Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 2215–2235

BIOORGANIC & MEDICINAL CHEMISTRY

Molecular Recognition of DNA by Small Molecules

Peter B. Dervan*

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Accepted 17 July 2001

Chemists, like artists, are able to construct new three-dimensional objects, molecules and materials that exist only in the mind of a person. I became interested in creating novel molecular shapes with properties different from those found in nature shortly after arriving at Caltech in 1973. One cannot design without the brushes and paint of the craft. Indeed, modern organic chemists, standing on the shoulders of the pioneering achievements of Woodward, Corey, Merrifield, and others are able to apply the power of synthetic chemistry and the logic of incremental change to the field of structure–function. In early 1973, I was inspired by the work of Lehn and Cram in the field of host–guest chemistry where early studies were largely conducted in organic solvents (e.g., cation–crown complexation). I decided that a pivotal path forward would be to understand in a predictive mechanistic sense how to create ensembles of weak bonds between synthetic ligands and biological macromolecules in water, the solvent of life.

All living organisms on planet Earth from bacteria, yeast, flower, fruit fly, mouse to man store the genetic information in a common molecule, the DNA double helix (Fig. 1) In 1973, there were no crystal structures of double helical DNA nor of protein–DNA complexes. We now know that the chemical principles by which nature's proteins read out and control the genetic information are chemically complex. After all, nature had billions of years to solve this recognition problem and used selection over time from a vast library of protein surfaces. I asked the question whether organic chemists could rationally create 'four chemical keys' which would distinguish each of the Watson–Crick A•T, T•A, G•C and C•G base pairs and when linked together read any contiguous predetermined sequence of DNA. This structure–function program would be different from 'anti-sense,' wherein the Watson–Crick pairing rules uncovered in 1953 are used for 'sense–strand' recogni-

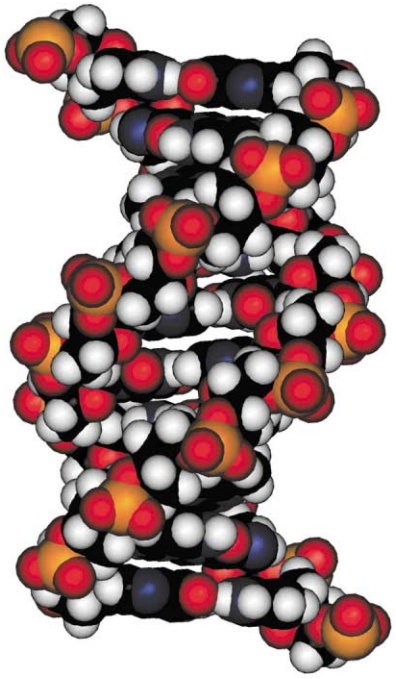


Figure 1. The DNA Helix.

*Corresponding author. Tel.: +1-626-395-6002; fax: +1-626-683-8753; e-mail: dervan@caltech.edu

0968-0896/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0968-0896(01)00262-0

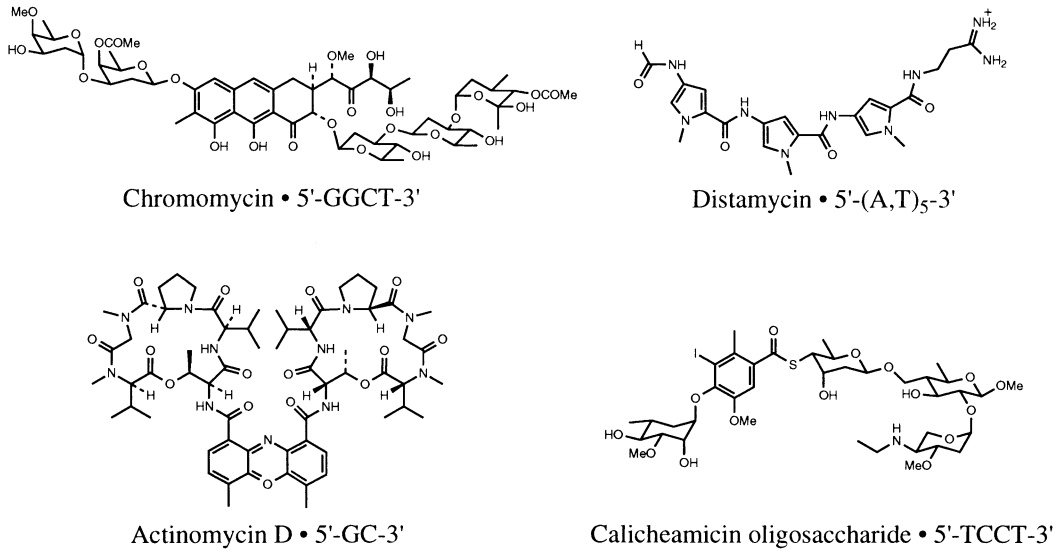


Figure 2. Natural products which bind DNA.

tion. Rather, this would be a small molecule design exercise to read the 'edges' of intact Watson–Crick pairs in either the major or minor groove of the DNA double helix. In one turn of the double helix (10 base pairs) all possible combinations of the four Watson–Crick base pairs afford 524,800 different sequences (Table 1). Rather than arriving at 524,800 different solutions to targeting each and every possible 10 base pair sequence, we would attempt to invent a universal chemical code with a common shape for recognizing any linear combination of Watson–Crick base pairs and assume there will be limitations for some set of sequences due to the sequence-dependent microstructure of DNA.

In keeping with the tradition of a personal essay that accompanies the Tetrahedron Prize, I will reflect on some of the events that led to our 25-year expedition to solve this latter approach. One has the benefit of looking back and understanding in hindsight which were pivotal decision points critical for the path forward. Indeed, there were five phases I can identify (1) formulating the questions, (2) methods development, (3) triple helix era, (4) pairing rules for minor groove recognition and (5) regulation of gene expression by synthetic ligands. In the next ten years I visualize an era of 'transcription therapy' wherein this basic research could underpin a new class of human medicines.

The Early Days at Caltech

I arrived at Caltech in Fall 1973 at the age of 28 with no research proposals and only 6 months of postdoctoral experience at Stanford. In the transition between completing my PhD work in physical organic chemistry with Jerry Berson at Yale University and my arrival at Stanford in January 1973, I was offered a job at Caltech. Gene Van Tamelen was supportive that despite my recent arrival I should take the unanticipated opportunity. I headed down to Pasadena as Assistant Professor and joined the Caltech organic group consisting of Jack Roberts, Bob Ireland and Bob Bergman. Dave Evans of UCLA would arrive 1 year later. I am forever grateful to these colleagues for their friendship and the time they invested in me. The breadth of my early publication record demonstrates a certain casting about for a new research direction. The project I found most satisfying during that time was the experimental demonstration by Tadao Uyehara and Don Santilli of 'biradical as a common intermediate' which resolved a long-standing debate between theory and experiment.[1–4] Eighteen years later my colleague Ahmed Zewail would directly probe the same intermediate at femtosecond resolution and validate our conclusion.[5]

During my first years at Caltech, I experienced how profoundly classroom teaching would influence my research. I taught Advanced Organic Chemistry, a

Table 1. Relation between binding site size and number of distinguishable sequences

Site size (*n*)	Unique sites (*N*)
2	10
3	32
4	136
5	512
6	2080
7	8192
8	32,896
9	131,072
10	524,800
11	2,097,153
12	8,390,656
13	33,554,432
14	134,225,920
15	536,870,912

reaction mechanisms course for the graduate students. Although my PhD was in physical organic chemistry, I never had the time to read in detail the original seminal papers of Winstein, Doering, Roberts, Olah and other pioneers of the field. Upon reading those key papers published mostly in the 1950′s, I realized that these researchers set the stage for several decades of new inquiry in the field of reactive intermediates. My research should not be closing problems but rather opening new areas. I decided to choose a research problem that seemed hopeless in terms of current understanding, but would be suitable for 30 years of scientific inquiry. Rather than study the covalent bond (though still imperfectly understood), I would study weak noncovalent bonds in the most challenging of solvents, water. Therefore, the 'synthetic objective' would be the three dimensional assembly of multiple specific noncovalent bonds in aqueous media. Biological molecules, such as proteins or nucleic acids, would be my target and small molecule synthesis combined with

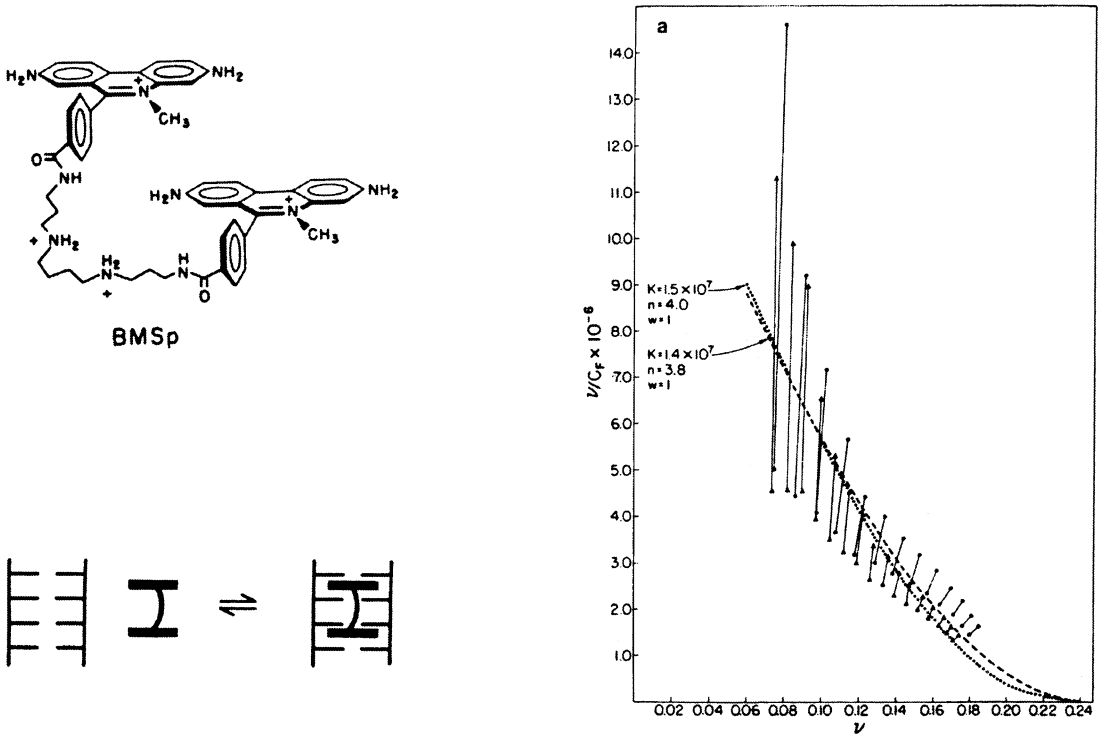


Figure 3. Bis(methidium) spermine, a synthetic bisintercalator, binds with high affinity to DNA.

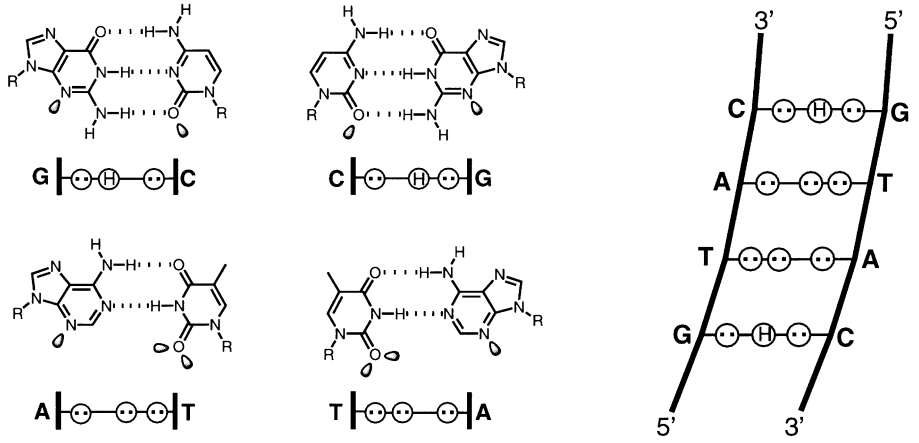


Figure 4. Molecular recognition of the minor groove of DNA. Minor groove hydrogen bonding patterns of Watson–Crick base pairs. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine.

physical characterization and the methods of biology would provide the experimental foundation. I would move from the gas phase world of hydrocarbon rearrangements to the aqueous world of nucleic acids and molecular recognition. I did not have the benefit of a biology or biochemistry course in college or graduate school. I was entering the interface of chemistry and biology unimpeded by preconceived notions and not wedded to any particular techniques. Sequencing of DNA was just being invented and the dream of the human genome effort was not yet articulated.

