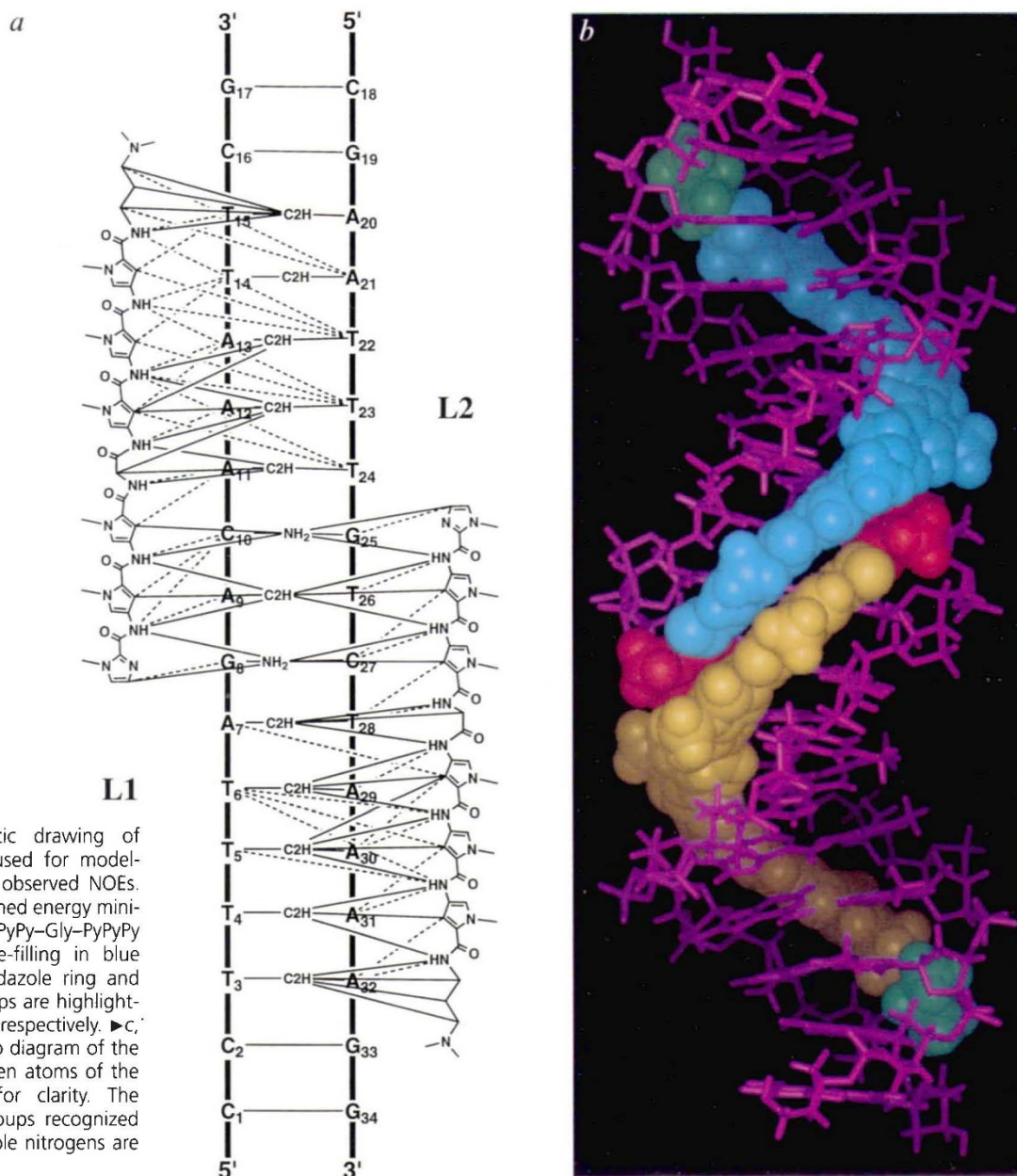

Fig. 2 a, Schematic drawing of observed restraints used for modelling, lines indicating observed NOEs. b, Model from restrained energy minimization with the ImPyPy–Gly–PyPyPy ligands shown space-filling in blue and yellow. The imidazole ring and the charged tail groups are highlighted in red and green, respectively. ►c, opposite page, Stereo diagram of the same model. Hydrogen atoms of the DNA are omitted for clarity. The guanosine amino groups recognized by the ligand imidazole nitrogens are highlighted.

5'-TGACT-3' site[6,7]. In this region each ligand is in close contact with just one strand of the DNA, while the PyPyPy segments of the ligand contact both strands. The positions of the intermolecular contacts clearly establish the 'slipped' positions of the two ligands with respect to one another, verifying the anticipated binding mode.

