Design of a Covalent Peptide Heterodimer for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA

Milan Mrksich and Peter B. Dervan*

Arnold and Mabel Beckman Laboratories of Chemical Synthesis California Institute of Technology Pasadena, California 91125

Received January 12, 1994

Downloaded via NORTHWESTERN UNIV on June 9, 2020 at 19:31:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

While 1:1 peptide–DNA models aided in the design of oligopeptides for recognition of long tracts of contiguous A,T base pairs of DNA, early efforts to design peptides for recognition of mixed A,T and G,C sequences met with limited success.[1] The synthetic peptides 2-PyN and 2-ImN, targeted to the 5′-TGTTA-3′ site as 1:1 complexes,[2] were found to bind the unanticipated sequence 5′-TGTCA-3′ as antiparallel side-by-side dimers in the minor groove.[3] These and other recent 2:1 peptide–DNA complexes offer a revised model for the design of ligands for sequence-specific recognition in the minor groove of DNA.[4–7] In addition to sequence-dependent minor groove width being a determinant of specificity, the 2:1 model allows specific contacts with *each* strand on the floor of the minor groove. Based on this new understanding, we report here a successful second generation molecule, a covalent peptide heterodimer, which specifically binds the designated sequence 5′-TGTTA-3′.

Although it has been shown that the two peptides distamycin (D) and 2-ImN bind the 5′-TGTTA-3′ site as an intermolecular antiparallel heterodimer, each peptide binds as well the sequences 5′-$(A,T)_5$-3′ and 5′-TGTCA-3′, respectively (Figure 1).[5] Recent results of peptide homodimers covalently linked through the central pyrrole rings suggested that a covalent peptide heterodimer should increase affinity for the 5′-TGTTA-3′ site.[6] The covalent peptide heterodimer 2-ImN-C_4-P3, wherein the peptides P3 and 2-ImN are connected through the nitrogens of the central pyrrole rings with a butyl linker, was synthesized in 12 steps from commercially available precursors (Figure 2).[8,9] A DNA fragment was constructed which contains a distamycin binding site (5′-TTTTT-3′), a 2-ImN homodimer binding site (5′-TGTCA-3′), and a 2-ImN/D heterodimer binding site (5′-TGTTA-3′), each separated by approximately 10 base pairs.[10] Quantitative DNase I footprint titration experiments on a ^{32}P end-labeled 447-base pair restriction fragment affords the binding affinities of D, 2-ImN, and 2-ImN-C_4-P3 for these three sites (Table 1).[11]

P3 binds the three sites in order of decreasing affinity: 5′-TTTTT-3′ > 5′-TGTTA-3′ > 5′-TGTCA-3′. 2-ImN binds these sites in the reverse order, 5′-TGTCA-3′ > 5′-TGTTA-3′ > 5′-TTTTT-3′. The covalent heterodimer 2-ImN-C_4-P3 specifically binds the 5′-TGTTA-3′ site, the new order being 5′-TGTTA-3′ > 5′-TGTCA-3′ > 5′-TTTTT-3′. The 3-fold lower affinity of 2-ImN-C_4-P3 for the 5′-TGTCA-3′ site is consistent with binding as an intramolecular dimer with a single hydrogen bond mismatch between one guanine amino group and the *N*-terminal pyrrole-carboxamide of the distamycin ligand (Figure 3b). Similarly, 2-ImN binds the 5′-TGTTA-3′ site with 3-fold lower affinity than the 5′-TGTCA-3′ site.[12] This correspondence suggests a *common free energy penalty of ~0.6 kcal/mol for a single*

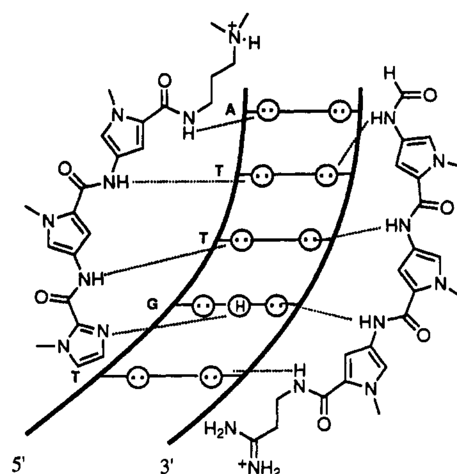

Figure 1. Heterodimeric binding model for the complex formed between 2-ImN and D with a 5′-TGTTA-3′ sequence.[5] Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of guanine. Putative hydrogen bonds are illustrated by dotted lines.