Methods Development

Thirty years ago, the field of small molecule–DNA recognition had few organic chemists and was led largely by biophysical chemists, such as Donald Crothers and Michael Waring, who utilized spectroscopic methods for characterizing 'drug-DNA' interactions, in particular Lerman's intercalation model.[6–10] (Fig. 2) I decided to apply the power of synthetic chemistry to create novel DNA binding molecules for DNA recognition. The title of this essay is taken from one of our first

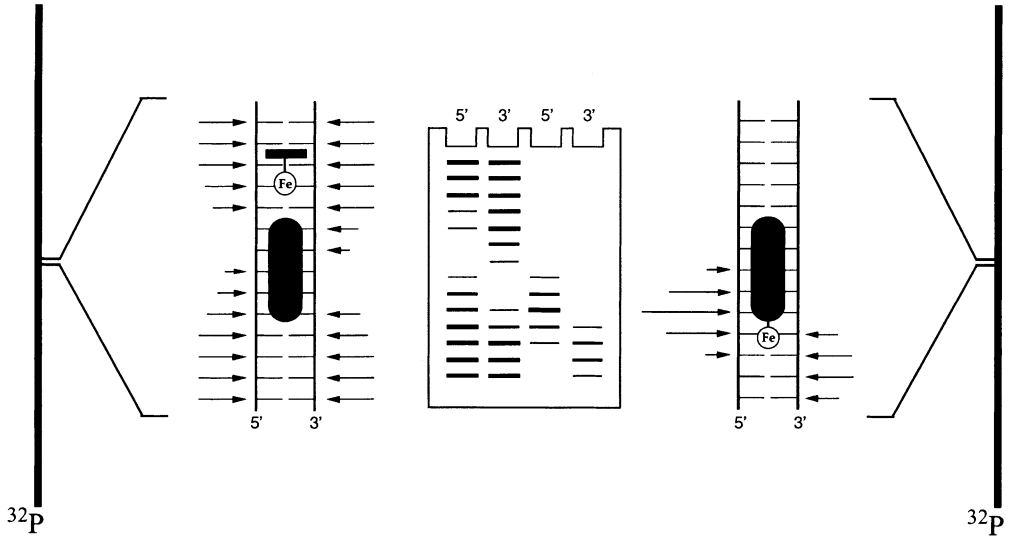


Figure 5. Schematic of (left) MPE·Fe(II) footprinting and (right) affinity cleaving techniques. (center) DNA cleavage products obtained on a denaturing gel with DNA end-labeled either or the 5′ or 3′ strands are shown.

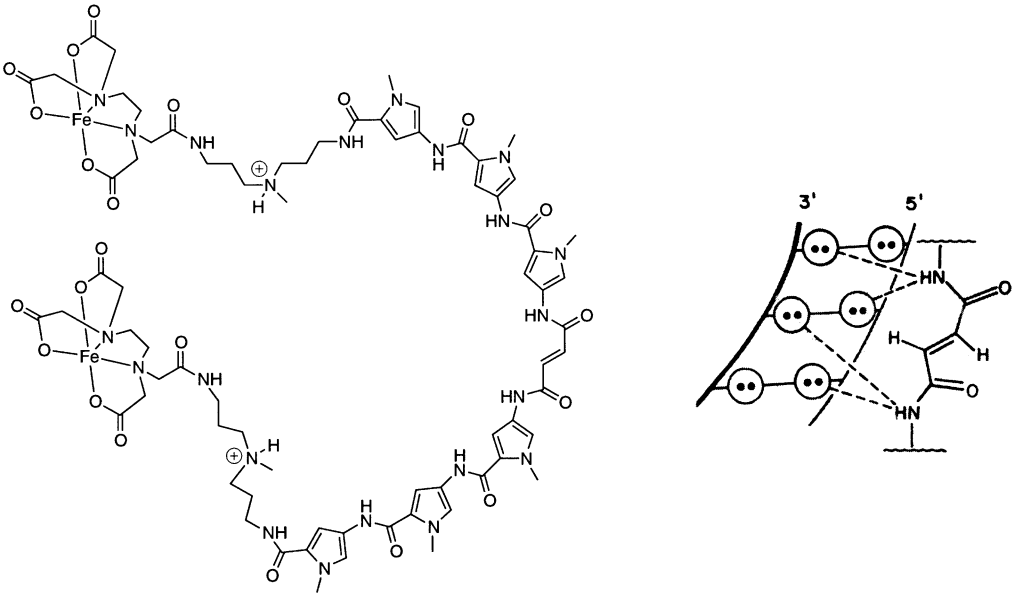
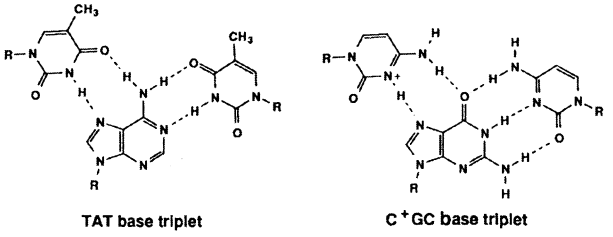


Figure 6. Distamycin dimer binds 9 base pairs of A,T rich DNA.

papers on molecular recognition of DNA published in 1978 which establishes a starting point for our thinking, the idea of connecting 'modules', in this case intercalators, that would bind cooperatively with high affinity to DNA with an increase of the binding site size[11,12] (Fig. 3). At the time we had no idea how to control sequence specificity. An examination of DNA binding natural products actinomycin, daunomycin, echinomycin and chromomycin revealed that they are structurally complex (Fig. 2). The simplest structure was the crescent-shaped distamycin,[10] comprising three *N*-methylpyrrole amino acids (Py) and known to bind in the minor groove of DNA at A,T rich sequences. The field of molecular recognition was sufficiently immature that design of sequence-specific DNA binding molecules for predetermined sequences would be at best an approximate exercise. We decided to focus on the minor groove of DNA (Fig. 4).

Together with my first bioorganic graduate students Mick Becker, Mark Mitchell, Tadhg Begley and Rick Ikeda, we struggled with labor intensive Scatchard plots and then we realized the path forward was to allow experiment to guide design by screening libraries of potential DNA binding sites using the powerful separation methods of gel electrophoresis. Inspired by the natural product bleomycin•Fe which cleaves DNA in a metal-mediated dioxygen-dependent reaction, Bob Hertzberg synthesized methidium with the iron chelator EDTA attached and demonstrated that MPE•Fe oxidatively cuts DNA very efficiently, but in a highly sequence neutral manner.[13] This was our initial effort to demonstrate the coupling of two different functions into one molecule separated as modular domains. Mike Van Dyke then utilized MPE•Fe as a reagent to introduce the footprinting method for small molecule–DNA recognition which allowed us to screen 150 base pairs of DNA in a single experiment (and, hence, a library of 150 potential binding sites).[14–16] (Fig. 5) Subsequently, Peter Schultz and John Taylor teamed up to attach Fe•EDTA to distamycin, which afforded the 'positive print' to a footprint 'negative' on sequencing gels.[17,18] (Fig. 5). The affinity cleavage provided a second method to scan large libraries of DNA sites.[19] As an unanticipated bonus, the cleaving patterns revealed orientation and groove location of the ligand–DNA complex.[17–19] During the next few years, Peter Schultz, John Griffin, Scott Youngquist and Jim Sluka synthesized several minor groove binding ligands based on the notion of combining 'modules' from different minor groove binding natural products such as netropsin and actinomycin.[19-23] (Fig. 6) We were able to demonstrate that connecting three modules of *N*-methylpyrrole-tetramers $(Py)_4$ with β alanine (β) created a minor groove ligand which bound an A, T tract 16 base pairs in size.[23]



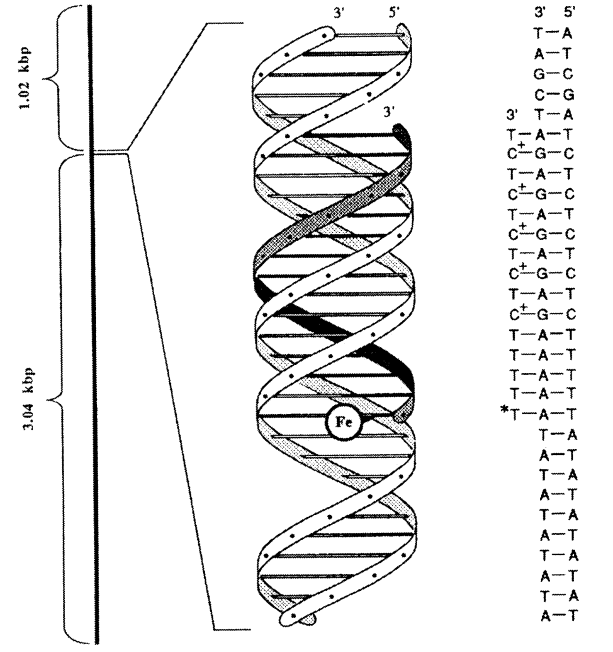


Figure 7. (Top) The pyrimidine triple helix motif. Isomorphous base triplets of TAT and CGC. The additional pyrimidine strand is bound by Hoogsteen hydrogen bonds in the major groove to the complementary purine strand in the Watson–Crick duplex.

The Triple Helix

Triple helical RNA and DNA had been postulated in the literature as early as 1957 by Felsenfeld, Davies and Rich, but never used for specific DNA targeting.[24] A breakthrough occurred in 1987 when coworker Heinz Moser, demonstrated that oligodeoxyribonucleotides (15 mer) could form stable specific triple helical complexes in the major groove of DNA under pH, temperature and salt conditions not too different from a living cell.[25] (Fig. 7). The discovery was made simultaneously by Claude Hélène's group in Paris,[26] and together the Pasadena and Paris laboratories put forward a huge effort to elucidate the full scope and limitations of the approach. The chemical significance was the cooperative assembly of 30 sequence specific hydrogen bonds in the major groove of the double helix.[25–27] In addition, the method of recognition in the major groove of DNA was modular. The code for the 'pyrimidine motif' was T binds AT and C binds GC (Hoogsteen TAT and C+GC triplets).[25–27] Peter Beal then demonstrated the orientation of the 'purine motif' wherein G binds GC and A binds AT.[28] (Fig. 8). However, despite heroic synthetic efforts by very able coworkers Linda Griffin, Dave Horne, Laura Kiessling, Eric Kool, Uli Stilz, Thomas Lehmann, Carol Wada and Jon Parquette, the triple helix approach remained limited largely to purine sequences.[29–33] Jim Maher in collaboration with

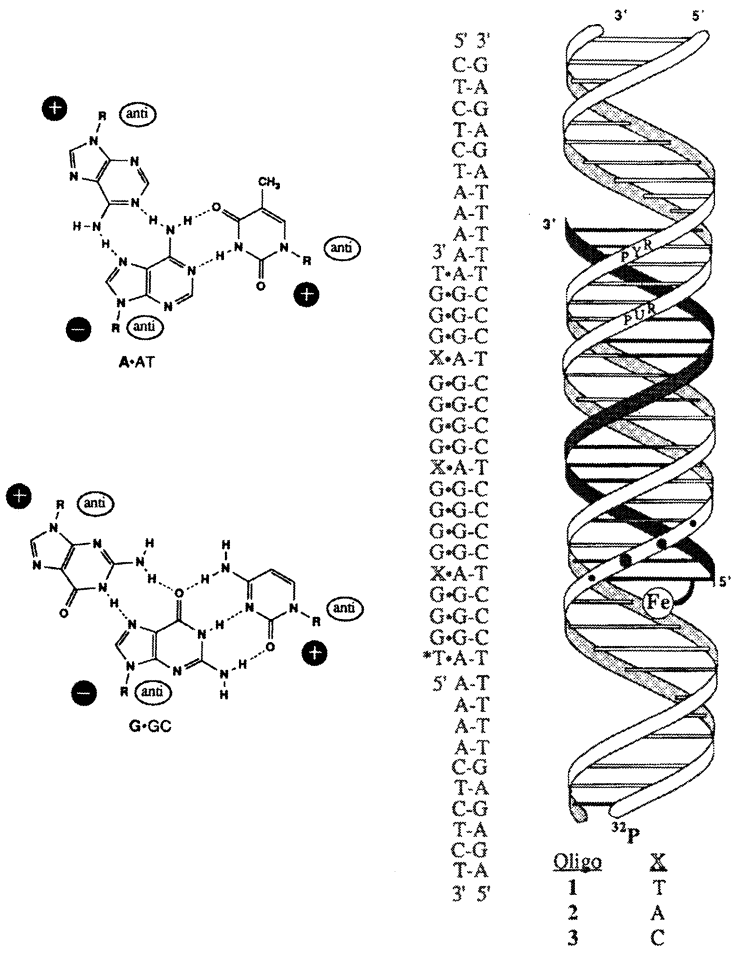


Figure 8. The purine triple helix motif. Isomorphous base triplets of AAT and GGC where the third stand is antiparallel to the purine Watson–Crick strand. Plus and minus indicate relative polarities of the phosphate-deoxyribose backbones.

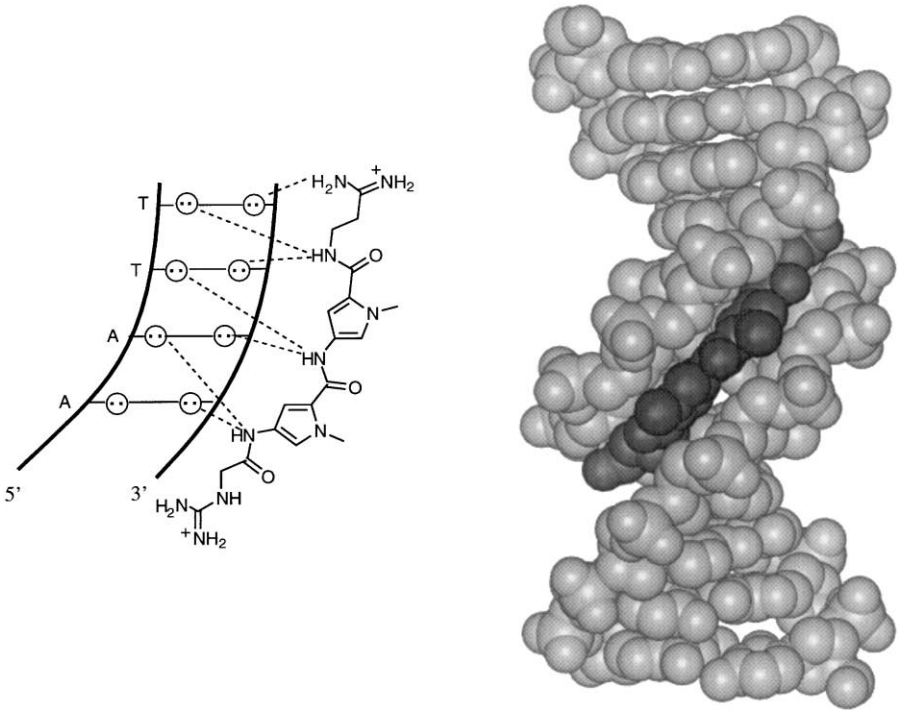


Figure 9. X-ray crystal structure of a 1:1 complex of netropsin: DNA.