A total of 98 intermolecular ligand-DNA, two ligand–ligand and 20 intra-ligand semiquantitative restraints were derived from NOESY data; these are indicated schematically in Fig. 2a. These restraints were used to model the complex by restrained energy minimization (Fig. 2b). The local characteristics of this model are very similar to those seen in related 1:1 and 2:1 complexes. Each imidazole nitrogen forms a specific hydrogen bond with the adjacent guanosine amino proton (Fig. 2c). Hydrogen bonds are also seen from ligand amides to DNA acceptor groups. The AATT sequence is complexed by the PyPyPy part of one polyamide, while AAAA binds the other PyPyPy. As in 1:1 complexes of distamycin[4] the pyrrole and amide protons of each PyPyPy moiety show NOE cross peaks to C1' protons on both DNA strands, verifying that they occupy the centre of the narrower minor groove of these regions. The glycine linker allows a smooth transition between the 1:1 and 2:1 binding modes. As can be seen from the model, the last pyrrole ring of the ImPyPy segment and the first pyrrole of the PyPyPy segment are almost perpendicular to one another. This rotation of the ligand is accommodated within the glycine segment. The importance of linkers in matching the helicity of multi-ring ligands and the DNA has been discussed previously[5]. For the complex described here the nature of the linker seems to be especially important in allowing the transition from a 2:1 to a 1:1 binding mode, maintaining good van der Waals contacts along the length of each molecule. A more rigid, planar linker may not allow the relatively sharp turn between ligand segments.

The NOESY data suggest that no major distortion of the DNA duplex is required to accommodate two ligand molecules. The patterns and intensities of NOEs within the DNA do not change significantly on ligand binding. It should be noted that intra-DNA NOEs were not used in the modelling, and the structure of the DNA within the complex remains near the starting B-form. While we did not attempt to determine the full range of DNA conformations consistent with the NMR data, the model in Fig. 2 agrees well with all experimental observations. The model suggests that transitions between narrow and normal grooves occur around the linker region, but comprise only a single base pair.

These results show clearly that 2:1 and 1:1 binding modes can be combined to form a new 'slipped motif' which allows polyamides to recognize mixed A,T/G,C sites greater than one turn of the DNA helix, significantly longer than previously possible. All of the features used for sequence specific recognition in previous polyamide•DNA complexes apply as well in this new motif. The present structural characterization does not fully address the question of specificity of these ligands. Another binding mode, in which the full length of each ligand overlaps, can be envisioned. Indeed the ImPyPy–Gly–PyPyPy polyamide also binds the eight base-pair sequences 5'-TGTTAAACA-3' and 5'-TGATATACA-3' as a 'full side-by-side dimer' with affinity similar to the slipped motif (J. Trauger and P.B.D.; R. Pelaez and D.E.W., unpublished results). One future challenge for the field of design–synthesis will be the modification of these polyamide structures to control precisely the binding motif (slipped *versus* full side-by-side) and hence the target DNA sequence and binding site size.

Methods

ImPyPy–Gly–PyPyPy and oligonucleotides were prepared as described previously[14,16]. NMR experiments
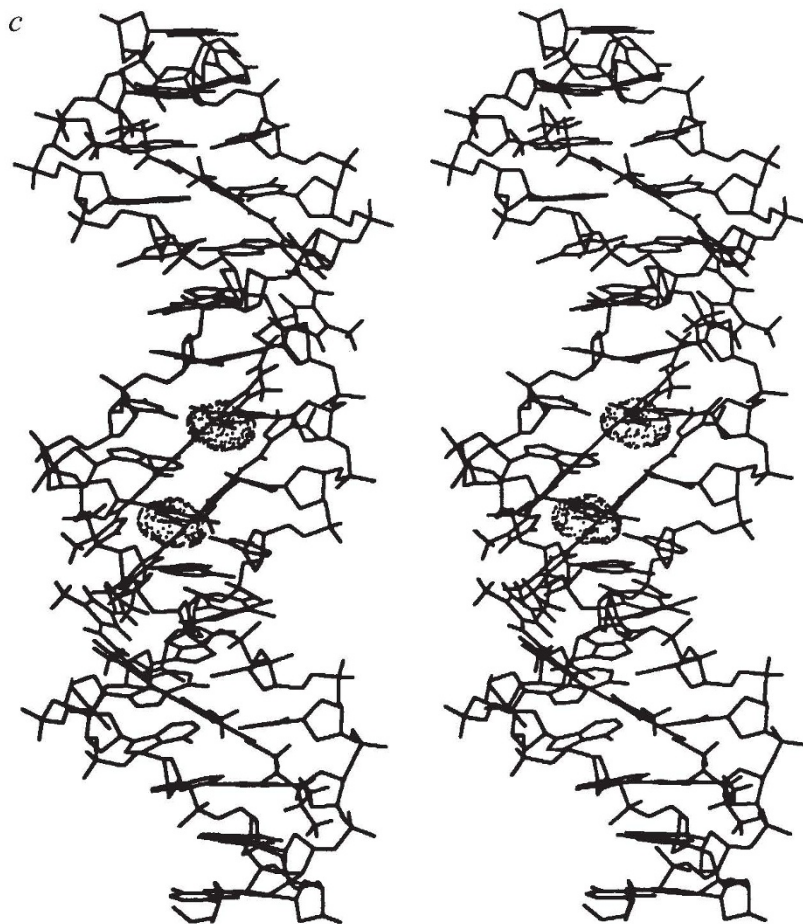
c

correspondence

were performed on a Bruker AMX-600 (Bruker Instruments, Billerica, MA). NOE volumes were derived from a 2D NOESY spectra in H_2O (2 mM duplex, pH 7.0, 10 mM phosphate, τ_{mix}=100 ms at 25 °C)[6,11].