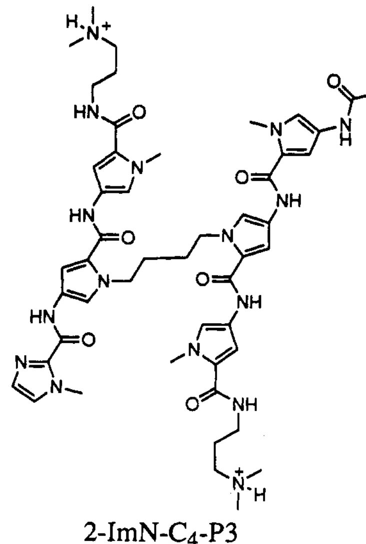

Figure 2. Covalent peptide heterodimer 2-ImN-C_4-P3 wherein the two peptides 2-ImN and P3 are connected through the nitrogens of the central pyrrole rings with a butyl linker.

(1) Dervan, P. B. *Science* **1986**, *232*, 464–471.
(2) Wade, W. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 1574–1575.
(3) (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783–8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586–7590. (c) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385–11389.
(4) (a) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723–5727. (b) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393–1399.
(5) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572–2576. (b) Geierstanger, B. H.; Jacobsen, J.-P.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *Biochemistry* **1994**, *33*, 3055–3062.
(6) (a) Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 9892–9899. (b) Dwyer, T. J.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, *115*, 9900–9906.
(7) (a) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911–5919. (b) Geierstanger, B. H.; Dwyer, T. J.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1993**, 115, 4474–4482.
(8) P3 is an analog of distamycin A wherein the *N*-terminal formamide and the *C*-terminal amidinium groups of the natural product are replaced with acetamide and dimethylammonium groups, respectively.
(9) ^1H NMR, ^{13}C NMR, IR and mass spectral data for 2-ImN-C_4-P3 and all intermediates are consistent with the assigned structures.