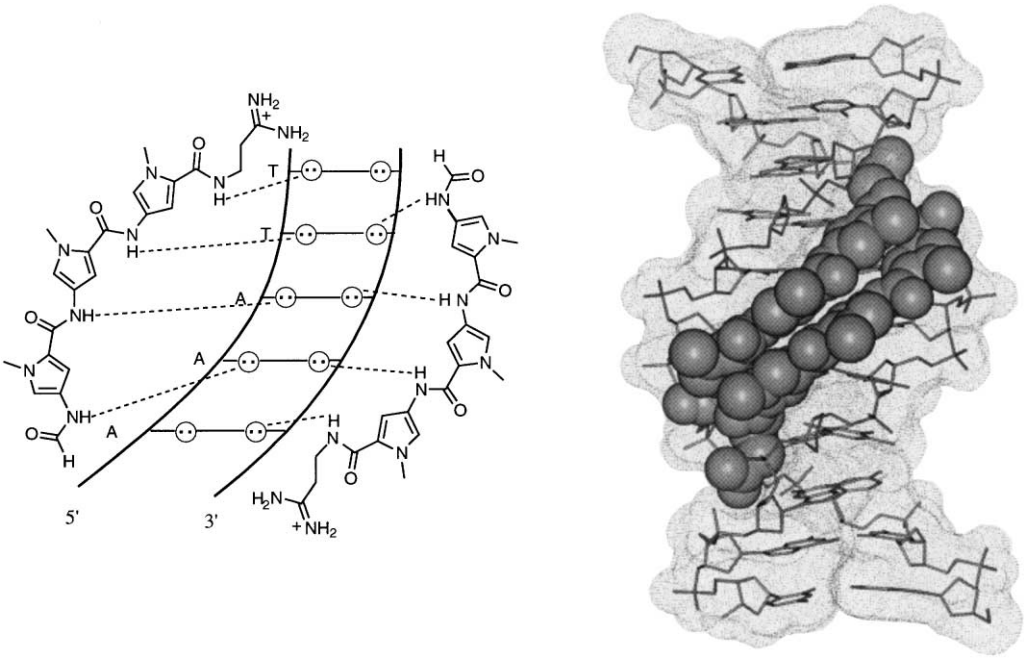


Figure 10. NMR structure of 2:1 complex of distamycin-DNA complex.

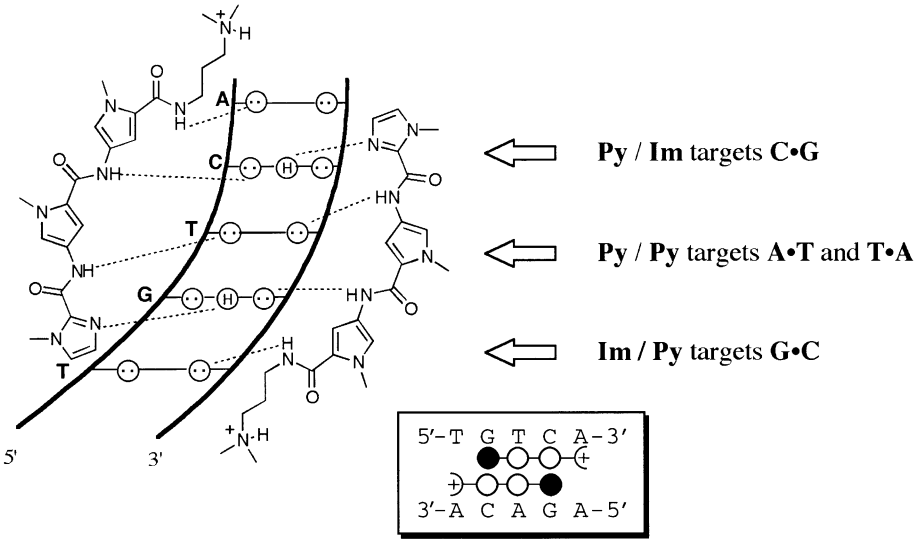


Figure 11. Model for antiparallel dimer ImPyPy binding 5′-TGTCA-3′. Unsymmetrical pair Im/Py distinguishes GC from CG and both from AT and TA.

Caltech biologist Barbara Wold was able to demonstrate that triple helix forming oligonucleotides inhibit DNA binding proteins and could modulate transcription in cell-free systems.[34,35] However, our group was concerned that the high molecular weight anionic oligomers would suffer from poor cellular uptake hindering transcription experiments in cell culture and therefore we continued our search for other modular organic subunits for targeting the grooves of DNA.[36–39] In retrospect, the triple helix approach was a benchmark for the field because it demonstrated that in principle, there was a chemical method different from nature's protein pool which could read the double helix with specificity sufficient to target single sites in gigabase size DNA.[40–44]

Pairing Rules for Minor Groove Recognition

Parallel to our triple helix efforts, we maintained our exploratory program in the minor groove, a necessary diversification in case the four base pair code in the major groove proved insurmountable. In the mid-1980s, Warren Wade constructed analogues of distamycin by single atom changes with the goal to alter the A,T binding specificity of the natural product. In a landmark paper, Richard Dickerson had obtained the first X-ray structure of a complex of netropsin and DNA[45] (Fig. 9). The 1:1 complex was a classic example of shape-selective recognition with the crescent PyPy bound snugly by the walls of the narrow minor groove of an A,T tract of

DNA. The NHs of the carboxamides pointed toward the minor groove floor of the helix making specific hydrogen bonds with the A•T and T•A base pairs (N3 of A and O2 of T). Dickerson and Lown suggested that imidazole (Im) replacing pyrrole (Py) rings might allow for steric reasons to read the exocyclic NH_2 of G,C base pairs in the 1:1 complex.[45,46] We were thinking along the same lines and set out to test this hypothesis.

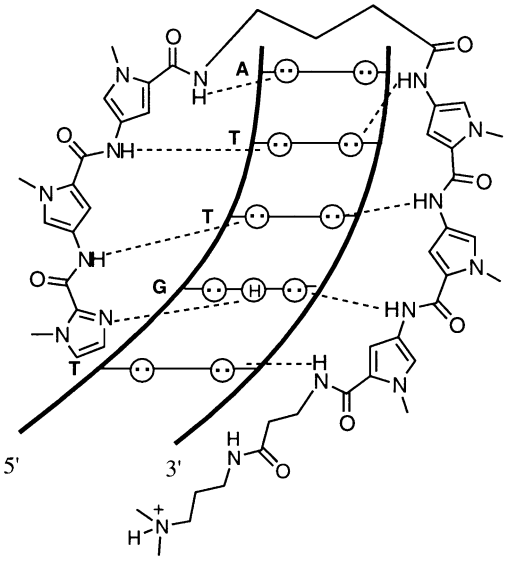


Figure 12. The hairpin motif. The amino and carboxy terminus of the antiparallel dimers are connected by γ-aminobutyric acid (γ).

Wade synthesized the novel polyamide, ImPyPy which was designed to bind, according to the 1:1 model, the sequence 5′-(G,C)(A,T)$_2$-3′ in a single orientation. Using MPE•Fe(II) footprinting, Wade scanned a library of sites on several DNA restriction fragments and found the molecule bound not the expected sequence but rather a new unanticipated five base pair sequence, 5′-(W)G(W)C(W)-3′ (where W = A or T).[47] G was always in the second position and C was always preferred in the fourth position. In addition, affinity cleavage experiments demonstrated C2 symmetry, opposite the single orientation expected for a 1:1 complex.[47] Quantitative footprint titration had not yet been introduced to the field of small molecule–DNA recognition and Warren decided to embark on a more rigorous characterization of ImPyPy binding the high affinity sites by this method. Quantitative footprint titration with ImPyPy (using MPE•Fe) could only be fit by a cooperative 2:1 isotherm at the high affinity sites.[47] The (ImPyPy)$_2$/DNA stoichiometry of binding was unanticipated. While Wade and I deliberated how to write up the work,[47] a remarkable NMR paper appeared from David Wemmer's laboratory at UC Berkeley. Wemmer demonstrated by NMR that the natural product distamycin (PyPyPy) could bind A,T sequences of DNA as an antiparallel 2:1 complex as well as 1:1[48] (Fig. 10). This observation provided a missing clue. Our synthetic ImPyPy binds as an antiparallel dimer as well, suggesting that pairs of Im/Py recognize G•C, Py/Im recognizes C•G and neither A•T or T•A[49] (Fig. 11). The Py/Py pair did not distinguish T•A and A•T. In one of my Todd Lectures in Cambridge, England in May 1989, I presented the antiparallel 2:1 model to explain the remarkable specificity observed for ImPyPy binding 5′-

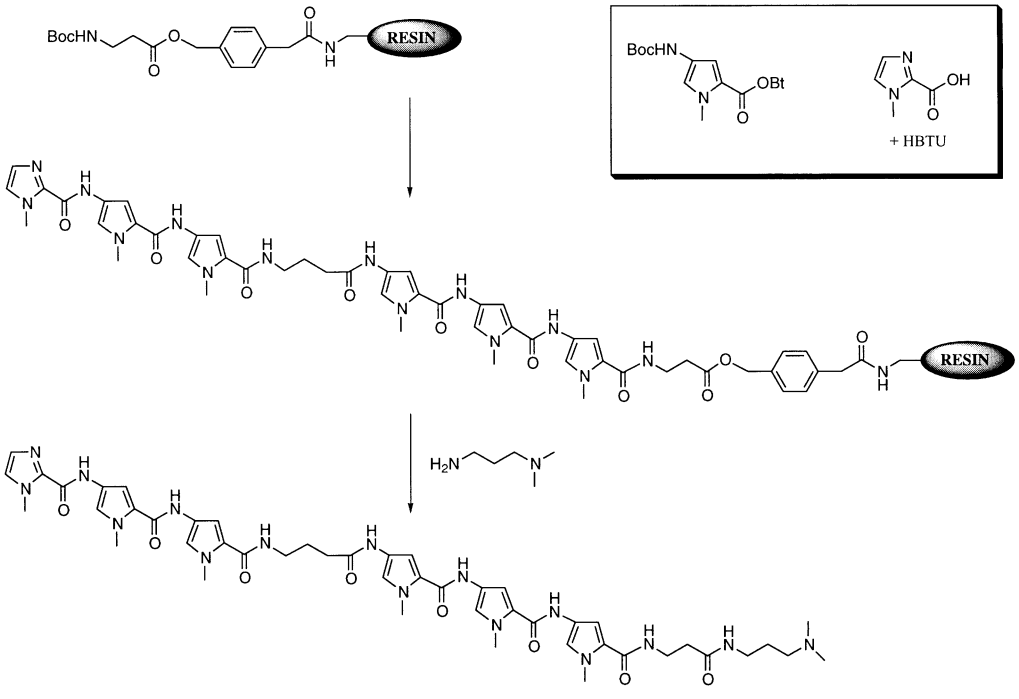


Figure 13. Solid phase synthesis of polyamides.