The model was obtained using InsightII (Biosym, San Diego). The initial DNA model was standard B-form DNA. Coordinates for ImPyPy–Gly–PyPyPy were derived by linking two modified ligands from a previously modelled complex[13]. Ligands were energy minimized using Discover (AMBER forcefield) in the extended conformation with partial charges calculated using MOPAC (AM1). Ligands were roughly oriented in the binding site with the imidazole rings near the guanosine amino groups. Restrained energy minimizations were performed utilizing 98

intermolecular ligand-DNA and 20 intra-ligand restraints[13]. Restraints were derived from NOESY data by classifying cross peak volumes[6] semi-quantitatively into three categories: strong (1.8–2.5 Å), medium (2.5–3.7 Å) or weak (3.7–4.2 Å) referencing to the volumes of cytosine H5-H6 cross peaks. The H4-1 and H5-1 protons of both ligands did not show NOEs to the glycine protons of the opposite ligand, so glycine methylene protons point away from the imidazole rings. To assure a linker conformation consistent with this observation two intermolecular ligand-ligand restraints (4.2–10 Å) between the pseudo atoms for the glycine methylene protons and the H4-1 protons of the opposite ligand were introduced. Hydrogen bonds for standard Watson–Crick base pairing were included

as NOE restraints[13]. The model fulfils all restraints to within 0.1 Å.

Bernhard H. Geierstanger[1], Milan Mrksich[2], Peter B. Dervan[2] and David E. Wemmer[1]

[1]Department of Chemistry, University of California, Berkeley, California 94720-1460, USA
[2]Arnold and Mabel Beckman Laboratories of Chemical Synthesis, California Institute of Technology, Pasadena, California 91125, USA

Correspondence should be addressed to D.E.W.

1. Zimmer C. & Wähnert U. *Prog. Biophys. Mol. Biol.* **47**, 31–112 (1986).
2. Taylor J.S., Schultz, P.G. & Dervan P.B. *Tetrahedron* **40**, 457–465 (1984).
3. Coll M., Fredrick C.A., Wang A.H.J. & Rich A. *Proc. Natl. Acad. Sci. USA* **84**, 8385–8389 (1987).
4. Pelton J.G. & Wemmer D.E. *Biochemistry* **27**, 8088–8096 (1988).
5. Dervan P.B. *Science* **232**, 464–471 (1986).
6. Mrksich M. et al. *Proc. Natl. Acad. Sci. USA* **89**, 7586–7590 (1992).
7. Wade W.S., Mrksich M. & Dervan P.B. *J. Am. Chem. Soc.* **114**, 8783–8794 (1992).
8. Wade W.S., Mrksich M. & Dervan P.B. *Biochemistry* **32**, 11385–11389 (1993).
9. Mrksich M. & Dervan P.B. *J. Am. Chem. Soc.* **115**, 2572–2576 (1993).
10. Geierstanger B.H., Mrksich M., Dervan P.B. & Wemmer D.E. *Science* **266**, 646–650 (1994).
11. Geierstanger B.H., Dwyer T.J., Bathini Y., Lown J.W. & Wemmer D.E. *J. Am. Chem. Soc.* **115**, 4474–4482 (1993).
12. Mrksich M. & Dervan P.B. *J. Am. Chem. Soc.* **115**, 9892–9899 (1993).
13. Geierstanger B.H., Jacobsen J.P., Mrksich M., Dervan P.B. & Wemmer D.E. *Biochemistry* **33**, 3055–3062 (1994).
14. Mrksich M., Parks M.E. & Dervan P.B. *J. Am. Chem. Soc.* **116**, 7983–7988 (1994).
15. Mrksich M. & Dervan P.B. *J. Am. Chem. Soc.* **117**, 3325–3332 (1995).
16. Pelton J.G. & Wemmer D.E. *Proc. Natl. Acad. Sci. USA* **86**, 5723–5727 (1989).
17. Wemmer D.E. et al. in *Structural Biology: The State of the Art, Proceedings of the Eighth Conversation on Biomolecular Stereodynamics* (eds. Sarma RH, S.M.) 301–323 (Adenine Press, Guilderland, NY, 1994).
18. Chen X., Ramakrishnan B., Rao S.T. & Sundaralingam M. *Nature Struct. Biology* **1**, 169–175 (1994).
19. Schultz, P.G. & Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **80**, 6834–6837 (1983).
20. Youngquist, R.S. & Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **82**, 25–65 (1985).
21. Youngquist, R.S. & Dervan, P.B. *J. Am. Chem. Soc.* **107**, 55–28 (1985).
22. Youngquist, R.S. & Dervan, P.B. *J. Am. Chem. Soc.* **109**, 75–64 (1987).