0002-7863/94/1516-3663$04.50/0 © 1994 American Chemical Society

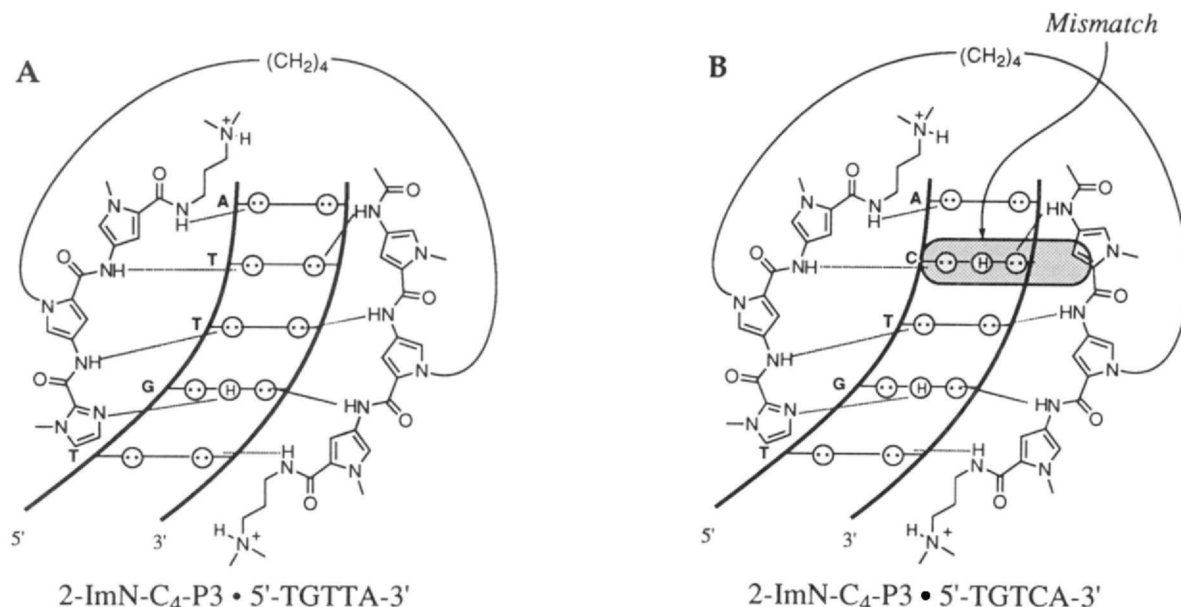

Figure 3. Model for the complex formed between the covalent peptide heterodimer 2-ImN-C_4-P3 with the 5'-TGTTA-3' site. (b) Model for the complex formed between 2-ImN-C_4-P3 with the 5'-TGTCA-3' site. The absence of a hydrogen bond acceptor on the peptides for recognition of the guanine amino group results in a single hydrogen bond mismatch in the complex.

hydrogen bond mismatch in the 2:1 complex. Finally, 2-ImN-C_4-P3 binds the "mismatch" 5'-TGTCA-3' site with greater affinity than does the matched intermolecular homodimer 2-ImN, suggesting that the favorable entropy gained upon linking the peptides overcomes the energetic penalty of the hydrogen bond mismatch.

In the past decade, the design of nonnatural minor groove binding peptides with high specificity for mixed DNA sequences has been driven by some combination of imperfect models,[13] powerful assays, and serendipity. The present example combines elements of design which may be general: placing complementary hydrogen bond donors and acceptors on side-by-side peptides to those of each DNA strand on the floor of the minor groove and improving specific affinity by covalently linking the peptides. Recognition of a GC base pair is afforded by a side-by-side pairing of imidazole and pyrrolecarboxamide.[3] Similarly, recognition of A,T base pairs is afforded by pairing of two pyrrolecarboxamides.[3,4] Perhaps this will provide an underpinning for the design of peptide analogs for sequence-specific recognition in the minor groove of *designated* DNA sites containing both A,T and G,C base pairs.

Table 1. Apparent First-Order Binding Constants (M^{-1})[a,b]

peptide	binding site		
	5'-TTTTT-3'	5'-TGTTA-3'	5'-TGTCA-3'
P3	1.6×10^6 (0.2)	1.7×10^5 (0.1)	$<1 \times 10^5$
2-ImN	$<5 \times 10^4$	4.7×10^4 (2.6)	1.5×10^5 (0.1)
2-ImN-C_4-P3	1.7×10^5 (0.5)	1.1×10^6 (0.3)	3.7×10^5 (0.6)

[a] Values reported are the mean values measured from four footprint titration experiments. Numbers in parentheses indicate the standard deviation for each data set. [b] The assays were performed at 22 °C, pH 7.0, in the presence of 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$.

Acknowledgment. We are grateful to the National Institutes of Health (GM-27681) for research support and the Ralph M. Parsons Foundation for a Graduate Fellowship to M.M.

(10) The plasmid pMM5 was constructed by ligation of an insert, 5'-TCGACATGACATTCGTCCACATTGTTAGACCACGATCG-TTTTTCGCATG-3' and 5'-CGAAAAACGATCGTGGTCTAACAAT-GTGGACGAATGTCATG-3', into pUC19 previously cleaved with *Sal*I and *Sph*I. The plasmid was digested with *Eco*RI, labeled at the 3'-end, and digested with *Afl*III. The 447-base pair restriction fragment was isolated by nondenaturing gel electrophoresis and used in all experiements described here.

(11) The quantitative footprint titration experiments and data analysis were performed as previously described,[6a] with unlabeled carrier DNA absent from all reactions.

(12) For characterization of the $(2\text{-ImN})_2$·5'-TGTTA-3' complex by two-dimensional NMR, see ref 5b.

(13) The lack of generality of early 1:1 peptide–DNA models[14] for mixed sequences is likely due to the fact that there is wide variation in minor groove width depending on sequence.

(14) (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376–1380. (b) Goodsell, D.; Dickerson, R. E. *J. Med. Chem.* **1986**, *29*, 727–733. (c) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590–5595.