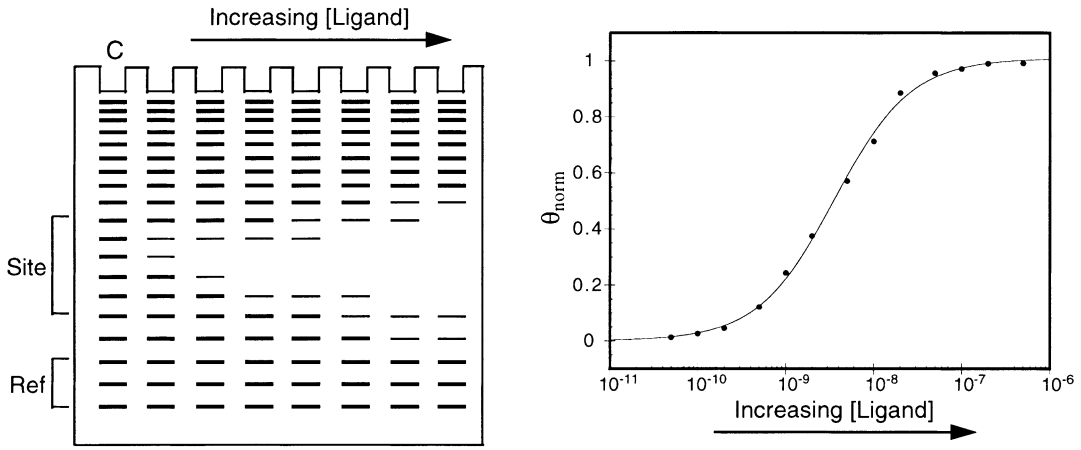
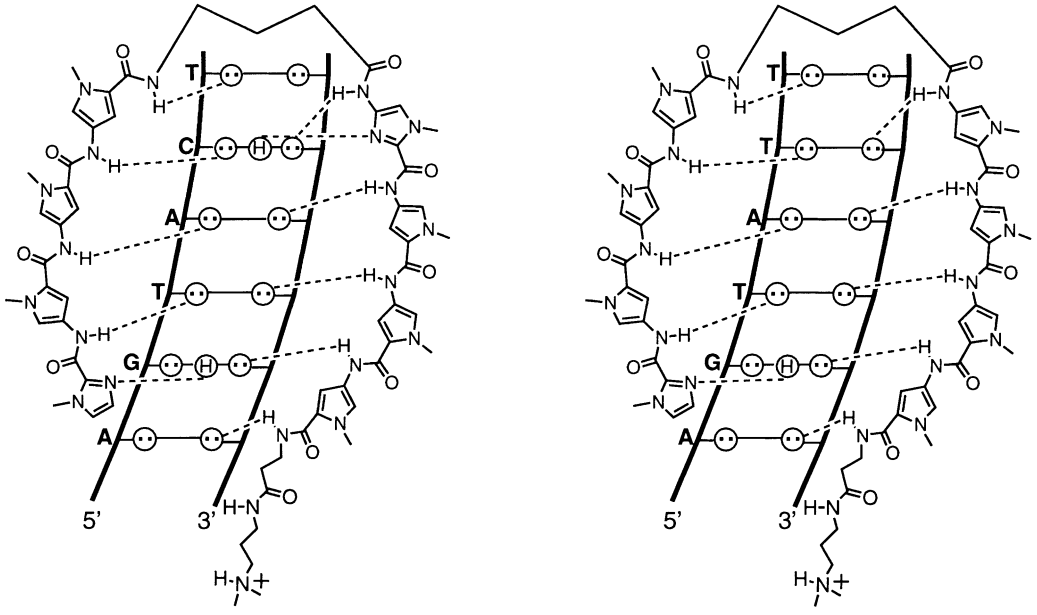


Figure 14. Quantitative footprinting titration analyses. (left) Cleavage pattern generated by quantitative DNaseI footprinting titration on a 3′ end-labeled DNA fragment in the presence of increasing ligand concentration (right) Langmuir binding titration isotherm obtained from DNaseI data.



Equilibrium Association Constants (M^{-1})

Binding Site	**Im**PyPyPy-γ-**Im**PyPyPy-β-Dp	**Im**PyPyPy-γ-**Py**PyPyPy-β-Dp
5'- AGTACT -3'	**3.7×10^{10}**	**5.0×10^{8}**
5'- AGTATT -3'	**4.1×10^{8}**	**3.5×10^{9}**

Conditions: 10 mM Tris•HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$, pH 7.0, 22°C

Figure 15. Eight ring polyamides bind match sequence with affinity and specificity similar to protein transcription factors.

WGWCW-3′, always G•C in the second position and C•G in the fourth position of the binding site. Olga Kennard, a good friend and leader in the field of nucleic acid structure, was skeptical that two molecules could fit side-by-side in the minor groove of DNA. Upon my return to the States, I decided we needed a structure proof before publication of our data. By now, Warren had graduated but a first-year student, Milan Mrksich synthesized a new batch of the polyamide for structure studies. In April 1991, I gave the Calvin lecture at UC

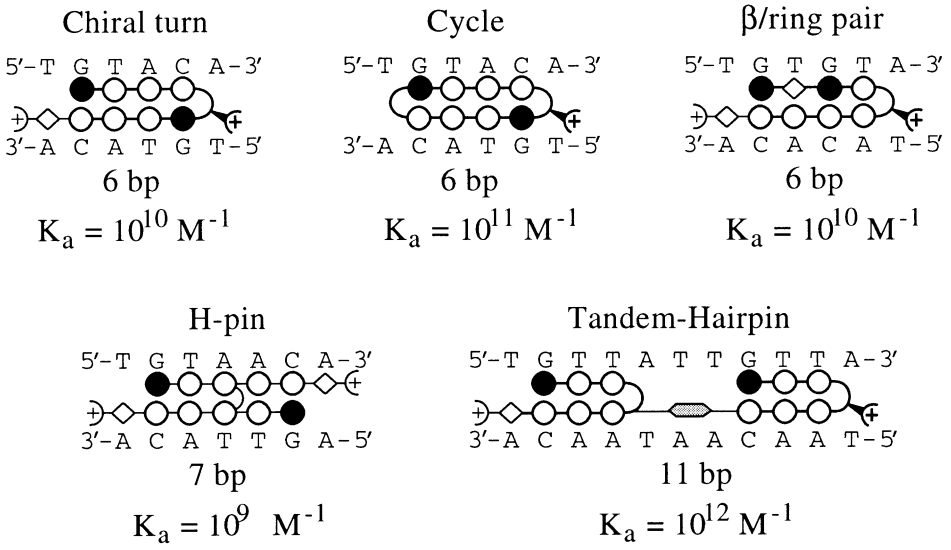


Figure. 16. Novel polyamide-DNA binding motifs with equilibrium association constants (K_a) shown. Chiral turn: amino-substitution at the α-position of the γ-turn residue leads to enhanced binding affinity (10-fold) without loss of specificity, higher orientational selectivity, and offers potential for further substitution. Cycle: cyclic polyamides show higher affinity with respect to the hairpin structure. β/ring pair: the β/ring pair relaxes the ligand curvature and allows the hairpin structure to adjust to the microstructure of non-B-form helices. In some cases, the binding affinity of the β/ring-polyamides is significantly higher than that of the ring/ring analogue. H-pin: compared to their non-linked analogues, H-pins exhibit higher binding affinity. Tandem hairpin: hairpin dimers recognize large DNA sequences with excellent binding affinity and specificity.

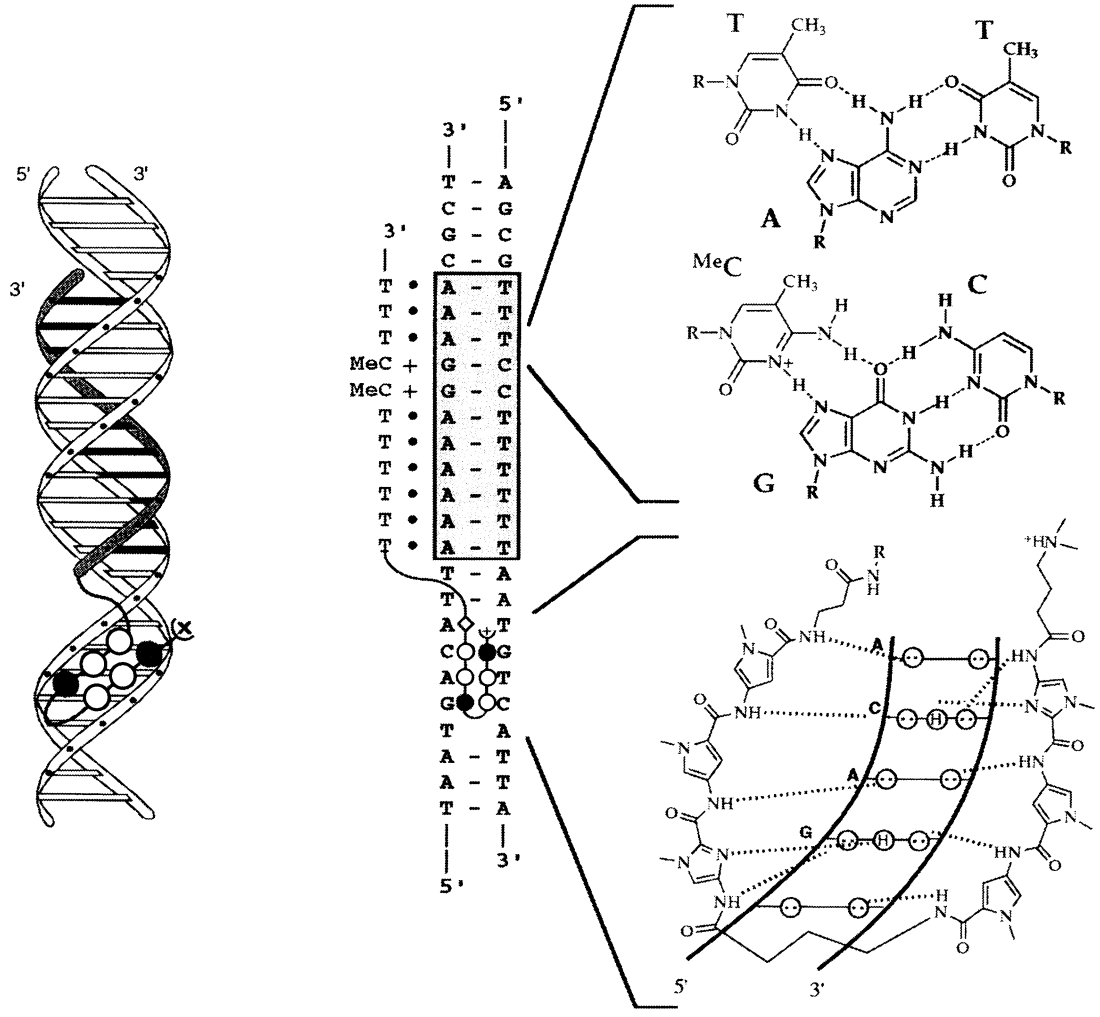


Figure 17. An oligonucleotide-polyamide conjugate targets simultaneously the major and minor grooves of DNA.

Berkeley and with great anticipation discussed our unpublished $(ImPyPy)_2$–DNA model with David Wemmer. The Berkeley team agreed to determine the structure by NMR. The specificity paper and the NMR structure proof were published in 1992.[49–51]

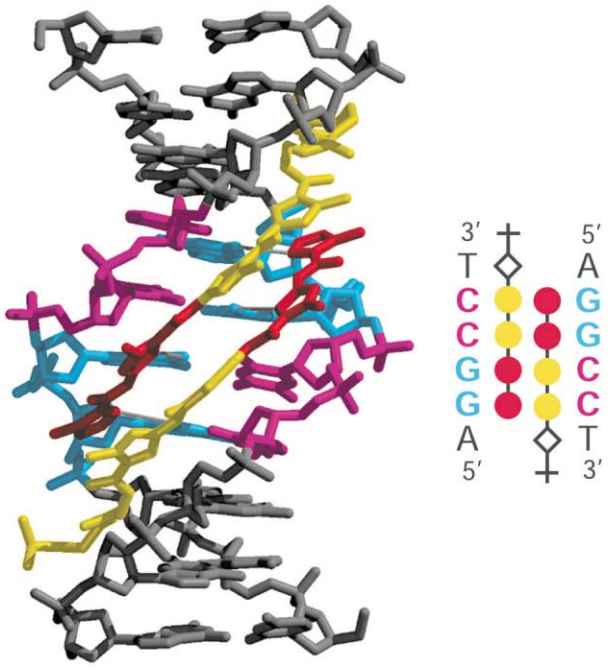


Figure 18. Stereodiagram of the polyamide-DNA complex (ImImPyPy-β-Dp)$_2$• 5′CCAGGCCTGG-3′. (Im-N-methylimidazole carboxamide, Py = N-methylpyrrole carboxamide, β = β-alanine, Dp = dimethylaminopripylamide).

Milan Mrksich then covalently linked the antiparallel dimers by two different strategies. He coupled antiparallel heterodimers[52] with a short methylene linker across the backbone which we called the H-pin motif.[53] Importantly, he coupled the amino- and carboxyl termini of the antiparallel dimers with an aliphatic amino acid (γ) to create a U-shaped motif which bound the minor groove of DNA with very high affinity and specificity.[54] The hairpin structure kept the rings unambiguously 'paired' avoiding slipped dimers[55] (Fig. 12) We used this binding motif in subsequent experiments to study polyamides with affinities and specificities similar to nature's proteins.

From that point forward, a stream of projects were completed by a highly motivated group of coworkers at Caltech to explore the scope and limitations of the 'pairing rules.'[55–79] The effort was accelerated in large part by Eldon Baird who moved our group from solution to solid phase synthesis of polyamides.[57] (Fig. 13) The reduction in synthesis time from months to days allowed us to dream of new motifs and test our ideas in very short order: extended hairpins,[60] cooperatively binding hairpins,[72] β/ring pairs,[73] chiral turns,[74] cycles,[75] hairpin dimers,[76] major/minor groove hybrids[80,81]— with the energetics of binding match and mismatch sites characterized by quantitative footprinting titration analyses (Figs 14–17). Ken Breslauer's laboratory provided an enthalpic and entropic dissection of the energetics revealing that the specificity of the Im/Py pair for GC was driven by a favorable enthalpic contribution.[65,79] In a breakthrough effort, Clara Kielkopf was successful in co-crystallizing an ImImPyPy dimer with a GGCC core sequence (Fig. 18). High resolution X-ray

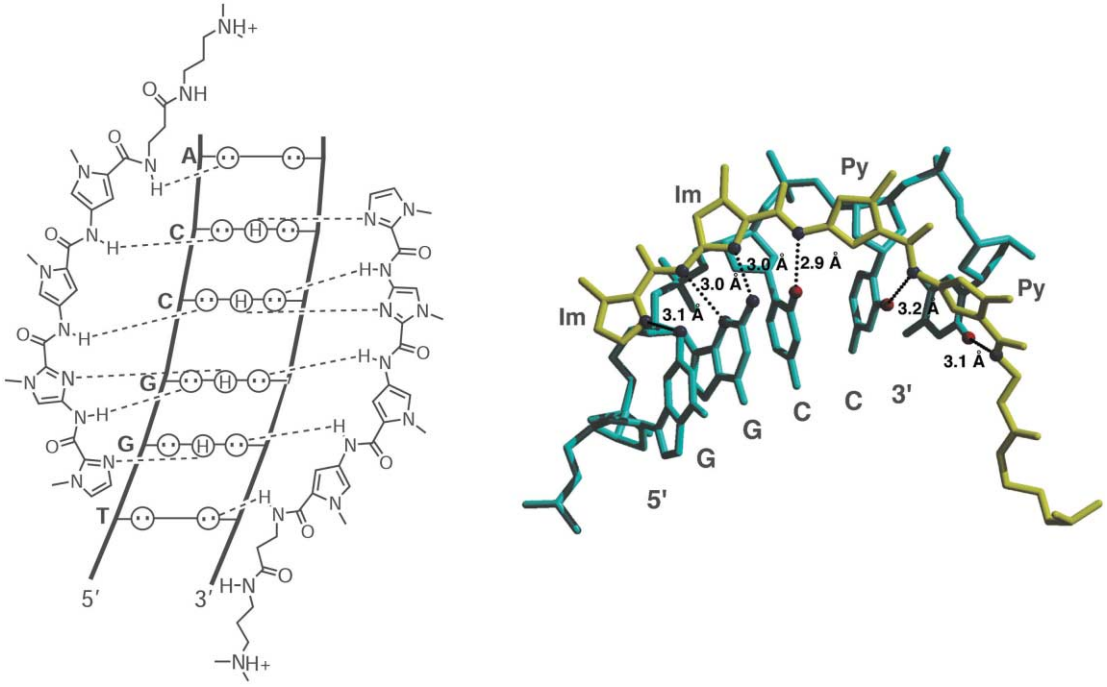


Figure 19. (left) Model of the polyamide bound as an antiparallel head to tail dimer. (right) Isolated view of one of the ImImPyPy-β-Dp molecules and hydrogen bonds with the adjacent DNA strand. The hydrogen bond lengths given are the average of the ncs-related strands.

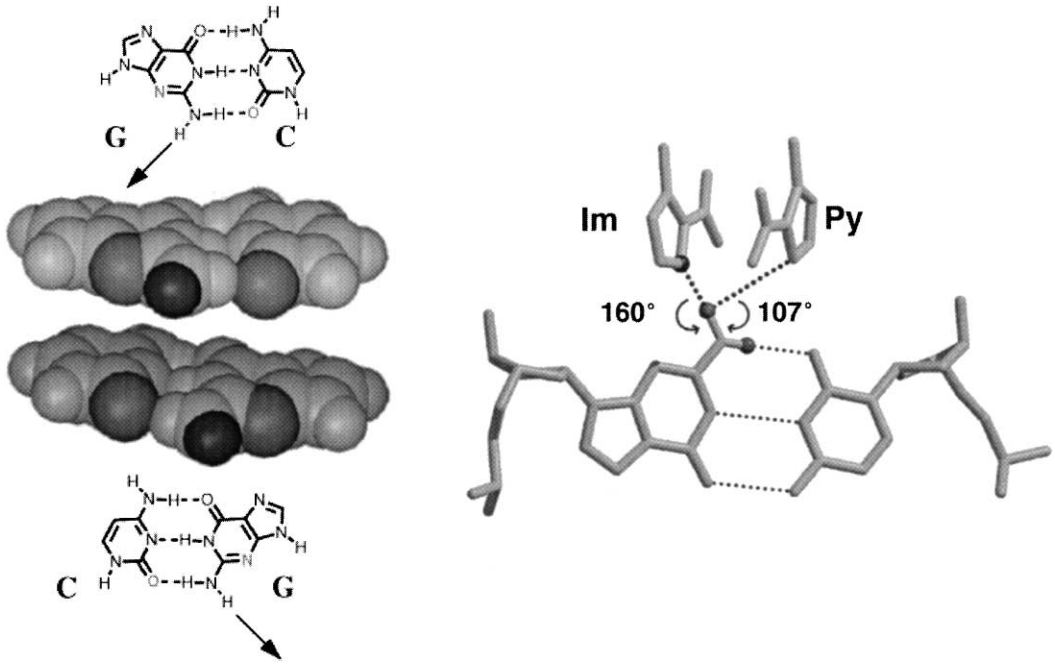


Figure 20. (left) Space filling model of GC versus CG. Note the exocyclic N-H of guanine points to one side of the minor groove. (right) Geometry of the Im-Py pair interacting with the G5•C16 base pair. An imidazole on the cytosine side of the base pair would be unable to form an optimum hydrogen bond with the hydrogen at the guanine exocyclic amine. This allows the Im-Py pair to discriminate the G•C base pair from C•G.

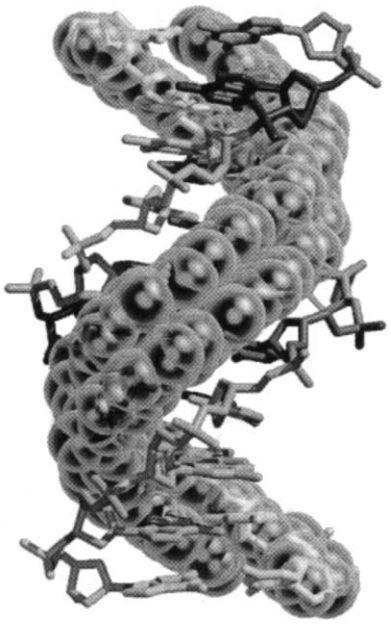


Figure 21. An extended polyamide helix. Least squares minimized, mean coordinates for a polyamide monomer were translated and rotated by the parameters of the polyamide helix. After a period of approximately seven residues, the 14 great twist of the polyamide with respect to B-form DNA manifests as clash with the walls of the minor groove.

analysis (with Doug Rees) of the 2:1 (ImImPy-Py)$_2$•DNA complex revealed in detail that the Im/Py pair makes three specific hydrogen bonds to a G•C base pair[82] (Figs 19 and 20). Although the rise per residue for each ring subunit matches each base pair step in the double helix, we found that the polyamides are over-curved suggesting an explanation why there is an energetic penalty after four or five contiguous rings are bound in the minor groove (Fig. 21). This energetic penalty was eliminated by introducing at strategic locations β alanine (β) as an aliphatic substitute for a Py ring.[55] We imagine the β residue acts like a spring to allow the crescent-shaped ligand to match the curvature of the DNA helix (Fig. 22). Sue Swalley demonstrated the β,β pair was specific for A,T base pairs and there was no sacrifice in specificity.[69] The use of the β/β pair as a flexible spacer unit and sequence-specific DNA binding element allows targeting of 11 to 16 base pairs with highly cooperative dimeric polyamides.[78]

Because of sequence dependent microstructure of the double helix, certain DNA sequences remain challenging target sites for hairpin polyamides. For example, the eight ring polyamide ImPyImPy-γ-ImPyImPy-β-Dp, designed to recognize 5′-TGCGCA-3′, binds only with a moderate affinity to its target site. Selective placement of an aliphatic β-alanine (β) residue paired side-by-side with either a Im or Py aromatic amino acid (β/Im, β/Py pair was found to compensate for sequence composition effects for recognition of the minor groove of DNA by hairpin polyamides. Im-β-ImPy-γ-Im-β-ImPy-β-Dp which incorporates Im/β pairings binds the 'problematic' 5′-TGCGCA-3′ sequence at sub-nanomolar concentrations.[73]

Selective amino-substitution of the prochiral α-position of the γ-turn residue relocates the cationic change from the hairpin carboxyl terminus. As a consequence of this

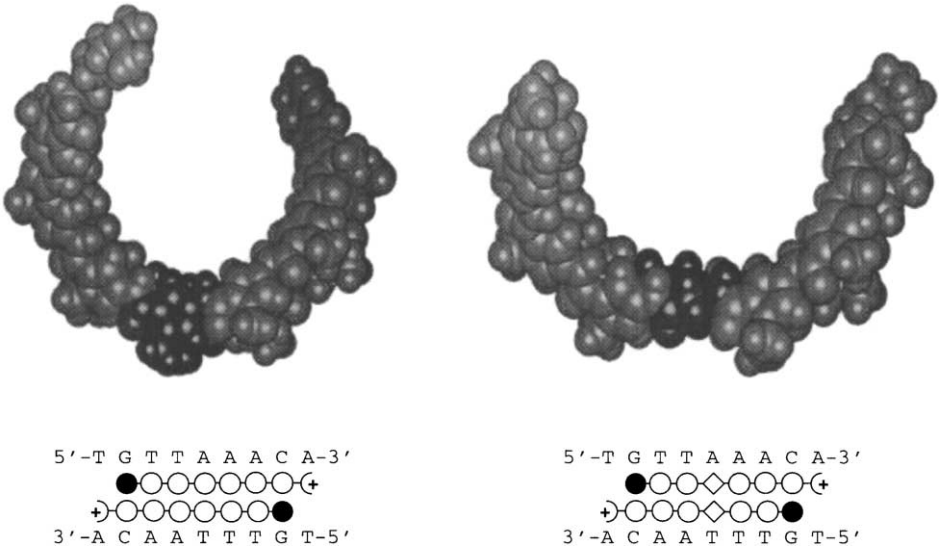


Figure 22. Space filling model of seven ring, ImPyPyPyPyPyPy-DP (left) and of 3-β-3 motif, ImPyPy-β-PyPyP-Dp (right). The 3-β-3 dimer curvature is lower than that of 3-Py-3 and fits better with the shape of the DNA helix.

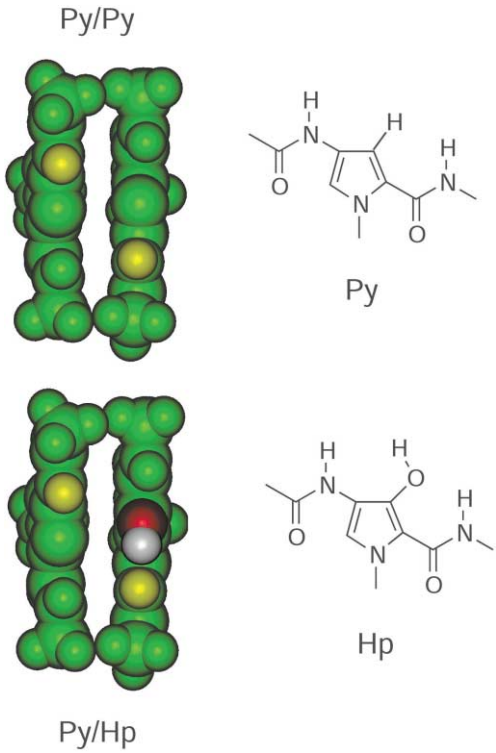


Figure 23. (top) Space filling model of the symmetrical Py/Py side-by-side antiparallel pair . (bottom) Space filling model of the unsymmetrical Py/Hp pair both viewed from the floor of the DNA helix.

substitution the affinity of the hairpins for the DNA target increases by about a factor of 10 without loss of specificity and a 'reversed' binding orientation becomes energetically less favorable.[74] The amino group at the turn unit allows further substitution as demonstrated in the synthesis of tandem hairpin dimers. Linked hairpin dimers not only increase the binding site size but significantly increases the binding affinity of the ligand as well. For example, a six ring/six ring tandem binds an 11 bp target site at picomolar concentration with excellent specificity compared to the double bp mismatch site.[76]

In the mid-1990s, Eldon Baird, Sarah White, Jason Szewczyk and Jim Turner then teamed up to complete the recognition code for distinguishing T•A and A•T base pairs.[83] From model building they designed a third aromatic ring, hydroxypyrrole (Hp) in order to introduce the unsymmetrical Hp/Py pair. We reasoned that a small substituent on one corner of the Py/Py pair would disfavor A over T (steric destabilization against A, but steric permissiveness at T). Hydroxyl was chosen as the substituent for reasons of size and the added possibility of one (or perhaps two!) hydrogen bonds to the lone pair electrons at O2 of T (Figs 23 and 24). The specificity paper[83] and a high resolution X-ray structure (with Rees and Kielkopf) were both published in 1998 and the 'four base pair code' was now formally complete [83–87] (Fig. 25–28). In retrospect, the Im/Py 'discovery' was made possible by our decision in 1982 to scan libraries of sites (footprinting, affinity cleaving), and not presume the validity of any prior models. In contrast, the Hp/Py pair to distinguish TA from AT was truly 'invented,' revealing the maturation of the field of molecular recognition of DNA from serendipity to successful design[83–89] (Tables 2 and 3).

Currently, we are interested to ask whether bifunctional DNA binding hairpins could covalently react with the minor groove of DNA and inhibit polymerase elongation during transcription. Aileen Chang and Nick Wurtz have synthesized a class of hairpin conjugates with alkylating agents attached based on analogues of

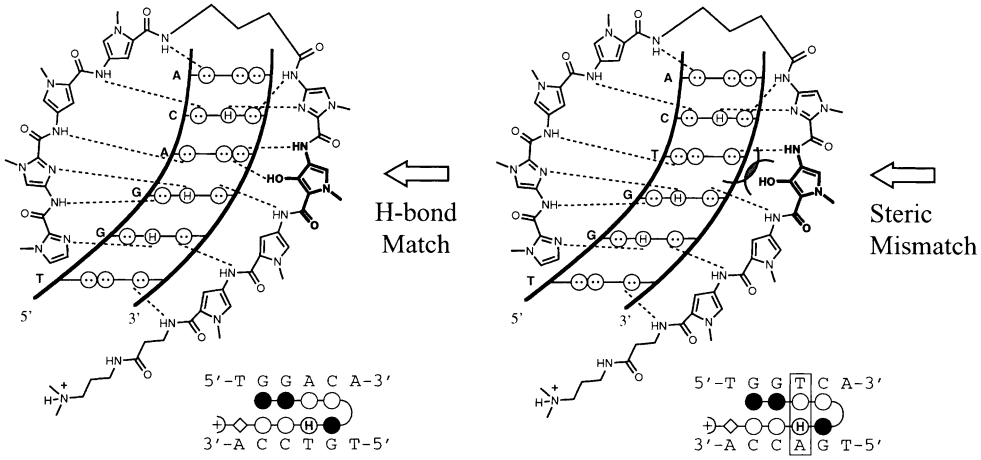


Figure 24. Models for Py/Hp binding AT, but not TA. (left) the Hp ring forms two hydrogen bonds with O2 of T and the hydroxyl fits in the cleft between A and T. (right) The Hp-OH has a steric clash with the A of the TA base pair.

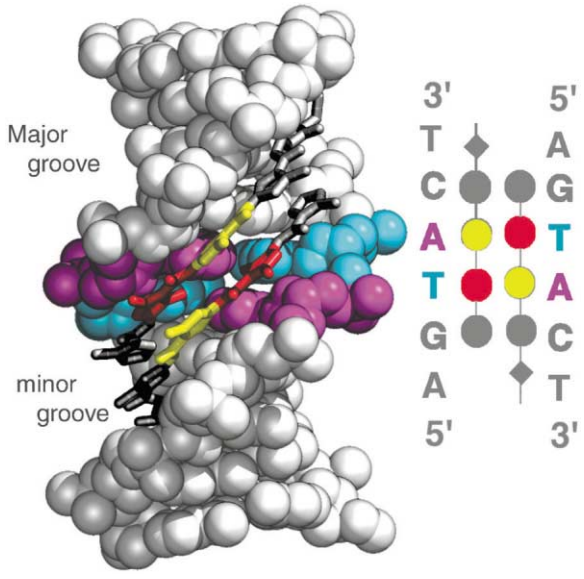


Figure 25. Space-filling model of $(ImHpPyPy)_2$•5′-CCAGTACTGG-3′. Adenosine is purple and thymidine cyan. The Hp is red and the Py paired with it is yellow.

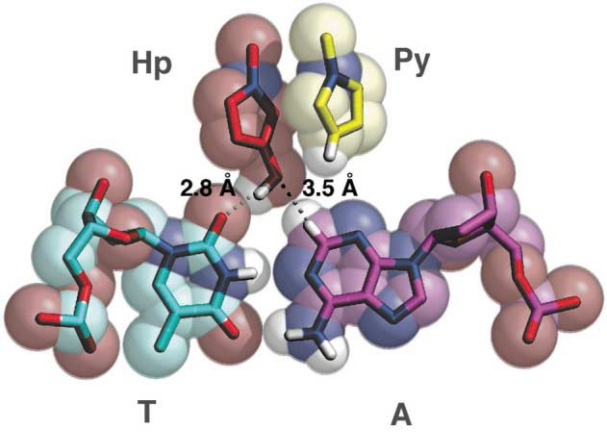


Figure 26. Space-filling model of the Hp/Py pair interaction with the T•A base pair shows that the Hp-OH tightly fits the cleft formed by the adenosine C2H.

the natural product CC1065 and chloroambucil.[90,91] At nM concentrations, we observe near quantitative cleavage at single nucleotide positions (Fig. 29). This paves the way for 'genetic surgery' by small molecules in coding regions of genes.

Regulation of Gene Expression by Small Molecules

There are approximately 35,000–50,000 genes in each human cell. It is now understood that Nature controls the expression in time and space of each of these genes

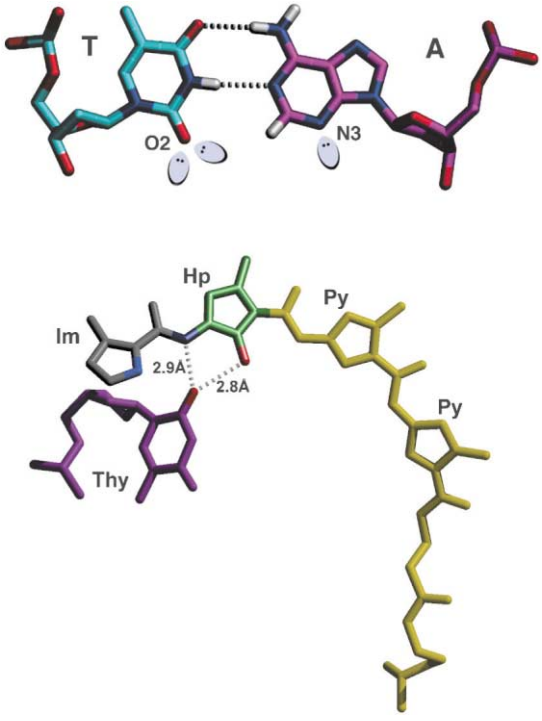


Figure 27. (top) Anatomy of the T•A base pair. Lone pair electron in the minor groove are shown as ovals and the Watson–Crick hydrogen bonds of the base pair as dotted lines. (bottom) The hydrogen bonds between ImHpPyPy and one strand of DNA, indicated by dashed lines. Note the two hydrogen bonds between Hp and O2 of T.

by a remarkable three dimensional switch, that is ~50–100 protein complex assembled on a few hundred specific base pairs of 'promoter' DNA sequence encoded upstream from the RNA polymerase start site and coding region. Protein transcription factors bind very specific sequences of DNA in the promoter region of each gene modulating the expression of that gene. Could a sequence-specific minor groove polyamide inhibit and complete with transcription factor binding and interfere

Table 2. Pairing code for minor groove recognition[a]

Pair	G•C	C•G	T•A	A•T
Im/Py	+	–	–	–
Py/Im	–	+	–	–
Hp/Py	–	–	+	–
Py/Hp	–	–	–	+

[a]Favored (+), disfavored (–).

Table 3. Pairing code β-alanine (β), Py/Py and Im/Im pairs.[a]

Pair	G•C	C•G	T•A	A•T
Im/β	+	–	–	–
β/Im	–	+	–	–
Py/β	–	–	+	+
β/Py	–	–	+	+
β/β	–	–	+	+
Py/Py	–	–	+	+
Im/Im	–	–	–	–

[a]Favored (+), disfavored (–).

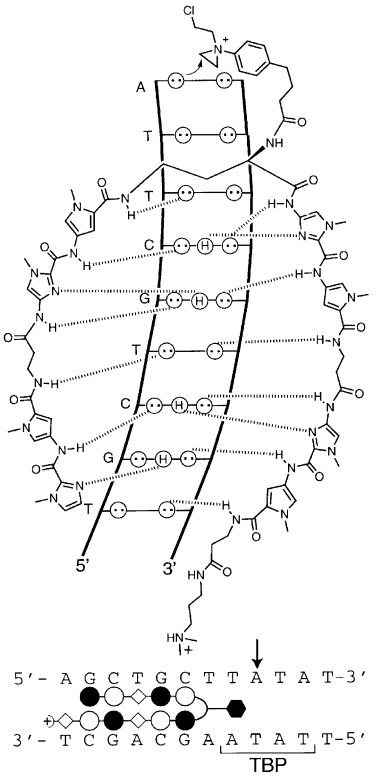


Figure 29. (top) Hydrogen-bonding model and alkylation mechanism of polyamide-CHL conjugate ImPy-β-ImPy-(R)^{CHL}-γ-ImPy-β-ImPy-β-Dp (2) bound to the minor groove of 5′AGCTGCT-3′. Circles with two dots represent the lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Model of polyamide conjugate bond to the match site 5′-AGCTGCT-3′. Black and white circles represent imidazole (Im) and pyrrole (Py) polyamide rings, respectively. Diamonds and hexagon represent β alanine (β) and CHL, respectively. (*R*)-2,4-diaminobutyric acid [(R)γ] and dimethylamino-propylamide (Dp) are depicted as a curved line and a plus sign, respectively. TBP, TATA box binding region.

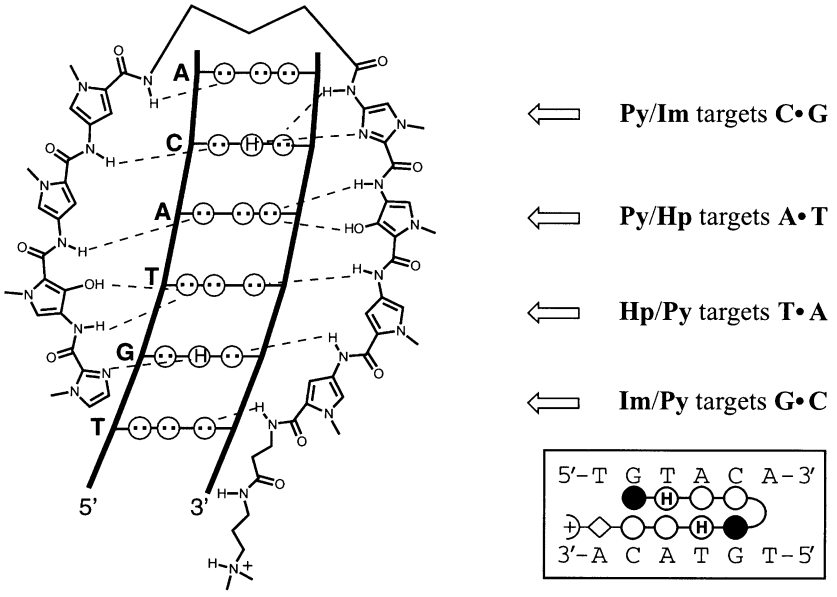


Figure 28. Molecular recognition of the minor groove of DNA. Minor groove hydrogen bonding patterns of Watson–Crick bps. Circles with dots represent long pairs of N(3) and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. The R group represents the sugar-phosphate backbone of DNA. Binding model for the complex formed between ImHpPyPy-γ-ImHpPyPy-β-Dp and a 5′-TGTACA-3′ sequence. Hydrogen bonds are shown as dashed lines. Im is black circles, Py is open circle and Hp is open circle with H.

with the protein complex that is responsible for the recruitment of the RNA polymerase II (Fig. 30)?[92–94] If so, we might reprogram with small molecules the software in the promoter leading to altered levels of gene expression. John Trauger and Eldon Baird, with biologist Joel Gottesfeld at Scripps, set out to demonstrate that synthetic hairpin polyamides could inhibit specific gene expression.

In a first case study the Caltech and Scripps team found that the DNA-binding activity of the 5S RNA gene-specific transcription factor TFIIIA was inhibited by an eight-ring hairpin polyamide that bound within the recognition site of the nine-zinc-finger protein in the minor groove. Transcription of 5S RNA genes by RNA polymerase III was suppressed in vitro and in cultured Xenopus kidney cells.[95] We then examined whether polyamides can specifically regulate genes transcribed by RNA polymerase II. The HIV-1 enhancer/promoter element contains binding sites for multiple transcription factors, among them Ets-1, LEF-1 and TBP. Two hairpin polyamides were designed to bind DNA sequences immediately adjacent to the binding sites for these transcription factors. These synthetic ligands specifically inhibit DNA-binding of each transcription factor and HIV-1 transcription in cell-free assays.[96] Don Mosier at Scripps demonstrated that when used in combination, the two polyamides inhibit virus replication by more that 99% in isolated human peripheral blood lymphocytes with no obvious cell toxicity[96] (Fig. 31).

The question arises whether one can activate as well as repress gene expression. Positive regulation of gene expression requires transcriptional activators. Activator proteins bind to DNA and recruit the transcriptional machinery to a proximal promoter, thereby stimulating gene expression. These steps define the initial regulatory decisions in a transcriptional circuit, and misregulation at any stage can result in a variety of human diseases. Activators achieve specificity in targeting genes by a DNA recognition module which binds to cognate DNA sequences near a promoter and in most cases binding specificity is further enhanced by dimerization. The key functional module, the activation region, is thought to bind several components of the transcriptional machinery. Many of these components of the machinery exist in large multi-subunit complexes which associate with the RNA polymerase II, and are known as the RNA polyamerase II holoenzyme. The holoenzyme along with a few additional factors, like TBP, constitutes the transcriptional machinery that is recruited by activators to most promoters in vivo. It is believed that weak

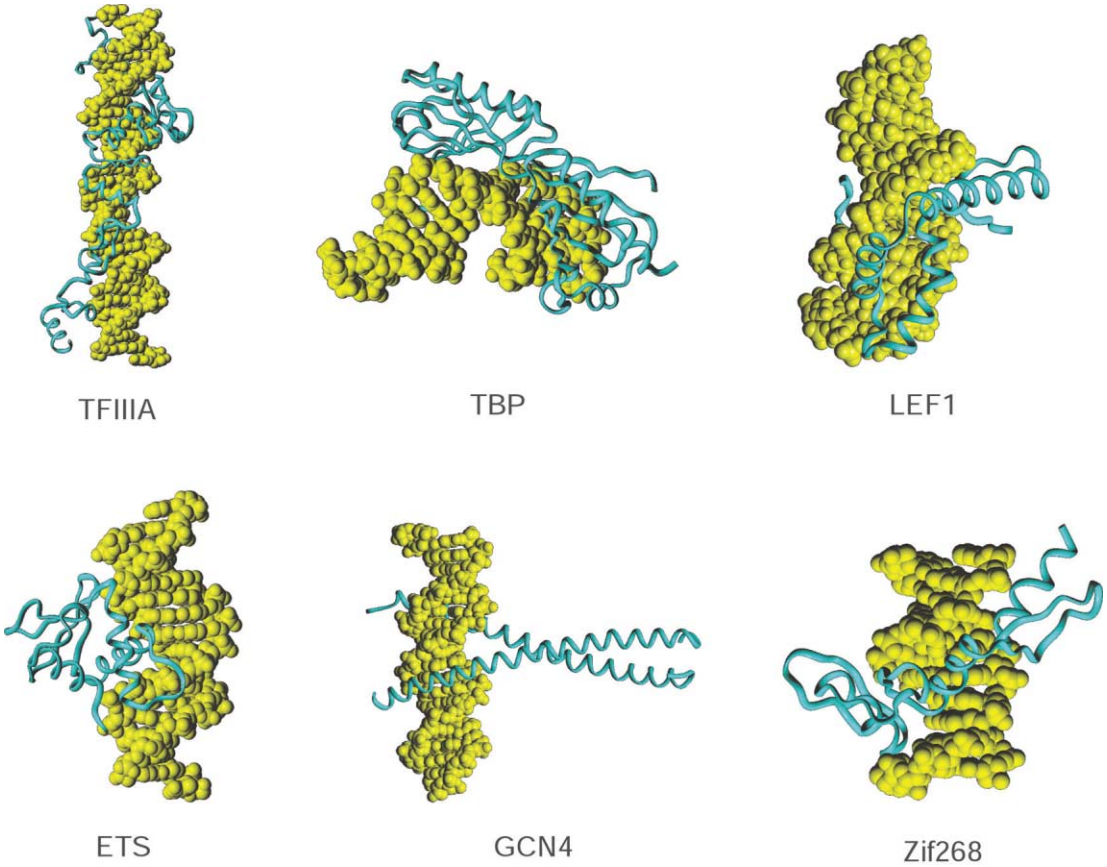


Figure 30. Examples of protein·DNA complexes as targets for inhibition by designed polyamides. TFIIIA is a zinc finger, TBP is a minor groove binding protein, LEF-1 is an HMG box, Ets-1 is a winged-helix-turn-helix, GCN4 is a bZIP protein, and Zif 268 is a zinc finger.

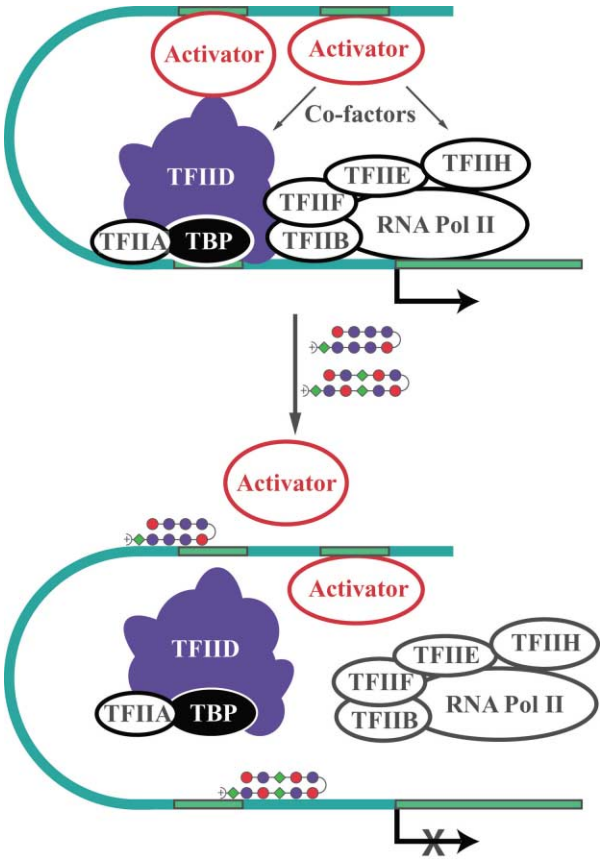


Figure 31. The inhibition of gene transcription by polyamides. (a) Model of the transcriptional machinery required for the initiation of gene transcription mediated by RNA polymerase II (RNA Pol II); (b) The addition of polyamides targeted to a DNA site proximal and overlapping a transcription factor (TF) site within the promoter region of a specific gene. As a consequence, transcription of the gene does not take place.

interactions between an activating region and several components of the holoenzyme result in high avidity 'multi-dentate' binding by the activator to the holoenzyme. In addition, acidic activating regions are believed to contact and recruit nucleosome modifying activities to promoters.

Attempts to generate artificial activators have relied on this principle of recruitment. In one example, a dimer of the natural product FK506 was used to couple a chimeric DNA binding protein to an activating region, thus upregulating a gene bearing upstream cognate binding sites.[97] More recently attachment of an activating region to a designed Zn(II) finger motifs was shown to up-regulate gene expression.[98] Our group teamed up with the Ptashne laboratory at Sloan Kettering to create artificial activators that are capable of targeting a wide variety of DNA sequences based upon sequence-specific DNA binding polyamides. Anna Mapp synthesized a hairpin polyamide tethered by a 36 atom straight chain linker to a short peptide (20 mer) activation domain (AH) of sequence PEFPGIELQELQELQALLQQ.[99] The conjugate (PA-L-AH), only 4.2 kDa in size, was demonstrated by biologist Aseem Ansari to upregulate transcription in a cell-free system.[99] As a prelude to cell culture experiments, we then set out to ask what is the smallest polyamide–peptide conjugate which is still functional for activation of gene expression. To create our minimal motif, we replaced the 20 amino acid AH peptide with 16 residues, DFDLDMLGDFDLDMLG, derived from the activator domain of the viral activatorVP16 and reduced the linker from 36 to 8 atoms. The second generation 'minimal' polyamide–peptide conjugate, 3.2 kDA in size, activated transcription with comparable efficiency to PA-L-AH (Fig. 32).[100]

Limitations

Will the pairing code allow targeting of all 524,800 ten base-pair sequences of DNA with the criteria of high affinity (K_D in subnanomolar range) and good specificity [K(match)/K(single base-pair mismatch) ratio greater than a factor of 10]? The answer is likely no, due to the sequence-dependent microstructure of double helical DNA. But having synthesized and characterized approximately 250 polyamides, we now estimate that we are able to target as much as 50% of the DNA sites on any promoter which will be sufficient to target most important transcription factors. Another key concern is whether the DNA minor groove in chromatin is accessible (Fig. 33). The genetic information of a single eukaryotic cell is stored in DNA molecules that are over 2 m in length, but compacted in the cell nucleus to nearly one millionth of this dimension. This is achieved by a hierarchical scheme of folding and compaction into chromatin, in which the DNA can still be manipulated during transcription, replication, and repair. At the first level of organization, two tight superhelical turns of DNA (147 bp in length) are wrapped around a disk-shaped protein assembly of eight histone molecules to form the nucleosome core particle (NCP). An additional 20–80 bp of linker DNA extends from the NCP, forming the fundamental repeating unit of chromatin, the nucleosome. Nucleosomes are subject to multiple higher order levels of organization, which require the presence of the linker histone H1. As the molecular substrate for most activities involving the genome, nucleosomes do not simply serve to compact DNA, but also affect accessibility of specific sequences, and actively interact with components of the molecular machinery in the nucleus. The ability of DNA-binding proteins to recognize their cognate sites in chromatin is restricted by the structure and dynamics of nucleosomal DNA, and by the translational and rotational positioning of the histone octamer. We tested six different pyrrole-imidazole polyamides as sequence-specific molecular calipers for DNA accessibility in nucleosomes.[101] Joel Gottesfeld and Christian Melander found that sites on nucleosomal DNA facing away from the histone octamer are fully accessible and that nucleosomes remain fully folded upon polyamide binding. Polyamides only failed to bind where sites are completely blocked by interactions with the histone octamer. Much of the DNA in the nucleosome is freely

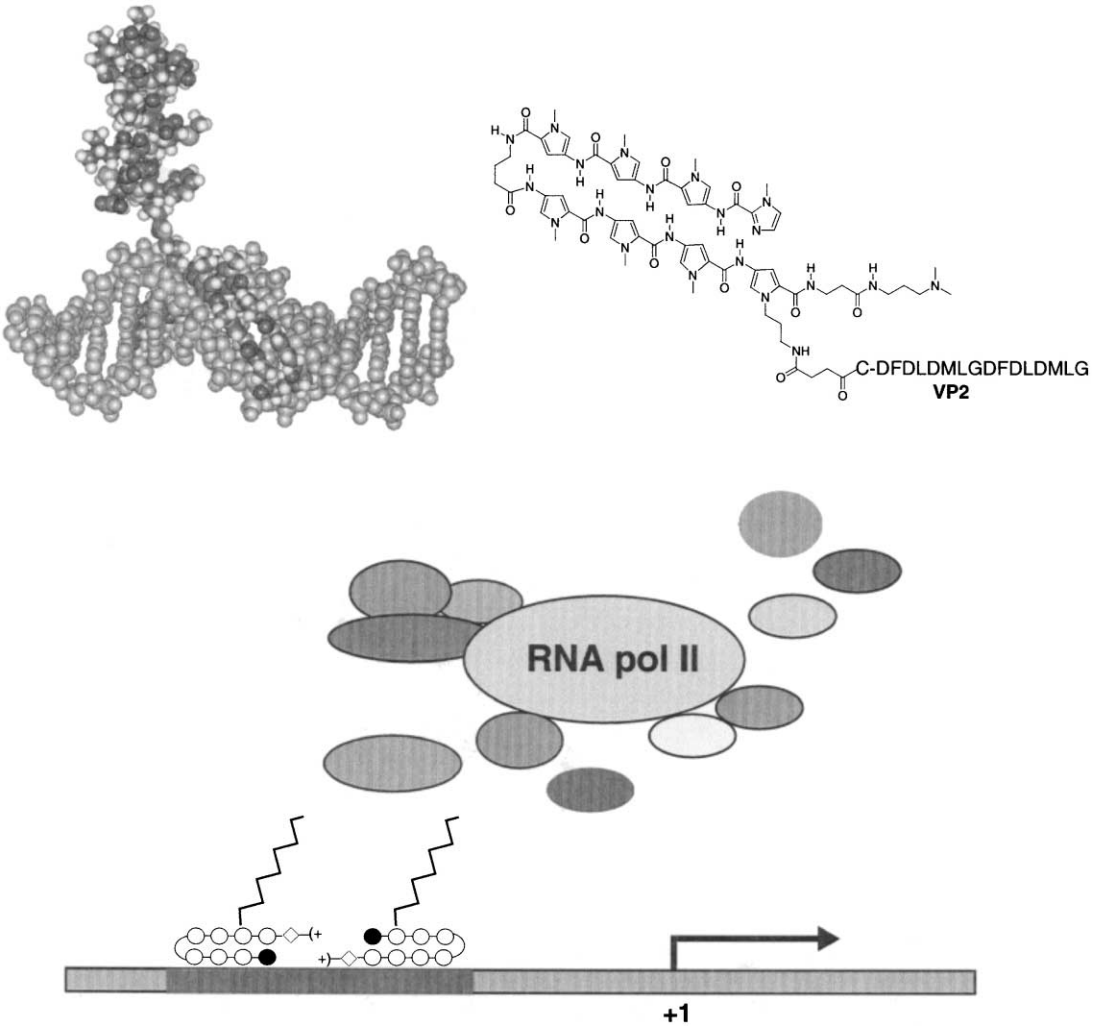


Figure 32. The activation of gene transcription by polyamide-peptide conjugates. The hairpin polyamide DNA binding domain binds to a specific promoter and the activation peptide domain VP2 binds and recruits parts of the transcription machinery.

accessible for molecular recognition in the minor groove.

Despite our early success targeting in cell culture,[95,96] we have experienced recently several examples where the inhibition results observed in cell-free systems do not occur in vivo.[102] We have taken one step backward and are now exploring the uptake and trafficking properties of polyamide-dye conjugates in living cells. Bobby Arora and Jason Belitsky find that for a variety of cell types, the polyamides are mainly in the cytoplasm, not the nucleus. Therefore, we are actively engaged in creating modified polyamides which are not only cell permeable, but will traffic to the nucleus as well.

Concluding Remarks

As I look to the future, deployment of significant resources will be necessary to answer the question whether this fundamental chemistry in the molecular recognition of DNA by small molecules will impact human medicine. Will polyamides some day form the basis of new classes of anti-infectives or new strategies for transcription therapy in the field of oncology?[103] This will require extensive development work in medicinal chemistry in parallel with high thorough-put biological screening. The pharmacokinetics, bioavailability and toxicity of polyamides remain to be established.

As I look to the future of the field of bioorganic chemistry, I see extraordinary intellectual vitality and excitement. Several of my former coworkers have followed my path into academics and it is remarkable how each one has created his or her own unique scientific vision at the interface of chemistry, biology and human medicine. Synthetic organic chemistry is one of the most power tools in modern science and, in the post genome world, organic chemists will continue to play a major role.

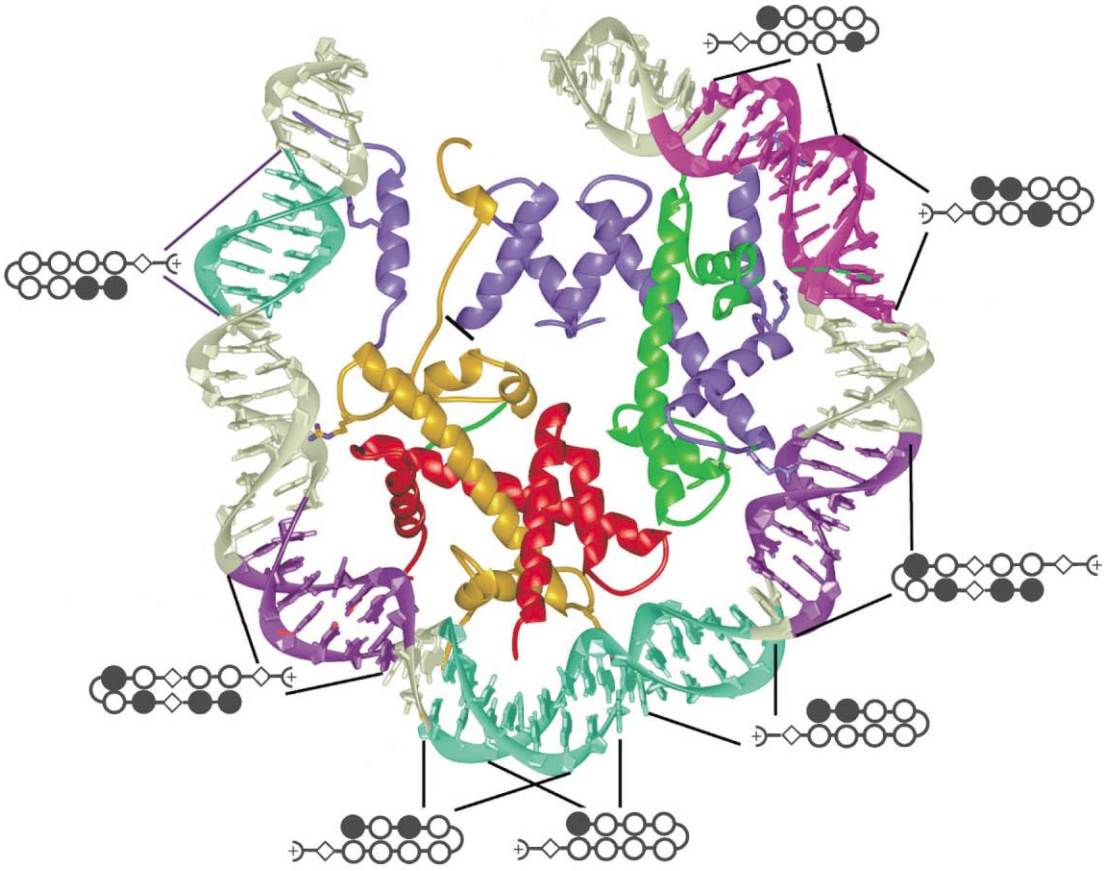


Figure 33. Polyamide binding to nucleosomal DNA. Ribbon diagram of one half of the nucleosome core particle (73 bp of DNA and associated proteins). H3 is shown in blue, H4 in green, H2A in yellow, H2B in red. Polyamide binding sites are shown as indicated.

Acknowledgements

The research described in this article was performed by a gifted and dedicated group of co-workers to whom I am indebted. I am grateful to the National Institutes of Health which has served as a primary source of funding. Three mentors have profoundly influenced me. Jerome A. Berson helped me discover the joy of research. I have been inspired by John D. Roberts' commitment to Caltech, a truly wonderful institution. Arnold O. Beckman, inventor and philanthropist, taught me the importance of giving back. I am the beneficiary of some wonderful collaborators who have enriched this entire research effort: David Wemmer (NMR structures), Doug Rees (X-ray structures), Joel Gottesfeld (gene inhibition), Mark Ptashne (gene activation), Karolin Luger (chromatin), Don Crothers (kinetics), Ken Breslauer (thermodynamics), Terry Beerman (cellular uptake). I thank Michael Waring and Claude Hélène for helpful discussions over the years.

References and Notes

1. Dervan, P. B.; Uyehara, T. *J. Am. Chem. Soc.* **1976**, *98*, 1262.
2. Dervan, P. B.; Uyehara, T.; Santilli, D. S. *J. Am. Chem. Soc.* **1979**, *101*, 2069.
3. Santilli, D. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1979**, *101*, 3663.
4. Dervan, P. B.; Santilli, D. S. *J. Am. Chem. Soc.* **1980**, *102*, 3863.
5. Pedersen, S.; Herek, J. L.; Zewail, A. H. *Science* **1994**, *266*, 1359.
6. Lerman, L. S. *J. Mol. Biol.* **1961**, *3*, 18.
7. Crothers, D. M. *J. Mol. Biol.* **1968**, *6*, 575.
8. Bresloff, J. L.; Crothers, D. M. *J. Mol. Biol.* **1975**, *95*, 103.
9. Waring, M. J.; Wakelin, F. P. C. *Nature* **1974**, *252*, 653.
10. Arcamone, F.; Penco, S.; Orezzi, P.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203*, 1064.
11. Dervan, P. B.; Becker, M. M. *J. Am. Chem. Soc.* **1978**, *100*, 1968.
12. Becker, M. M.; Dervan, P. B. *J. Am. Chem. Soc.* **1979**, *101*, 3664.
13. Hertzberg, R. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 313.

14. Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. *B. Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 5470.
15. Van Dyke, M. W.; Dervan, P. B. *Biochemistry* **1983**, *22*, 2373.
16. Van Dyke, M. W.; Dervan, P. B. *Science* **1984**, *225*, 1122.
17. Schultz, P. G.; Taylor, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1982**, *104*, 6861.
18. Taylor, J. S.; Schultz, P. G.; Dervan, P. B. *Tetrahedron* **1984**, *40*, 457.
19. Dervan, P. B. *Science* **1986**, *232*, 464.
20. Schultz, P. G.; Dervan, P. B. *J. Am. Chem. Soc.* **1983**, *105*, 7748.
21. Griffin, J. H.; Dervan, P. B. *J. Am. Chem. Soc.* **1986**, *108*, 5008.
22. Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1985**, *107*, 5528.
23. Youngquist, R. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1987**, *109*, 7564.
24. Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023.
25. Moser, H.; Dervan, P. B. *Science* **1987**, *238*, 645.
26. Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhoub, N.; Decout, J.-L.; Thuong, N.-T.; Lhomme, J.; Hélène, C. *Nucleic Acids Res.* **1987**, *15*, 7749.
27. Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059.
28. Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360.
29. Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967.
30. Horne, D. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 2435.
31. Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976.
32. Stilz, H. U.; Dervan, P. B. *Biochemistry* **1993**, *32*, 2177.
33. Lehmann, T. E.; Greenberg, W. A.; Liberles, D. A.; Wada, C. K.; Dervan, P. B. *Helv. Chim. Acta* **1997**, *80*, 2002.
34. Maher, L. J., III; Wold, B. J.; Dervan, P. B. *Science* **1989**, *245*, 725.
35. Maher, L. J., III; Dervan, P. B.; Wold, B. J. *Biochemistry* **1992**, *31*, 70.
36. Dervan, P. B.; Youngquist, R. S.; Sluka, J. P. In *Stereochemistry of Organic and Bioorganic Transformations*; Bartmann, W., Sharpless, K. B., Eds.; Verlag Chemie: Weinheim, 1986; p 221.
37. Dervan, P. B. In *Design of Enzymes and Enzyme* Models; The Robert A. Welch Foundation Conference on Chemical Research XXXI. Houston, TX, Nov 2–4, 1987; p 93.
38. Dervan, P. B. In *Nucleic Acids and Molecular Biology*; Eckstein, F. and Lilley, D. M. J., Eds., Springer-Verlag, Heidelberg, 1988; Vol. 2, p 49.
39. Dervan, P. B. In *Oligonucleotides as Inhibitors of Gene Expression*; Cohen, J. S., Ed.; MacMillan, London, 1990; p 197.
40. Strobel, S. A.; Dervan, P. B. *Science* **1990**, *249*, 73.
41. Strobel, S. A.; Dervan, P. B. *Nature* **1991**, *350*, 172.
42. Strobel, S. A.; Doucette-Stamm, L. A.; Riba, L.; Housman, D. E.; Dervan, P. B. *Science* **1991**, *254*, 1639.
43. Strobel, S. A.; Dervan, P. B. *Methods Enzymol.* **1992**, *216*, 309.
44. Povsic, T. J.; Strobel, S. A.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 5934.
45. Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376.
46. Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408.
47. Wade, W. S. PhD Thesis, California Institute of Technology, **1989**.
48. Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723.
49. Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783.
50. 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.
51. Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
52. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572.
53. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 3663.
54. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983.
55. Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160.
56. Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1995**, *117*, 3325.
57. Baird, E. E.; Dervan, P. B. J. *Am. Chem. Soc.* **1996**, *118*, 6141.
58. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6147.
59. Parks, M. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6153.
60. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1996**, *3*, 369.
61. White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
62. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
63. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198.
64. Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6981.
65. Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8306.
66. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953.
67. White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756.
68. Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 7636.
69. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600.
70. White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569.
71. Greenberg, W. A.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1998**, *4*, 796.
72. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem. Int. Ed.* **1998**, *37*, 1421.
73. Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219.
74. Herman, D. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.
75. Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121.
76. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1999**, *5*, 975.
77. Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113.
78. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534.
79. Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. *Biochemistry* **1999**, *38*, 2143.
80. Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Angew. Chem.* **1996**, *35*, 1487.
81. Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6778.
82. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struc. Biol.* **1998**, *5*, 104.

83. White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
84. Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Science* **1998**, *282*, 111.
85. White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260.
86. Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621.
87. Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557.
88. Ellervik, U.; Wang, C. C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 9354.
89. Melander, C.; Herman, D. M.; Dervan, P. B. *Chem. A. Eur. J.* **2000**, *6*, 4487.
90. Wurtz, N. R.; Dervan, P. B. *Chem. Biol.* **2000**, *7*, 153.
91. Chang, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2000**, *122*, 4856.
92. Bremer, R. E.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1998**, *5*, 119.
93. Winston, R. L.; Ehley, J.; Baird, E. E.; Dervan, P. B.; Gottesfeld, J. M. *Biochemistry* **2000**, *39*, 9092.
94. Gottesfeld, J. M.; Turner, J. M.; Dervan, P. B. *Gene Expression* **2000**, *9*, 77.
95. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
96. Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12890.
97. Ho, S. N.; Biggar, S. R.; Spencer, D. M.; Schreiber, S. L.; Crabtee, G. R. *Nature* **1996**, *382*, 822.
98. Beerli, R. R.; Dreier, B.; Barbas, C. F. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1495.
99. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3930.
100. Ansari, A. Z.; Mapp, A. K.; Nguyen, D. H.; Dervan, P. B.; Ptashne, M. *Chem. Biol.* **2001**, *8*, 583.
101. Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615.
102. Chiang, S.-Y.; Bürli, R. W.; Benz, C. C.; Gawron, L.; Scott, G. K.; Dervan, P. B.; Beerman, T. A. *J. Biol. Chem.* **2000**, *275*, 24246.
103. Pandolfi, P. P. *Oncogene* **2001**, *20*, 3116.

Peter B. Dervan, PhD is the Bren Professor of Chemistry at the California Institute of Technology.

Peter B. Dervan is a native of Boston, Massachusetts where he was born in June 1945. After receiving his BS degree in chemistry in 1967 at Boston College, he began research in physical organic chemistry under the supervision of J. A. Berson. After earning his PhD degree at Yale University in 1972, he spent one-half year at Stanford as a postdoctoral fellow in the laboratory of G. Van Tamelen. In 1973, he went to Pasadena to take up a faculty appointment at the California Institute of Technology. He resides near the Caltech campus in San Marino with his wife Jacqueline K. Barton, daughter Elizabeth and son Andrew.

Dervan's research interests are at the interface of chemistry and biology, in particular, studies directed toward understanding the chemical principles for the sequence-specific recognition of the genetic material, DNA. The development of a code to distinguish each of the four Watson–Crick base pairs by small molecules might allow the regulation of gene expression in living cells. In the post genome era where many diseases are related to specific gene expression profiles, the ability to reprogram a human cell could have profound implications for human medicine.

Dervan is a member of the National Academy of Sciences, the Institute of Medicine, the American Academy of Arts & Sciences, and a Foreign Member of the French Academy of Sciences. His awards include the Harrison Howe Award (1988), Arthur C. Cope Award (1993), Willard Gibbs Medal (1993), Nichols Medal (1994), Maison de la Chimie Foundation Prize (1996), Remsen Award (1998), Kirkwood Medal (1998), Alfred Bader Award (1999), Linus Pauling Medal (1999), Richard C. Tolman Medal (1999), and the Tetrahedron Prize (2000).

Dervan is considered an outstanding teacher, having won several excellence in teaching awards from the Caltech undergraduates. He has mentored over 140 predoctoral and postdoctoral co-workers. In addition to teaching, research and administrative duties, he serves on several Scientific Advisory Boards for the pharmaceutical and biotechnology industries, as well as the Robert A. Welch Foundation.